Essays in Political Economy and on Networks

by

Nathan Joseph Canen

BEcon, Fundação Getulio Vargas (FGV/EPGE), 2011
MEcon, Fundação Getulio Vargas (FGV/EPGE), 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

The Faculty of Graduate and Postdoctoral Studies
(Economics)

The University of British Columbia
(Vancouver)

September 2018

© Nathan Joseph Canen, 2018
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled: Essays in Political Economy and on Networks, submitted by Nathan Joseph Canen in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics.

Examining Committee:
Francesco Trebbi, Economics Supervisor
Paul Schrimpf, Economics Supervisory Committee Member
Sergei Severinov, Economics University Examiner
Richard Johnston, Political Science University Examiner
Arianna Degan, Economics, UQAM External Examiner
Abstract

This thesis studies topics in political economy and the economics of networks.

In Chapter 2, we present and structurally estimate a model of endogenous network formation and legislative activity of politicians. Employing data on social and legislative effort of members of the 105th-110th U.S. Congresses (1997-2009), we find that there are substantial complementarities between the efforts of politicians, both within and across parties.

Chapter 3 considers the econometrics of incomplete information games on networks. This chapter develops a tractable empirical model of linear interactions where each agent, after observing part of his neighbors’ types, not knowing the full network of how information is transmitted, uses linear best responses. This allows the researcher to perform asymptotic inference without having to observe all the players in the game or having to know precisely the sampling process. The usefulness of this procedure is shown with an application to the provision of public goods across municipalities in Colombia.

Chapter 4 studies the sources of party polarization in the U.S. Congress. Polarization is not just the result of changes in the ideology of individual legislators, but also of changes in the ability of political parties to discipline (whip) their members and of the deliberate agenda setting by their leadership. This chapter evaluates quantitatively the importance of these three components in driving polarization through a novel identification approach based on previously untapped whip count data and a structural model of legislative activity.

In the final chapter, I turn my attention to the voters’ side in political economy models. Surveys, polling data and media reports indicate that voters often choose whom to vote for at different stages in the political campaign. I develop a model of costly information acquisition that rationalizes these observations. The model implies a key tradeoff between the cost of acquiring information, and the gain such information brings. Under this framework, I show that information blackouts (i.e. forbidding release of campaigning or polling information before the election) generates welfare losses of around 1-2%.
Lay Summary

This thesis is based on a series of contributions to the fields of political economy and the economics of networks. In Chapter 2, we explore how politicians form social connections in Congress. We provide an empirical model that allows an in depth analysis of these strategic decisions. Chapter 3 provides methodological contributions to the study of strategic interactions when agents may be connected on a network. In particular, we focus on tools for models of incomplete information where agents are better informed about their neighbors than of others. Leveraging on new data and a new theoretical framework, Chapter 4 shows how observed polarization may be driven by three separate factors: the polarization of individual ideologies, changes in legislative agenda and party discipline. We show how to empirically decompose its sources. Finally, I look at the timing of information on voter’s decisions through an empirical model of information accumulation.
Preface

Chapter 2 of this thesis is coauthored with Matthew O. Jackson and Francesco Trebbi. All coauthors contributed equally to all aspects of the project.

Chapter 3 of this thesis is based on joint work with Kyungchul Song and Jacob Schwartz. All coauthors contributed equally to all aspects of the project, including developing the model, the Monte Carlo simulations and the empirical example. Part of this work also appears in Jacob Schwartz’s Ph.D dissertation at the University of British Columbia. He has given his consent in sharing this chapter.

Chapter 4 of this thesis is coauthored with Chad Kendall and Francesco Trebbi. All coauthors contributed equally to all aspects of the project.
# Table of Contents

Abstract .......................... iii
Lay Summary ........................ iv
Preface ............................. v
Table of Contents ........................ vi
List of Tables ........................ x
List of Figures ........................ xii
Acknowledgments ........................ xiii

1 Introduction ........................ 1

2 Endogenous Networks and Legislative Activity ........................ 3
  2.1 Introduction ........................ 3
  2.1.1 Relation to the Literature ........................ 5
  2.2 The Model ........................ 6
  2.2.1 Legislators, Parties, and Partisanship ........................ 6
  2.2.2 Legislative effort and Preferences ........................ 8
  2.2.3 A Micro-Foundation Built Upon Reelection Preferences ........................ 9
  2.2.4 Solving For Equilibrium ........................ 12
  2.2.5 Pareto Efficient Efforts ........................ 14
  2.2.6 Preliminaries to Estimation ........................ 15
  2.3 Data ........................ 17
  2.4 Estimation ........................ 23
### 2.4.1 Moment Equations

23

### 2.4.2 Estimation via GMM

25

### 2.5 Results

26

#### 2.5.1 Fit and Discussion

30

### 2.6 Assessment and Counterfactuals

32

#### 2.6.1 Stability of Equilibrium and Other Equilibrium Properties

34

#### 2.6.2 Counterfactuals

36

### 2.7 Conclusions

40

### 3 Estimating Local Interactions Among Many Agents Who Observe Their Neighbors

42

#### 3.1 Introduction

42

#### 3.2 Strategic Interactions with Information Sharing

47

##### 3.2.1 A Model of Interactions with Information Sharing

47

##### 3.2.2 Predictions from Rationality

50

##### 3.2.3 Belief Projection and Best Linear Responses

52

##### 3.2.4 The External Validity of Network Externality

57

#### 3.3 Econometric Inference

61

##### 3.3.1 General Overview

61

##### 3.3.2 Asymptotic Theory

70

#### 3.4 A Monte Carlo Simulation Study

71

#### 3.5 Empirical Application: State Presence across Municipalities

73

##### 3.5.1 Motivation and Background

73

##### 3.5.2 Empirical Set-up

76

##### 3.5.3 Model Specification

77

##### 3.5.4 Results

78

#### 3.6 Conclusion

80

### 4 Unbundling Polarization

81

#### 4.1 Introduction

81

#### 4.2 Model

86

##### 4.2.1 Preliminaries

87

##### 4.2.2 Preferences

87

##### 4.2.3 Information and Timing

88

#### 4.3 Analysis

90

##### 4.3.1 Roll Call Votes

90
A Appendix to Chapter 2 .............................. 163
  A.1 Proofs ........................................... 163
    A.1.1 Best Response Dynamics ..................... 169
  A.2 Parametric Identification, with set identification of \( \phi \) .......... 170
    A.2.1 The need for exponential measurement errors even when \( \alpha \) is parametric. ........................................... 173
  A.3 Rewriting the Model in terms of Moment conditions over \( i \) ........ 174
  A.4 Details on Estimation ................................ 175
    A.4.1 OLS and plug-in Approach as Starting Values for Optimization .. 176
    A.4.2 Computation of the Optimal Weighting Matrix and of Standard Errors 176
  A.5 Computation of Comparative Statics ................................. 177
  A.6 Identification and Estimation using second moments of the proxies of \( (s^*_i, x^*_i) \) ........ 179
    A.6.1 Estimation with Second Moment Conditions on the Proxies of \( s^*_i, x^*_i \) 180
    A.6.2 Results under Restrictions on the Second Moments of the Proxies for \( (s^*_i, x^*_i) \) .................................................. 181
  A.7 Identification with Nonparametric \( \alpha \) ............................. 183

B Appendix to Chapter 3 ..................................... 185
  B.1 Construction of the Estimator for the Asymptotic Covariance Matrix . . 185
  B.2 Proofs of Theorems 3.2.1 - 3.3.1 .................................... 188

C Appendix to Chapter 4 ..................................... 205
  C.1 Proofs ........................................... 205
  C.2 The choices from Party “R” ..................................... 218
    C.2.1 Agenda Setting ......................................... 219

D Appendix to Chapter 5 ..................................... 221
  D.1 Additional Tables and Figures ............................... 221
  D.2 Proofs ........................................... 229
  D.3 Israeli Political System ................................. 233
    D.3.1 Israel in 2006 ........................................ 233
  D.4 First Stage Estimation of the Priors ............................ 234
  D.5 Optimization Routine ...................................... 234
List of Tables

Table 2.1 Summary Statistics ........................................... 22
Table 2.2 Main Results, Specification 1 ................................. 27
Table 2.3 Differences in the Distributions of $\alpha_i$ Across Parties ........................................... 28
Table 2.4 Main Results, Specification 2 ................................. 31
Table 2.5 Model Fit: Correlation of Estimated Network of the Model to the Cosponsorship Networks in the Data ........................................... 33
Table 2.6 Estimated and High Effort Equilibria ......................... 35
Table 2.7 Counterfactual in $\gamma$: Predicted Probability of Bill Approval ........................................... 37
Table 2.8 Counterfactuals in $c$: Predicted (Proportional) Change in the (Mean) Probability of Bill Approval ........................................... 38
Table 2.9 Counterfactuals in $\alpha$: Looking at the Changes in (Ex-Ante) predicted probability of Emergency Crisis bills in the 110th Congress, if the Republicans who lost their seats remained ........................................... 41
Table 3.1 The Characteristics of the Payoff Graphs ..................... 58
Table 3.2 The Degree Characteristics of the Graphs Used in the Simulation Study ........................................... 73
Table 3.3 The Empirical Coverage Probability and Average Length of Confidence Intervals for $\beta_0$ at 95% Nominal Level ........................................... 74
Table 3.4 The Empirical Coverage Probability and Average Length of Confidence Intervals for $a' \rho_0$ at 95% Nominal Level ........................................... 75
Table 3.5 State Presence and Networks Effects across Colombian Municipalities ........................................... 79
Table 4.1 Summary Statistics on Bill Selection .......................... 114
Table 4.2 Number of Whips per Party .................................... 114
Table 4.3 Main Estimates from the First Step ........................... 115
Table 4.4 Decomposition of Polarization in Ideologies and Whipping ........................................... 116
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>Total Number of Cosponsorships per Congressional cycle</td>
<td>18</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Number of Cosponsorships Within and Across Parties per Congressional cycle</td>
<td>18</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Correlation between the raw data of log(1+Words) in Floor Speeches and Cosponsorship decisions.</td>
<td>20</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Correlation between the raw data of Roll Call Effort and Cosponsorship decisions.</td>
<td>21</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Distribution of Estimated $\alpha$ over Time</td>
<td>26</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Estimated Probability of Approval - Democrats, Congresses 105-110</td>
<td>29</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>Estimated Probability of Approval - Republicans, Congresses 105-110</td>
<td>30</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Network Externality Comparison Between Equilibrium and Behavioral Models: Erdős-Rényi Graphs</td>
<td>60</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Network Externality Comparison Between Equilibrium and Behavioral Models: Barabási-Albert Graphs</td>
<td>61</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Degree Distribution of $G_P$</td>
<td>76</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Average Network Externality from being a Department Capital</td>
<td>80</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Timeline</td>
<td>89</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Example of Value Functions</td>
<td>96</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Votes with the Majority Party at Whip Counts and Roll Calls</td>
<td>112</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Estimated ideologies, $\theta_i$, per Party over Time</td>
<td>112</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Estimated Ideologies from the Model $\theta_i$ compared to DWNominate</td>
<td>113</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Estimated $y^{\text{max}}$ per Party over Time</td>
<td>113</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>Distribution of Estimated Priors (Precision)</td>
<td>143</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>Model Fit</td>
<td>147</td>
</tr>
</tbody>
</table>
Acknowledgments

I am deeply indebted to my supervisor, Francesco Trebbi, for all of his dedication and support. I am particularly thankful for his immense generosity, including the many hours I have learned while exchanging ideas; invaluable to my growth. I would also like to thank my committee members, Hiro Kasahara and Paul Schrimpf, for their continuous support throughout the years. I am also very grateful to my examining committee, Richard Johnston, Sergei Severinov and, especially, Arianna Degan, for their valuable comments and thorough reading of this dissertation.

I have benefited from the support of many other UBC faculty and staff. In particular, I would like to thank Maureen Chin for all her help. Furthermore, I thank the generous funding from UBC and the Vancouver School of Economics.

I have learned much from my other co-authors Matthew O. Jackson, Chad Kendall, Jacob Schwartz and Kevin Song, for whom I would like to express my gratitude. These have included long conversations that have resulted in a deeper understanding of the nature of economics and econometric models. I take these lessons on with me. Support has also come from my friends and colleagues. I couldn’t have wished for a better environment within the School of Economics. I would particularly like to thank Anderson, Anujit, Denis, Gaëlle, Hugo, Jacob, João, Pierluca, Rogerio, Tom and Tzu-Ting, who have shared this path with me at different times. To my friends outside of economics, I owe so much to your emotional support.

Last but not least, I would like to thank my family. My parents, Ana and Alberto, and my sister, Doris, have always been present, encouraging and nurturing. They have inspired me to pursue my dreams, to do my best, while grounded in what really matters. The Fortes-Kursan family provided me a family away from home. Finally, thanks to Cadi, who has been with me through the best and the worst.
Chapter 1

Introduction

This thesis explores topics in political economy and the economics of networks. In particular, it investigates questions related to the organization of political institutions, such as the social networks formed by politicians, the organization and disciplining of politicians by their party leaders, and the differential impact of information on the timing of voting decisions. To do so, I also introduce new methodologies and new data that can address these questions.

The second chapter of my dissertation, in joint work with Matthew O. Jackson and Francesco Trebbi, addresses how politicians form networks in Congress. These networks can be beneficial for the approval of a politician’s legislation, as they can gather support from their own network which may help to pass a bill. Previous work had focused on analyzing the shape of the network of politicians, with connections measured by bill cosponsorships. However, the decision of whom to connect with is itself a choice, taken while anticipating future returns of such connections. In our model, we characterize the choices of social effort (used to connect with others) and legislative effort, when passing a bill requires both. The model provides a rich set of testable predictions. These include who becomes more connected, how the amount of effort changes according to the cost of effort, as well as the impacts of changing the composition of Congress. The model is then estimated, and we discuss how changes to the composition of Congress can affect the passage of major bills, such as the Emergency Economic Stabilization Act in Congress 110.

Chapter 3 provides new methodological tools for the estimation of such models. This chapter is entitled “Estimating Local Interactions Among Many Agents Who Observe Their Neighbors” and is based on joint work with Jacob Schwartz and Kyungchul Song. We study games on networks when there might be incomplete information: agents are better informed
about their neighbors than someone who they are not connected to. Our results provide an estimator and statistical tests with good properties. One key issue addressed with our methodology is that usual sampling techniques do not work with networks: a random sample of a network will look very different than the original network studied. This is because each member’s connections are unlike another’s. Another key issue addressed is how to estimate models of games on networks, even when the researcher does not observe the full network in the data. This has multiple applications in political economy, as illustrated in an empirical application of the provision of public goods across municipalities in Colombia.

Having studied the choices of individual legislators in a previous chapter, Chapter 4 then studies the decisions and incentives within parties. This is done with a particular focus on understanding the sources of polarization, in joint work with Chad Kendall and Francesco Trebbi. We量化 the importance of agenda setting, drifting ideologies and increasing party discipline in explaining observed polarization in roll call votes. Separately identifying such effects is nontrivial: we only observe outcomes that are a result of all of them together. To solve this, we use new data with a new theoretical framework. The new data comes from whip counts conducted by party leaders, which gauges legislators’ positions on a bill before they actually vote them on the Congress floor. We model the inner workings of party functioning in the U.S. Congress, with a theoretical model that addresses both agenda setting and party discipline. Crucially, our empirical model allows us to disentangle party effects from party discipline, agenda setting and ideology.

My final chapter then looks at voters’ decisions. I start from the observation that voters make decisions on whom to vote for at different points in time. While many of them decide on who to vote for in the last day or the last week of the campaign, others know all along who they are going to vote for. This means they are willing to commit on their decision, even when they might learn useful new information in the future. To explain these differences, I propose a model of costly information accumulation. For some voters, stopping early is due to high costs of information, while for others, the benefit of information is small. The model allows me to disentangle what drives the stopping times observed for different subgroups in the data. I then look at the impacts of a blackout policy, by which information is banned from voters in a day or a week before the election. Such a policy, although aiming for fairness, harms voters as it affects only those who still wish to accumulate information.
Chapter 2

Endogenous Networks and Legislative Activity

2.1 Introduction

Deliberative bodies, especially larger ones, rely on informal interactions in order to function productively. Information percolates through informal political networks and individuals form relationships with each other in order to craft and pass legislation. Because of the salient role of interpersonal ties in the legislative process, its study dates at least to the 1930s ([154]), but only in the last fifteen years has this area of research come to more prominence ([115]).

The challenge of simultaneously modeling network formation and political decision-making may be a reason for this delayed uptake, and it still presents a significant hurdle for advancing our understanding of how legislatures internally operate. The construction and analysis of the model presented here helps fill this gap.

Our model generalizes the tractable and powerful framework of [36]. In their model agents (who in our setting are to be thought of as elected representatives) choose both how much socializing to do with other politicians and how much legislative effort to exert. Socializing efforts result in randomly formed relationships that increase the success of legislative efforts, and so social and legislative efforts are complements. Importantly, social and legislative efforts are also complementary to those of the other politicians with whom a given politician has ties, both within and outside of his/her party. In [36], however, relationships form completely at random. This misses realistic biases in interaction (such as

---

1This chapter is a joint work with Matthew O. Jackson and Francesco Trebbi.
homophily) that characterize many social environments, especially ones political in nature. In our generalization, we allow social ties to form at a different rate within compared to across groups – so that, for instance, legislators can collaborate with members of their own party at a different rate than with members of the opposition.

We structurally estimate our model employing data on cosponsorship and legislative effort of members of House of Representatives from the 105th-110th U.S. Congresses.

A first empirical finding is that the complementarities among politicians are significant and stable across our sample period. The social marginal multiplier on legislative effort is estimated to be between a tenth and a third of the direct incentive for legislative effort. This means that a nontrivial fraction of incentives for efforts of politicians are driven by what other politicians are doing. In summary, socializing appears vital to bill passage and, in addition, this has been consistently true over time.

We then examine differences between Democrats and Republicans. We find that the two parties have different base payoffs from passing legislation, both in terms of average and variance across party members (both are higher for the Democrats). These differences lead to higher levels of social and legislative efforts of both types by Democrats, all else held equal.

Further, our generalization allows social interaction to be biased towards one’s own party. This turns out to be quantitatively important, as it allows for asymmetries in behavior across political parties, which clearly appear in the congressional data we explore in our application. Specifically, we find evidence that partisan bias is an empirically relevant feature and a model with biased interactions fits the data significantly better than a model with no bias. We also show evidence, however, that social interaction in the U.S. Congress is far from being an exclusively partisan affair. The data appears more nuanced than the common narrative of a completely balkanized Congress, segregated along party lines, that has emerged from recent literature mostly based on post-1980 congressional roll call evidence ([128]; [69]). Intermediate levels of partisanship in our environment – a partisan bias is in the range of 8 – 10 percent which we can estimate in an extended version of our model, for example – fits the data better than a fully partisan model where 100 percent of interactions are exclusively within party. Intuitively, it is hard to reconcile the thousands of bipartisan cosponsorships observed in recent congressional data with an hypothesis of unmitigated polarization between parties. Our conjecture is that the stark posturing required by the media,

\[^2\text{As it will be clear in the analysis that follows, the parameter multiplying the full product of social and legislative efforts ranges from 0.05 to 0.08, which when multiplied by other legislators’ efforts of 3 to 4.5, and social efforts around 1, leads to a multiplier of 0.1 to 0.32. This is compared to direct incentives for legislative activity ranging from 1.0 to 1.4.}\]
the focus on divisive language ([78], [77]), and some metrics of formal political activity may miss some dimensions of bipartisan interaction that more informally may take place among legislators.

In the final part of this chapter, we assess the specific equilibrium at play within each Congress among the multiple equilibria typically present in this class of games.\(^3\) We show that the estimated equilibria are interior and stable and that social effort is (inefficiently) under-provided in all Congresses.

Finally, our estimated model enables us to perform counterfactuals. We show large responses to changes in the relative costs of socializing. In addition, we examine whether the amount of legislation that was generated in the emergency response to the 2008-09 financial crisis would have changed if the Democrats had not taken over the House, but it had remained as it was before the crisis. Here we find a quantitatively small change (between 3 – 5 percent) in the amount of legislation – so even though the preferences changed, the difference would be in the fractions of social to legislative efforts, but not in the end outcome of this particular episode.

### 2.1.1 Relation to the Literature

From the theoretical perspective, this chapter contributes to a literature that examines peer-influenced behavior when accounting for the endogeneity of networks.\(^4\) In particular, we generalize the model of [36] to include biases (homophily) in network formation. The partisan biases in interactions matter significantly in our empirical application, and should also in a variety of other applications where social interaction may have in-group versus out-group components.

This chapter also contributes to the earlier literature that showed that social networks matter in legislative environments. For instance, [74] used a connectedness measure based on cosponsorship to show that more connected members of Congress are able to get more amendments approved and have more success on roll call votes on their sponsored bills.\(^5\) Again using cosponsorship links, [48] show that Congress can be understood as a small-world network, as it appears subdivided in multiple dense parts tied together by some intermediaries. These network features correlate with legislative productivity over time (number of important laws passed, as defined by [123]).\(^6\)

---

\(^3\) For a complete review of approaches to empirical models with multiple equilibria, see [57].

\(^4\) See [83], [82], [121], [96], [15], [97] (and see [94], [95], [98] for surveys of the network formation and games on networks literatures).

\(^5\) A similar study is [175].

\(^6\) There is other work on cosponsorship. For example, [8], [109], and [31] study the incentives for cosponsorship.
The network analysis of legislation is growing, and provides increasing evidence that social relationships matter substantially and are causal in nature. For example, [106] shows a correlation between bill survival and weak ties of the sponsor for eight state legislatures and for the US House of Representatives. [50] employ identification restrictions aimed at ascertaining causal effects of networks on voting behavior (using the quasi-at-random seating arrangements of Freshman Senators). [70] studies the role of exogenously shifted social connections within the European Parliament also using seating arrangements.

Importantly, all of these papers take networks as exogenous. In contrast, our endogenous analysis of the network enables us to see how incentives to socialize differ across parties and have changed over time.

The rest of this chapter is organized as follows. Section 2.2 derives the empirical model and discusses identification. Section 2.3 presents the data. Section 2.4 illustrates the estimation and the moment conditions. The estimation results and the assessment of the model’s fit are reported in Section 2.5. Section 2.6 contains the main counterfactuals and discusses the equilibria of the model from an empirical perspective. Section 2.7 concludes.

2.2 The Model

2.2.1 Legislators, Parties, and Partisanship

The legislature, henceforth referred to as “Congress”, is composed of \( n \) members and, for simplicity, we focus on one chamber (e.g. the House). \( \mathcal{N} = \{1, 2, \ldots, n\} \) represents the set of politicians in Congress. Clearly, although we refer to Congress, the model applies to a variety of deliberative bodies, legislatures, committees, and organizations more generally.

The set of politicians is partitioned into parties, with a generic party denoted \( P_\ell \).

Each party \( P_\ell \) has a level of partisanship \( p_\ell \). In particular, members of party \( \ell \) spend a fraction of their interaction, \( p_\ell \), at exclusively party \( \ell \) events, so only mixing and meeting with own-party events, and the remainder, \( 1 - p_\ell \), at events in which they mix with others in different settings (focusing on ideological similarity, tenure, etc.). Beyond their role in social networks, [170] study the signaling content of cosponsorships, noting that cosponsorship is a cheap way of signaling to the median voter about one’s congressional activity. They identify three different explanations for cosponsorships and their possible signaling impact: (i) bandwagoning (signaling strong support for the bill), (ii) ideology, and (iii) expertise. They find a null to moderate effect of cosponsorship on bill success, as measured as successive progress of the bills through Congress hurdles. [104] instead point out that the timing of cosponsorships would indicate that it is not as much a signaling to voters, as to other politicians (for example, they show that extremists seem to cosponsor earlier). Still in the context of bill sponsorships, [10] find correlations with legislative productivity (i.e. the bill passing through different stages in Congress) for Congress member who sponsor more bills and use more floor time (albeit at a declining marginal rate).
members of other parties. This can include party and caucus meetings, joint sessions, fund-raising events, committee works, social gatherings and formal events, etc. For our period of interest, examples of party-specific events are closed sessions called Party Conferences for Republicans and Party Caucuses for Democrats (their respective chairs represent the number 3 position of official party leadership rankings).

Politician \( i \) is from party \( P(i) \), and \( p(i) \) denotes the level of partisanship of politician \( i \)'s party.

In our empirical analysis, there are two parties, 1, 2, and then we index the \( n \) politicians so that the first \( q \) of them belong to \( P_1 = \{1; \ldots ; q\} \) and the remainder to \( P_2 = \{q + 1; \ldots ; n\} \). Let \( q \geq n/2 \), so that party 1 is the majority party.

### Socializing

Each politician chooses an effort level of how much he or she socializes (i.e., how much time he or she spends interacting with other politicians), denoted \( s_i \in \mathbb{R}_+ \). It is via this socializing that he or she forms connections with other politicians.

The network \( \{g_{i,j}\}_{i,j \in \mathcal{N}} \) that arises from the vector of social efforts, \( s \), is described by

\[
g_{i,j}(s) = s_is_jm_{i,j}(s), \tag{2.1}
\]

where if \( j \in P(i) \) then

\[
m_{i,j}(s) = p(i)\frac{p(j)}{\sum_{k \in P(i), k \neq i} p(k)s_k} + (1 - p(i))\frac{(1 - p(j))}{\sum_{k \neq i} (1 - p(k))s_k}, \tag{2.2}
\]

and if \( j \notin P(i) \) then

\[
m_{i,j}(s) = (1 - p(i))\frac{(1 - p(j))}{\sum_{k \neq i} (1 - p(k))s_k}. \tag{2.3}
\]

So, politicians meet their own party members in two ways: at their own events and at bipartisan events. They meet members of the other party only at the bipartisan events. Politicians are met with the relative frequency with which they are present at events.

---

7This is for the case in which \( s_j > 0 \) for at least two people in each party. If other agents are not putting in social effort, then there can be nobody to match with, and then some of these equations do not apply (they divide by 0, and if all \( s \) are 0). In those cases the matching is described as follows. If at most one \( s_j > 0 \), then set \( m_{ij} = 0 \) for all \( i,j \) and the entire network equal to 0. If there are at least two people with \( s_j > 0 \), but also at least one party with \( s_j > 0 \) for no more than one agent, then set \( m_{ij} = g_{i,j} = 0 \) for all members of a party that has nor more than one \( s_j > 0 \), but use the remaining above specified equations for \( g_{i,j} \)'s for any other combinations.
In Appendix, we show that:

$$
\sum_{j \neq i} s_i s_j m_{ij}(s) = s_i,
$$

(2.4)

so that the total number of connections that $i$ makes is proportional relative to $s_i$.

When $p_\ell = 0$ for each $\ell$, then this simplifies to coincide with the model of [36]. When $p_\ell = 1$ for each $\ell$, instead, each party is completely cut off from the other. Then, within each party again [36] applies.

### 2.2.2 Legislative effort and Preferences

The other choice of politicians is their legislative effort $x_i \in \mathbb{R}_+$. The benefits from legislative efforts are described by:

$$
\alpha_i x_i + \phi \sum_{j \neq i} s_i s_j m_{ij}(s)x_i x_j.
$$

(2.5)

As in a large class of models, of which [36] is a salient instance, there is a direct benefit from private effort, with idiosyncratic weight $\alpha_i$. In addition, there are complementarities in legislative efforts between politicians who have formed connections: the more effort they both expend, the more likely their legislation is to pass. The size of this interaction effect is governed by the parameter $\phi$, whose quantification is a relevant goal in the empirical analysis that follows.

Both forms of effort are costly for a politician. The cost of legislative effort is given by $\frac{c}{2} x_i^2$, with $c > 0$, and the total cost of socializing is given by $\frac{1}{2} s_i^2$. The parameter $c$ governs the relative cost of legislative effort to social effort.

Taken together, the politician’s preferences are the amount of legislation that he or she produces less the costs of legislative and social efforts. This is given by:

$$
\tilde{u}_i(x_i, x_{-i}, s_i, s_{-i}) = \alpha_i x_i + \phi \sum_{j \neq i} s_i s_j m_{ij}(s)x_i x_j - \frac{c}{2} x_i^2 - \frac{1}{2} s_i^2.
$$

(2.6)

---

8Although we use cosponsorships of Congressional bills as proxies for socializing efforts, we choose not model the directed cosponsoring by $i$ of any specific Congress member $j$ (nor other $i$, $j$ directed linkages). Rather, we posit a generic social effort $s_i$. Pairwise cosponsorships between $i, j \in \mathcal{N}$ have been extensively investigated in a literature largely led by [74]. Here, we make use of this information in Section 2.5, where we consider how well our model-predicted links $g_{i,j}$ fit the auxiliary information of $i,j$ pairwise cosponsorships as a form of external validation of our approach.
2.2.3 A Micro-Foundation Built Upon Reelection Preferences

There are many different ways in which we could justify the preferences in (2.6), as the natural presumption is that politicians care to maximize the legislation that they pass. Here, we posit a realistic microfoundation, which we will bring directly to the data.

Politicians care about being reelected and can affect the probability of being reelected by exerting effort in Congress and by building connections instrumental to having specific legislation passed (e.g. policy favorable to the politician’s constituents).

Each politician anticipates these effects on his/her reelection chances. More specifically, each congressional cycle has two periods, 1 and 2, where the second period provides the reelection incentives that drive activity in the first period. Politicians are career motivated and exert costly efforts with the aim of increasing their chances of being reelected.

In period 1, each Congress member can present a policy proposal, which for brevity we refer to as a “bill”. The bill consists of a policy goal the Congress member intends to fulfill, for instance passing a statute targeted to his or her constituency, landing a subsidy, or obtaining an earmark beneficial to firms in the home district. We describe below how getting \(i\)’s policy goal fulfilled maps into an increase in \(i\)’s chances of being reelected.

Suppose a politician’s utility is given by:

\[
U_i = Pr(\text{reelected}) - \frac{c}{2}x_i^2 - \frac{1}{2}s_i^2. \tag{2.7}
\]

The choice of \(x_i\), the level of legislative activity exerted by \(i\), affects the support for \(i\)’s legislation, \(Y_i\), through a function:

\[
Y_i = \varepsilon_i x_i \left( \sum_{j \in \mathcal{N}} g_{i,j}(s)x_j \right). \tag{2.8}
\]

Both \(i\)’s own legislative effort, \(x_i\), and that of his or her connections in the network, \(\sum_{j \in \mathcal{N}} g_{i,j}(s)x_j\), matter for the ultimate support received by \(i\)’s bill.

\(Y_i\) is stochastic and depends also on a random shock \(\varepsilon_i\), assumed to be standard Pareto distributed with scale parameter \(\gamma > 0\) and i.i.d. across politicians. We assume that \(\varepsilon_i\) is realized after the choice of \(x\), the vector of \(x_j\) across all politicians \(j \in \mathcal{N}\). Because \(\varepsilon_i\) is a shock following the realized legislative support, \(i\) must take expectations over its value when choosing \((x_i, s_i)\). Also notice that each link between politician \(i\) and \(j\) is an endogenous function of the social efforts of everybody else, hence the dependency \(g_{i,j}(s)\) on \(s\), the vector of \(s_j\) efforts across all politicians \(j \in \mathcal{N}\).
The bill is approved if $Y_i > m$, where $m > 0$ is a generic institutional threshold\textsuperscript{9} The probability of having the bill approved is thus given by:

$$Pr(Y_i > m) = Pr\left(\frac{m}{x_i (\sum_{j \in \mathcal{N}} g_{i,j}(s)x_j)} > e_i\right)$$

$$= \left(\frac{\gamma}{m}\right) \left(\sum_{j \in \mathcal{N}} g_{i,j}(s)x_j\right) x_i,$$

where we use the distributional assumption on $e_i$.\textsuperscript{10} Actual passage of the bill sponsored by $i$ is represented by the indicator function $I_{[Y_i > m]}$. We interpret $x_i$ as the observable legislative effort by $i$, instrumental to the approval of $i$’s bill, and we postulate that voters prefer politicians exerting higher legislative effort to lower effort. We also allow for voters to care about whether in fact the bill passes conditional on effort. That is, we allow for the political principals (the voters) to reward their agent $i$ for effort $x_i$, networking $s_i$, and ultimately luck $e_i$.

To get reelected, the politician must have an approval rate in his/her electoral district that is sufficiently large. Similarly to [18], the electoral approval rate of $i$ is modeled as a variable $V_i$:

$$V_i = \rho V_{i,0} + \zeta I_{[Y_i > m]} + \alpha x_i + \eta_i$$

where $\eta_i$ is assumed to be a mean zero electoral shock, uniformly distributed on $[-0.5, 0.5]$, and where $V_{i,0} \geq 0$ stands for the baseline approval rate before the start of the term (i.e. before period 1 in the model). Hence, this set-up allows for approval rates to be persistent, but also to react when a politician is capable of getting a bill approved $I_{[Y_i > m]}$ (with a gain $\zeta$) or when $i$ exerts high legislative effort $x_i$, $\rho > 0$ measures persistence in approval rates, which may be due to the politician’s characteristics (such as incumbency advantage, committee membership, majority party affiliation). The parameter $\zeta$, which could be equal to zero empirically, governs the relative importance of a bill actually passing vis-à-vis legislative effort. The direct effect of $x_i$ is captured by $\alpha_i$, which is $i$-specific and may depend on party affiliation, majority status, congressional delegation, etc. Finally, notice that, while there is no direct value to the voters of the politician having more socializing, the value of $s_i$ matters

\textsuperscript{9}Naturally, $m$ can be function of a simple majority requirement or even supermajority restrictions.

\textsuperscript{10}More generally, one can take $Y_i$ to represent the average approval rate of $i$’s multiple bills. In this case, each $b$ is a separate bill by a politician $i$. The conditions for our model are unchanged, as long as bills are not strategically introduced (i.e. specifically shocks $\varepsilon_b$ are still i.i.d. within $i$).
implicitly, being instrumental in getting legislation approved.

In period 2, i is reelected if his/her electoral approval level, \( V_i \) is larger than an electoral threshold \( w < 1 \). So, the probability of being reelected is given by:

\[
Pr(\rho V_{i,0} + \zeta I_{(Y_i > m)} + \alpha_i x_i + \eta_i > w) = \min\{(1 + 0.5 - w) + \rho V_{i,0} + \zeta I_{(Y_i > m)} + \alpha_i x_i, 1\},
\]

where we have used the distributional assumption on \( \eta \). The above expression is non negative, since its terms are non negative. If the first expression in the brackets is larger than 1, then the probability of reelection is 1. We proceed with the empirically relevant case of when reelection is uncertain (i.e. a politician does not know for sure whether (s)he will lose or win).

Note that, in period 1 when making his effort decisions, \( i \) does not know the value of \( \epsilon_i \). So taking the expectation over \( \epsilon \) of the above implies an expected probability of reelection, when choosing \((s_i, x_i)\), given by:

\[
Pr(\text{reelected}) = \mathbb{E}_\epsilon Pr(\rho V_{i,0} + \zeta I_{(Y_i > m)} + \alpha_i x_i + \eta_i > w) \equiv \min\{(1 + 0.5 - w) + \rho V_{i,0} + \zeta \gamma \sum_{j \in N_{g_i}} g_{i,j}(s) x_i x_j + \alpha_i x_i, 1\},
\]

where \( \mathbb{E}_\epsilon I_{(Y_i > m)} = Pr(Y_i > m) \), as given by (2.9).

Replacing (2.11) into the utility function (2.7) yields:

\[
u_i(x_i, x_{-i}) = (1.5 - w) + \rho V_{i,0} + \frac{\zeta}{m} \sum_{j \in N_{g_i}} g_{i,j}(s) x_i x_j + \alpha_i x_i - \frac{c}{2} x_i^2 - \frac{1}{2} s_i^2
\]

Since the terms \((1.5 - w)\) and \( \rho V_{i,0} \) do not affect the maximization problem, an equivalent utility function is:

\[
\tilde{u}_i(x_i, x_{-i}, s_i, s_{-i}) = \alpha_i x_i + \phi \sum_{j \neq i} s_j s_{m_{ij}}(s) x_i x_j - \frac{c}{2} x_i^2 - \frac{1}{2} s_i^2
\]

\(^{11}\)When the probability of reelection is either 0 or 1, we have that \( s_i = x_i = 0 \). This can be seen from equation (2.7). If a politician \( i \) cannot influence his reelection prospects, he will not undertake costly effort. This is contrary to what is observed in the data, as described in Section 2.3, as well as theories of legislative behavior (e.g. [122]).
where $\phi = \zeta \frac{L}{m}$. This is the specification given in (2.6).

### 2.2.4 Solving For Equilibrium

We examine the pure strategy Nash equilibria of the game in which all politicians simultaneously choose $s_i$ and $x_i$.

The first order conditions with respect to $s_i$ and $x_i$ that characterize the best response of politician $i$ imply that interior equilibrium levels of $(s_i^*, x_i^*)$ must satisfy:

$$s_i^* = \phi \sum_{j \neq i} s_j m_{ij}(s^*) x_j^*$$

and

$$cx_i^* = \alpha_i + \phi \sum_{j \neq i} s_i^* s_j m_{ij}(s^*) x_j^*.$$  

(2.15)

We rewrite (2.14) as

$$\frac{s_i^*}{x_i^*} = \phi \sum_{j \neq i} s_j m_{ij}(s^*) x_j^*,$$  

(2.16)

To fully characterize equilibria, we work with the same approximation as in [36]. In particular, we work “at the limit”, when the number of politicians grows. In particular, we solve for equilibrium under the assumption that $\sum_{j \neq i} s_j m_{ij}(s^*) x_j^*$ is the same for all $i$ of the same party.

This implies that $\frac{s_i^*}{x_i^*}$ is the same for all agents within a party. Using (2.16) in (2.15) yields:

$$cx_i^* = \alpha_i + s_i^* \phi \sum_{j \neq i} s_j m_{ij}(s^*) x_j^*$$

$$= \alpha_i + \frac{s_i^*}{x_i^*}$$

Dividing through by $x_i^*$ implies that

$$c = \frac{\alpha_i}{x_i^*} + \frac{s_i^*}{x_i^*}.$$  

(2.17)

Since $\frac{s_i^*}{x_i^*}$ is the same for all agents within a party, (2.17) implies that $\frac{\phi}{x_i^*}$ is the same for all

---

12 Note that second derivatives are everywhere negative.

13 Alternatively, this could be justified via a continuum of politicians of each type, or by examining an epsilon equilibrium with a large $n$. 

12
agents within a party. This further implies that:

\[ x_i^* = \alpha_i X_{P(i)}, \]  

(2.18)

for some \( X_{P(i)} \). In addition, the fact that \( s_i^* \) is the same for all agents within a party, implies that

\[ s_i^* = \alpha_i S_{P(i)}, \]  

(2.19)

in equilibrium for some \( S_{P(i)} \).

To get explicit expressions for our empirical analysis of Congress, we now specialize the analysis to the case of two parties.

For each party \( j = 1, 2 \) define

\[
A_j = \sum_{i \in P_j} \alpha_i, \\
B_j = \sum_{i \in P_j} \alpha_i^2.
\]

**Proposition 2.2.1.** The (interior) Nash equilibria of the limit game of this model are positive solutions to the system given by:

\[
x_i^* = \alpha_i X_{P(i)}, \text{ and} \\
s_i^* = \alpha_i S_{P(i)},
\]

(2.20)

where

\[
\frac{S_1}{X_1} = \phi \left( \frac{p_1 B_1 X_1}{A_1} + \frac{(1 - p_1)^2 B_1 S_1 X_1 + (1 - p_1)(1 - p_2) B_2 S_2 X_2}{(1 - p_1) A_1 S_1 + (1 - p_2) A_2 S_2} \right), \\
\frac{S_2}{X_2} = \phi \left( \frac{p_2 B_2 X_2}{A_2} + \frac{(1 - p_2)^2 B_2 S_2 X_2 + (1 - p_1)(1 - p_2) B_1 S_1 X_1}{(1 - p_1) A_1 S_1 + (1 - p_2) A_2 S_2} \right),
\]

(2.22)

(2.23)

\[
cX_1^2 = X_1 + S_1^2, \quad cX_2^2 = X_2 + S_2^2.
\]

(2.24)

All proofs appear in Appendix.

If \( p_1 = 1 \) or \( p_2 = 1 \), then things reduce to the case of two separate parties with no interaction across them. That is, they are two copies of the model in [36]. Similarly, if \( p_1 = p_2 = 0 \) then there is no impact of party affiliation, and again the model simplifies to that of [36]. The novel case is when at least one partisanship level is positive, yet both levels
are below 1. This biases the interaction of at least one party, leaving room for interaction across parties. In this case there will be both social mixing across different parties and partisanship in socializing.

Generally, there are multiple equilibria. For instance, there is always an (unstable) equilibrium in which \( s_i = 0 \) for all \( i \). In that case, since no other politician provides effort, a given politician’s efforts results in no connections and so the best response is also to provide no effort.

In addition, a sufficient condition for existence of an interior equilibrium is as follows.

**Proposition 2.2.2.** A sufficient condition for the existence of an interior equilibrium is

\[
\frac{2c^{3/2}}{3\sqrt{3}} \geq \phi \max \left[ \frac{B_1}{A_1}, \frac{B_2}{A_2} \right].
\]  

(2.25)

In this setting with two parties and nontrivial partisanship, there will generally be either two or four interior equilibria (except at a degenerate set of values where the system switches from two to four additional equilibria).\(^{14}\)

### 2.2.5 Pareto Efficient Efforts

Before proceeding with the empirical analysis, we comment on the Pareto inefficiency of the equilibrium outcomes of the model. This is relevant for a welfare analysis that checks whether there is over-provision or under-provision of social and legislative effort in the strategic setting.

Generally, the fact that there are positive externalities in efforts – in particular in legislative efforts – implies that there is under-provision of effort. In particular, the Pareto optimal social and legislative effort levels are unbounded: any finite level of efforts are Pareto dominated by some higher levels.\(^{15}\) Hence, all equilibria are characterized by an “under-provision” of efforts.

To see this, we first note that the interaction term in equation (2.13) multiplies three

---

\(^{14}\)These equilibria correspond to when both parties exert high levels or low levels of social efforts, and then for some parameters there are also two additional equilibria in which one party does medium-high and the other does medium-low socializing.

\(^{15}\)The Pareto analysis in [36] only applies if actions are bounded at some small enough finite level. The second derivatives in their proof flip signs if actions are large enough. Thus, there is a local maximum of a weighted sum of utilities that is interior (which is the one identified in their analysis of first-best actions), but the global maximum is actually unbounded. With a strict enough bound on efforts, there would exist an interior maximizer. Effectively, once the efforts are large enough, then the interaction effects dominate the costs. One needs to constrain efforts to be below that level in order to get an Pareto optimal effort solutions. Note however that even with bounds, the equilibrium efforts tend to be inefficient, given the externalities.
variables together: $s_i$, $x_i$ and $x_j$. This has a cubic function property on efforts: doubling all efforts produces an eight times higher interaction effect. Meanwhile, the costs on the social effort ($s_i$) and legislative effort ($x_i$) are quadratic. Hence, doubling those only quadruples costs. It is then direct to check that the gains from increasing effort grow faster than their costs, which implies the following result.

**Proposition 2.2.3.** Every finite profile of efforts is Pareto dominated by some larger level of efforts.

This implies that, although higher payoffs are possible, the selfish attention to individual costs limit the amount of effort that is produced in any equilibrium.\(^{16}\)

If we were to cap effort levels at some high level, then there would exist Pareto optimal efforts bound by the caps.

The message here is that generally, given the complementarities and positive externalities, there is underprovision of effort. A political party, or a government, could help overcome some of the inefficiencies, for instance, by subsidizing meetings and interactions.

### 2.2.6 Preliminaries to Estimation

Generally, effort levels $s_i^*, x_i^*$ are not measured exactly and are observed with noise. For instance, the bill cosponsorships often used as the basis for the construction of political networks are end products that miss other forms of socializing (e.g. close-doors meetings, fund-raisers, and so on). Similarly, although we can partially observe legislative effort through standard proxies (e.g. times the Congress member was present on the floor for speeches, presence in roll call voting, or number of bills written\(^{17}\)), these are imperfect proxies for the legislative efforts that politicians exert. Thus, we account for measurement error in our analysis.

Let $\mathcal{N}_\tau$ denote the politicians comprising Congress $\tau$ and note that this is a set which varies across different $\tau$.\(^{18}\) Introducing classical measurement error, for politician $i$ in Congress $\tau$, we observe:

\(^{16}\)It should be noted, that with high enough effort levels, then the best responses increase without bound in response to increases in others’ efforts. There is no equilibrium, because of the unbounded feedback. Again, if one imposed a cap, then there would be an equilibrium at those capped levels in the model.

\(^{17}\)Both highlighted as important for legislative success in [10].

\(^{18}\)The data is observed for multiple Congresses and we provide identification results for parameters specific to each Congress. This means we allow our parameters to differ across different Congresses and we can construct time-series estimates of the parameters.
\[
\tilde{s}_{i,\tau} = s^*_i e^{-\lambda_i \tau}, \quad (2.26)
\]
\[
\tilde{x}_{i,\tau} = x^*_i e^{-\nu_i \tau}. \quad (2.27)
\]

\(s^*_i\) denotes what is chosen, but it is hit with independent noise and \(\tilde{s}_i\) is observed, and similarly for \(\tilde{x}_i\). The measurement error, conditional on this observation (and on the data we have), is mean zero, and independent of all the other measurement errors across individuals and time. We do not need to impose that the measurement errors in both types of effort have the same distribution.

From Proposition 2.2.1, \((S_1, S_2, X_1, X_2)\) are completely determined by the parameters that govern the system. Then, all individual choices are functions of parameters and of the set of types \(\{\alpha_j\}_{j \in \mathcal{N}_\tau}\).

Let:
\[
\alpha_{i,\tau} = e^{z_{i,\tau}^\prime \beta_{P(i)}}, \quad (2.28)
\]

where \(z_{i,\tau}\) indicates a vector of individual observables\(^1\) (e.g. ideology, tenure, committee membership), and \(\beta_{P(i)}\) are party-specific parameters that will be estimated\(^2\).

The information we employ in the analysis is the following. Let \(\{y_{i,\tau} = I(Y_{i,\tau} > m)\}\) indicate whether each bill was approved or not, where \(i \in \mathcal{N}_\tau\) and \(\tau\) is a given Congressional cycle. \(\{\tilde{s}_{i,\tau}\}\) indicates the (log of hundreds of) cosponsorship decisions per politician \(i \in \mathcal{N}_\tau\). This is our proxy for the equilibrium social effort \(\{s^*_{i,\tau}\}\). The use of logs and rescaling allows us to keep this effort proxy in the same scale as our proxy variable for legislative effort. \(\{\tilde{x}_{i,\tau}\}\) indicates a vector of observable proxies for legislative effort \(\{x^*_{i,\tau}\}\).

As discussed in more detail in the following section, this is constructed using data on floor speeches (word counts per politician during a term) and roll call presence/votes. We employ a procedure (Non-Negative Matrix Factorization, see [166]), to reduce the dimensionality of this set of proxies to a single dimension\(^3\).

---

\(^1\)Unobservables are already be present when we introduce measurement errors. Note that if we had \(\alpha_i = e^{z_{i,\tau}^\prime \beta + \eta_i}\), we could rewrite equation (2.26), using the equilibrium results, simply as:
\[
\tilde{s}_i e^{\lambda_i - \eta_i} = e^{z_{i,\tau}^\prime \beta} S_{P(i)}, \quad (2.29)
\]

and a redefinition of the measurement error to \(\lambda_i - \eta_i\) (still mean-zero and i.i.d.) would suffice in returning to the model presented in the main text, as long as \(A_1, A_2, B_1, B_2\) were not functions of \(\eta_i\).

\(^2\)Identification of the model does not rely on the parametrization of \(\alpha\), as we prove in Appendix. However, this is useful for estimation purposes. A nonparametric \(\alpha\) would require us to estimate a parameter \(\alpha_i\) for each politician in each Congress, when we only observe one set of \((\tilde{s}_i, \tilde{x}_i)\) per period.

\(^3\)A complete data description section follows below.

Sponsorship of bills is already included, as we use the separate bills independently. Further details on this,
As we perform our analysis within a Congress, we suppress the notation $\tau$. We assume that a single pure strategy Nash equilibrium, as defined in Proposition 2.2.1, is played in each Congress. We do not impose, however, that the same equilibrium is played across different Congresses, rather we characterize the equilibrium played empirically in Section 2.6.

Given $\{y_i, \tilde{s}_i, \tilde{x}_i, z_i\}_{i \in N}$, we estimate the parameters $(c, \phi, \zeta, \gamma, p_1, p_2, \beta_1, \beta_2)$. For identification, we set $m = 1$, so that the random variable $\varepsilon_i$ is scaled in terms of the institutional threshold. The basis for identification is Proposition 2.2.1 and the systems of equations that it provides. For identification of the parameters of our model it is not necessary to identify the full set of equilibria, but instead just to use the implications that we are observing some (interior) equilibrium. More precisely, we show that, given the observed data, one can uniquely pin down the equilibrium that is played: although there are multiple interior pure strategy Nash equilibria, conditional on observing $\{\tilde{s}_i, \tilde{x}_i\}_{i \in N}$, there is only one set of values that is most consistent with it. Formal identification of our model, given the information available to the econometrician, is demonstrated in Appendix.

2.3 Data

We use the cosponsorship data from [74], compiled from the Library of Congress, covering the 105th to the 110th United States Congress (from 1997 to 2009). This data contains cosponsorship decisions by politician, and within that data, who sponsors and who cosponsors each bill. It also contains information on whether the each bill was approved in Congress or not (we focus on passage in the House of Representatives). Figure 2.1 shows that measures of inter-connectedness of Congress, for example the total number of cosponsorship links in legislative acts across members of the House ([74]), have been steadily increasing. Figure 2.2 then breaks down how cosponsorships vary within and across parties.

Per Congressional cycle, we compute the log of how many hundred bills each politician cosponsors, which is the variable Cosponsorships. This function of cosponsorships acts as an empirical proxy for the social effort $\{s^*_i, \tau\}_{i \in N}$. We note here that cosponsorship differs from bill sponsorship. Sponsoring a bill refers to the introduction of a bill for consideration (and can be done by multiple legislators drafting the bill, the “sponsors”). Instead, cosponsorships refer to the decision of adding one’s name as a supporter of the bill (becoming a “cosponsor” of the bill). This decision per se does not involve any writing of legislation. Cosponsorships are prevalent in Congress, as the data and procedure to lead to the effort proxy are in the next Section.
The figure shows the evolution of the total number of (unique) cosponsorships during a congressional cycle (i.e. anytime a politician has cosponsored another in a directed way) over time.

The figure shows the evolution of the total number of (unique) cosponsorships within and across parties during a congressional cycle (i.e. anytime a politician has cosponsored another in a directed way) over time.

can be seen in Table 2.1, and the presence of cosponsorship across party lines is still quite common, notwithstanding the trends in polarization discussed in [69], as evident from the time series in Figure 2.2.

The individual bill success outcome (i.e. if the bill passes or not) maps into \(y_{i,\tau} \in N\). We then use the sponsorship information to link the outcome of the bill to the network characteristics and individual decisions.
To compute our proxies for legislative effort, \(\{x^i_{r,\tau}\}_{i \in N, \tau}\) we first collect data on Roll Call voting and floor speeches in Congress. Data for Roll Call voting comes from VoteView. We compute an index, for each politician and for each term in Congress, as the times the Congress member voted as a proportion of total Roll Call votes. This measure, which we call Roll Call Effort, is defined as \(1 - \text{(number of times } i \text{ was “Not Voting”/ total Number of Roll Call votes in a Congress)}\).

Following [10], we also use data on floor speeches as a measure of individual legislative effort. To do so, we compile the amount of words that each Congress member used in his/her floor speeches across the duration of one term (we call this variable Words). Our Floor Speeches variable is constructed as \(\log(1 + \text{Words}_{i,\tau}/300)\). We log and rescale this variable to a scale comparable to other legislative activities. That is, we divide the number of words by 300, which is the congressional limit for each short speech – House Rules explicitly limit one minute speeches to 300 words.\(^{23}\) Data on floor speeches comes from [78], available on ICPSR.\(^{24}\)

That these measures of social interaction and legislative activity may be germane to one another is evident from the significant and positive raw correlation of link formation and proxies of legislative activity and effort, for instance floor speeches in Figure 2.3 and Roll Call Effort in Figure 2.4. This complementarity between effort choices is fully consistent with our theoretical setup.

We proceed to construct \(\{\tilde{x}^i_{r,\tau}\}_{i \in N, \tau}\), by using both Roll Call Effort and Floor Speeches. An appropriate combination of these variables can be obtained through dimensionality reduction methods. Since effort should be non-negative, we employ a procedure that guarantees positive values (i.e. we do not use methodologies like principal components analysis that involve a centering of data and negative values).\(^{25}\) We employ Non-Negative Matrix Factorization (NMF), a dimensionality reduction procedure which imposes constraints so

\(^{23}\)One minute speeches are usually conducted at the beginning of the legislative day. They allow politicians to address the chamber on a topic of their choice, in an unrestricted way. During these short speeches, congressmembers often share information about bills and amendments for the day, pay tributes or share their views on policy. As described in [156], this is a very valuable tool for politicians, including junior ones. It allows congressmembers to be seen and heard easily, as these speeches are often broadcast due to their short length during electoral campaigns.

\(^{24}\)As there are changes in the composition of Congress within a term, for instance due to death or resignation among other reasons, we have some observations whose cosponsorship numbers and word counts do not correspond to a full term. To mend this, we scale up values proportionally to the recorded behavior while in Congress. In other words, if a politician leaves halfway through his term, we double the values of these observations.

\(^{25}\)Our qualitative results still hold if we use either of these variables individually. However, the magnitudes of the estimates change due to the different scales of Roll Call Effort (between 0 and 1) and the floor speech data (in hundreds of words).
Figure 2.3: Correlation between the raw data of log(1+Words) in Floor Speeches and Cosponsorship decisions.

The figure shows the positive correlation between proxies for socializing (log number of cosponsorships) and legislative effort (log number of words in floor speeches). The graph presents the variables in raw form, without rescaling or removal of members with low cosponsorship. The raw correlation is 0.174. In red, we present a LOWESS (locally weighted scatterplot smoothing) fit, with bandwidth (span) equal to 0.9, fitting the relationship between the variables. We do remove, as described in the Data section, observations that have total words equal to zero, which are mostly due to death/resignations in that term.

that the resulting elements are all non-negative. [168] provide a discussion of this methodology. NNMF works by factorizing a matrix, call it $A$, into two positive matrices $W,H$, under a quadratic loss function. The product $WH$ is an approximation to $A$ of smaller dimension, as there are less columns in $W$ than rows in $A$. We then use the main factor in $W$ as our proxy.

We also use observable characteristics, namely ideology (measured by DWNominate from VoteView), tenure (how many terms a politician has served in Congress, with data coming from the Library of Congress), and committee memberships.

Data on committee memberships comes from the work of [162]. To quantify the value of the committees a politician is in, we use the Grosewart measure ([90]). [90] and [161] estimate a cardinal value of how much an assignment to a given committee is valuable to politicians. Such estimates are based upon data on how often politicians accept transfers from one committee to another. The more desirable committees are those that politicians accept to be transferred to often, but rarely accept to be transferred away from. The Grosewart
The figure shows the correlation between proxies for socializing (Cosponsorships) and legislative effort (times the politician is Voting in Roll Call). The graph presents the variables in raw form, without rescaling or removal of members with low cosponsorship. We present a LOWESS (locally weighted scatterplot smoothing) fit, with bandwidth (span) equal to 0.9, fitting the relationship between the variables. This is shown to illustrate that a low correlation across both variables on the full support is driven by the observations with Roll Call Effort close to 1. For example, a positive correlation of 0.135 holds for the data with Roll Call Effort less than 0.95. For our estimation, this will be the useful variation in the Roll Call Effort variable that approximates legislative effort.

Summary statistics for all our variables of interest can be found for reference in Table 2.1.

We restrict the data to Congresses 105th-110th for multiple reasons. First, the data we employ to compute effort from floor speeches is only available from the 104th Congress onwards. Second, the 104th Congress (corresponding to the Republican Revolution) provides a structural break in the analysis of Congressional behavior. With multiple changes

\[ \text{measure sums up the values of the committees in which a politician is present. We use the estimates given in [16] for our study, since they are the updated values for the period we study.}^{26} \]

\[ \text{Below, we also consider an alternative measure for committee memberships. There, we construct dummy variables for whether a politician has been assigned to a given committee during that congressional term. We then focus on the main committees for parsimony: Appropriations, Energy and Commerce, Oversight and Government Reform, Rules, Transportation and Infrastructure, and Ways and Means. We also include a variable Leadership of whether the politician was the Speaker, the Majority or Minority Leader, or the Majority or Minority Whip.} \]
Table 2.1: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>105</th>
<th>106</th>
<th>107</th>
<th>108</th>
<th>109</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cosponsorships</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>185.74</td>
<td>234.57</td>
<td>229.79</td>
<td>226.75</td>
<td>230.74</td>
<td>269.65</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>85.79</td>
<td>102.91</td>
<td>127.03</td>
<td>124.08</td>
<td>119.48</td>
<td>135.90</td>
</tr>
<tr>
<td><strong>Floor Speeches (Words)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>32938.63</td>
<td>36282.23</td>
<td>27906.61</td>
<td>33490.47</td>
<td>33985.21</td>
<td>37416.96</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>38503.19</td>
<td>39234.14</td>
<td>34421.74</td>
<td>42334.30</td>
<td>45922.73</td>
<td>51212.574</td>
</tr>
<tr>
<td><strong>Roll Call Effort</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.9620</td>
<td>0.9524</td>
<td>0.9556</td>
<td>0.9505</td>
<td>0.9605</td>
<td>0.9551</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.0514</td>
<td>0.0574</td>
<td>0.0579</td>
<td>0.0665</td>
<td>0.0380</td>
<td>0.0497</td>
</tr>
<tr>
<td><strong>Ideology (DWNominate)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.0674</td>
<td>0.0695</td>
<td>0.0865</td>
<td>0.1116</td>
<td>0.1276</td>
<td>0.0784</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.4428</td>
<td>0.4549</td>
<td>0.4682</td>
<td>0.4823</td>
<td>0.4966</td>
<td>0.5031</td>
</tr>
<tr>
<td><strong>Tenure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.8439</td>
<td>5.1839</td>
<td>5.4498</td>
<td>5.6073</td>
<td>6.0479</td>
<td>6.0584</td>
</tr>
<tr>
<td><strong>Grosewart</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.2725</td>
<td>0.2797</td>
<td>0.2896</td>
<td>0.2352</td>
<td>0.3046</td>
<td>0.3180</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.0815</td>
<td>1.1207</td>
<td>1.1224</td>
<td>1.1545</td>
<td>1.1591</td>
<td>1.1654</td>
</tr>
<tr>
<td><strong>Approval of House Bills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.1087</td>
<td>0.1246</td>
<td>0.0981</td>
<td>0.1138</td>
<td>0.0957</td>
<td>0.1285</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.3758</td>
<td>0.3782</td>
<td>0.3092</td>
<td>0.3439</td>
<td>0.3690</td>
<td>0.3687</td>
</tr>
<tr>
<td><strong>Number of Politicians N</strong></td>
<td>442</td>
<td>435</td>
<td>440</td>
<td>439</td>
<td>438</td>
<td>445</td>
</tr>
<tr>
<td><strong>Number of Bills</strong></td>
<td>4874</td>
<td>5681</td>
<td>5767</td>
<td>5431</td>
<td>6436</td>
<td>7340</td>
</tr>
</tbody>
</table>

The table presents summary statistics for the variables used in the structural estimation, across Congresses. Roll Call Effort is defined as the proportion of Roll Call votes that the politician does not appear as “Not Voting”. Number of words said in floor speeches aggregates the number of words said by a politician across all his speeches in a term. Cosponsorships and number of words are scaled to full term length (i.e. if a politician leaves mid-office and is replaced mid-office; then both him and the replacement have those variables multiplied by 2.). For estimation, we remove the observations (bills and politicians) we do not have or cannot match to identifying numbers, and those with less than 3 Cosponsorships (see the Data section). These are mostly Congressmen who substitute others mid-term. Data used for bills is House bills (H.R.).
to Congressional composition and structure during the 104th, it becomes hard to compare
the costs and socializing of this specific Congress to others, preceding or following, without
having to further delve into the exceptionality of this particular congressional cycle, which
is not the aim of this work.27

Finally, we perform an additional trimming of the data across all Congresses. We drop
7 (of the 2681) observations in which politicians that have cosponsorship figures less than
3 bills over a full term. From our identification equations, we must use Congress members
that have at least one cosponsored bill. For those with less than 3 cosponsorships (and
given that most politicians cosponsor in the hundreds), scaling is also inappropriate. We
also remove a set of 19 observations, that have the number of words in Floor Speeches set
to 0 in the data of [78]. These observations relate almost exclusively to a politician who
either resigned or died during that term.28

2.4 Estimation

2.4.1 Moment Equations

Let \( \tilde{z}_i = [1, I_i \in P_2, \tilde{z}_i I_i \in P_2] \), where \( I_i \in P_2 \) denotes a dummy variable of whether politician \( i \) is in
Party 2. Further define: \( \beta_s = [\log(S_1), \log(S_2) - \log(S_1), \beta_1, \beta_2 - \beta_1] \), \( \beta_x = [\log(X_1), \log(X_2) - \\
\log(X_1), \beta_1, \beta_2 - \beta_1] \).

The moment conditions necessary to identify and estimate the model’s parameters are29

---

27 In addition, without ad-hoc modifications to the estimating model specifically designed to accommodate
the idiosyncrasies of the 104th Congress, this lack of stability would also likely undermine any effort of struc-
tural estimation.

28 Such as Representatives Jo Ann Davis in the 110th Congress, Sony Bono in the 105th, or resignations
as Representative Bobby Jindal in the 110th. Since the data is zero, the rescaling above does not prove to be
adequate, so we drop these observations.

29 See Appendix for a full derivation of Identification and a description of the moment conditions. We also
prove in Appendix that identification holds in a nonparametric version of the model, where we do impose the
functional form for \( \alpha \).
\[ \mathbb{E}\tilde{z}_i (\log(\tilde{s}_i) - \tilde{z}_i \beta_s) = 0 \] 
\[ \mathbb{E}\tilde{z}_i (\log(\tilde{x}_i) - \tilde{z}_i \beta_x) = 0 \] 
\[ \mathbb{E} \left( 2(\log(\tilde{s}_i) - \log(\tilde{x}_i)) - \log \left( c - \frac{1}{X_1} \right) - \left( \log \left( c - \frac{1}{X_2} \right) - \log \left( c - \frac{1}{X_1} \right) \right) I_{i \in P_2} \right) = 0 \] 
\[ \mathbb{E}\log P(y_i = 1) - \log \left( \frac{1}{\xi} \right) - \log(\tilde{s}_i^2) = 0. \] 
\[ S_1 = \phi X_1 (B_1 S_{1,1} m_{11} + B_2 S_{2,2} m_{12}) \] 
\[ S_2 = \phi X_2 (B_2 S_{2,2} m_{22} + B_1 S_{1,1} m_{12}). \] 

These moment conditions are based on rewriting the equations of Proposition 2.2.1 using our parameterization for \( \alpha \) and measurement errors, given in equations (2.28) and (2.26). They allow us to identify \((c, \xi, S_1, S_2, X_1, X_2, \beta_1, \beta_2)\) and set identify \( \phi \). Specifically, lack of point identification of \( \phi \) is the result of lack of point identification of \( p_1 \) and \( p_2 \). The parameters \( p_1 \) and \( p_2 \) enter nonlinearly in equations (2.34) and (2.35) through \( m_{ij}(s) \), identifying a ridge of \((p_1, p_2)\) pairs satisfying (2.34) and (2.35).

In Appendix, we demonstrate how to obtain point identification of \( \phi \) when we impose additional restrictions on the proxy variables \((\tilde{s}_i, \tilde{x}_i)\). The restrictions there are on the second moments of the effort proxies, similarly in spirit to random coefficient models. Such restrictions are justified under the assumption that more partisanship may result in noisier measurement of social interactions and legislative effort. This alternative approach is heavier in terms of assumptions and for this reason we do not adopt it for the derivation of our main results in Section 2.5. Instead, we report estimated values for all parameters (including \( p_1, p_2 \)) under these additional assumptions and specification in Appendix.

For our main empirical exercise, we let Party 1 denote the Democratic Party (with its variables denoted by the subscript \( Dem \)) and Party 2 denote the Republican Party (analogously denoted with a \( Rep \) subscript).

We carry out the estimation process using a two step procedure. In the first step, we compute estimates for the parameters \((c, \xi, S_{Dem}, S_{Rep}, X_{Dem}, X_{Rep}, \beta_{Dem}, \beta_{Rep})\) from the moment equations (2.30) - (2.33) above, via the Generalized Method of Moments.

In the second step, we use the first-step estimates to derive a set estimate for \( \phi \). This is

---

\( ^{30} \)As second moments restrictions allow to obtain point estimates of \( p_1 \) and \( p_2 \), we use such estimates in the last part of Section 2.5 for assessing additional implications of our model and validation.
done by using equations (2.34) and (2.35). We grid all pairs \((p_{Dem}, p_{Rep}) \in [0, 1] \times [0, 1]\), and, employing the estimated \(A_{Dem}, A_{Rep}, S_{Dem}, S_{Rep}\) from the first step, we calculate the values of \(m_{ij}\) for each pair \((p_{Dem}, p_{Rep})\). The set estimate for \(\phi\) are all the values that satisfy equations (2.34) and (2.35) for any pair \((p_{Dem}, p_{Rep})\).

Concerning the information of whether a bill passed or not \(\{y_i, \tau\}_i \in N\), the model is agnostic on how many bills a politician proposes. Because a good fraction of members of Congress sponsor multiple bills, however, we work with \(L > N\) bills in the actual data. This is easily accommodated in the estimation. Recall that \(\varepsilon\) are i.i.d. across time and bills. For each politician \(i\), all \(i\)'s bills have the same associated network \(g_{i,j}\), as it comes from the same politician and his same network and effort choices (as well as those of his network). The different \(\varepsilon\) realizations, however, represent different bill qualities or institutional arrangements within politician, meaning that the same politician may have one bill approved and not another. The dimensionality of the problem can be decreased by simply averaging out each bill’s success by politician. This is made possible by the fact that equation (2.33) holds for all bills, implying that it must hold for all politicians as well. Hence, we use the average pass rate of bills for politician \(i\) as its estimate of the probability of bill approval.

2.4.2 Estimation via GMM

To estimate the model, we replace equations (2.30)-(2.33) by their empirical counterparts and stack them into a vector of the form \(\frac{1}{n} \sum_{i=1}^{n} g(\tilde{s}_i, \tilde{x}_i, y_i, z_i; \theta)\). Since all moments have expectations taken over \(\lambda_i, \nu_i\), which are i.i.d. and mean zero for all politicians, the empirical counterpart replaces the expectation operator by the mean over \(i\). Furthermore, we average over the approval rates for bills for each politician to get the estimated probability

\[ \frac{1}{n} \sum_{i=1}^{n} \tilde{z}_i (\log(\tilde{s}_i) - \tilde{z}_i \beta_s) = 0, \]  
\[ \frac{1}{n} \sum_{i=1}^{n} \tilde{z}_i (\log(\tilde{x}_i) - \tilde{z}_i \beta_x) = 0, \]

or in matrix form:

\[ \frac{1}{n} Z' (\log(\tilde{s}) - Z \beta_s) = 0, \]  
\[ \frac{1}{n} Z' (\log(\tilde{x}) - Z \beta_x) = 0, \]

where \(Z\) stacks up \(z_i\), and \(\log(\tilde{s}), \log(\tilde{x})\) stack up \(\log(\tilde{s}_i), \log(\tilde{x}_i)\) respectively.

31That is, the expectation operator has one observation for each politician, and averages across all politicians. For example, the empirical counterparts to (2.30)-(2.31) are:

\[ \frac{1}{n} \sum_{i=1}^{n} \tilde{z}_i (\log(\tilde{s}_i) - \tilde{z}_i \beta_s) = 0, \]  
\[ \frac{1}{n} \sum_{i=1}^{n} \tilde{z}_i (\log(\tilde{x}_i) - \tilde{z}_i \beta_x) = 0, \]
of approval at the politician level.

We then minimize the quadratic form:

\[
\left( \frac{1}{n} \sum_{i=1}^{n} g(\tilde{s}_i, \tilde{x}_i, y_i, z_i; \theta) \right) \cdot W \left( \frac{1}{n} \sum_{i=1}^{n} g(\tilde{s}_i, \tilde{x}_i, y_i, z_i; \theta) \right), \tag{2.40}
\]

where \( \frac{1}{n} \sum_{i=1}^{n} g(\tilde{s}_i, \tilde{x}_i, y_i, z_i; \theta) \) is given by stacking up the empirical counterparts of equations (2.30)-(2.33), for a total of \( 2k + 2 \) equations (\( k \) being the dimensionality of \( z_i \)). \( W \) is the weighting matrix, which can be taken as the identity matrix (an inefficient choice), or the optimal weighting matrix for the GMM.

Given these first stage estimates, we then estimate the set of feasible values for \( \phi \) as previously described. Further details about the empirical implementation are discussed in Appendix.

2.5 Results

Table 2.2 presents our parameter estimates. Figure 2.5 and Table 2.3 show the distributions of the estimated \( \alpha_i \) (Figure 2.5) over time and by party (Table 2.3). These distributions appear stable across Congresses.

**Figure 2.5: Distribution of Estimated \( \alpha \) over Time**

Splitting the samples by party, we observe important differences in the estimated distributions of Republicans and Democrats. Democrats have a higher average and dispersion, while Republicans have tighter distributions. This implies different social effort patterns.
<table>
<thead>
<tr>
<th>Congress</th>
<th>105</th>
<th>106</th>
<th>107</th>
<th>108</th>
<th>109</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>0.327</td>
<td>0.324</td>
<td>0.359</td>
<td>0.365</td>
<td>0.352</td>
<td>0.336</td>
</tr>
<tr>
<td></td>
<td>(0.0089)</td>
<td>(0.0083)</td>
<td>(0.0102)</td>
<td>(0.0114)</td>
<td>(0.0095)</td>
<td>(0.0085)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>[0.0476, 0.0591]</td>
<td>[0.0487, 0.0660]</td>
<td>[0.0537, 0.0762]</td>
<td>[0.0520, 0.0724]</td>
<td>[0.0526, 0.0709]</td>
<td>[0.0546, 0.0698]</td>
</tr>
<tr>
<td></td>
<td>(0.918)</td>
<td>(1.889)</td>
<td>(1.364)</td>
<td>(1.041)</td>
<td>(1.323)</td>
<td>(1.060)</td>
</tr>
<tr>
<td>$S_{Dem}$</td>
<td>0.876</td>
<td>1.054</td>
<td>1.006</td>
<td>0.970</td>
<td>1.009</td>
<td>1.156</td>
</tr>
<tr>
<td></td>
<td>(0.0329)</td>
<td>(0.0362)</td>
<td>(0.0355)</td>
<td>(0.0342)</td>
<td>(0.0312)</td>
<td>(0.0295)</td>
</tr>
<tr>
<td>$S_{Rep}$</td>
<td>0.850</td>
<td>0.978</td>
<td>0.936</td>
<td>0.847</td>
<td>0.881</td>
<td>0.994</td>
</tr>
<tr>
<td></td>
<td>(0.0380)</td>
<td>(0.0406)</td>
<td>(0.0502)</td>
<td>(0.0422)</td>
<td>(0.0471)</td>
<td>(0.0590)</td>
</tr>
<tr>
<td>$X_{Dem}$</td>
<td>3.423</td>
<td>3.586</td>
<td>3.155</td>
<td>3.132</td>
<td>3.297</td>
<td>3.723</td>
</tr>
<tr>
<td></td>
<td>(0.130)</td>
<td>(0.127)</td>
<td>(0.117)</td>
<td>(0.131)</td>
<td>(0.113)</td>
<td>(0.0938)</td>
</tr>
<tr>
<td>$X_{Rep}$</td>
<td>4.056</td>
<td>4.422</td>
<td>4.168</td>
<td>3.813</td>
<td>3.885</td>
<td>4.093</td>
</tr>
<tr>
<td></td>
<td>(0.180)</td>
<td>(0.186)</td>
<td>(0.220)</td>
<td>(0.193)</td>
<td>(0.209)</td>
<td>(0.262)</td>
</tr>
<tr>
<td>Ideology</td>
<td>-0.572</td>
<td>-0.493</td>
<td>-0.710</td>
<td>-0.766</td>
<td>-0.679</td>
<td>-0.509</td>
</tr>
<tr>
<td></td>
<td>(0.0888)</td>
<td>(0.0825)</td>
<td>(0.0839)</td>
<td>(0.0927)</td>
<td>(0.0818)</td>
<td>(0.0701)</td>
</tr>
<tr>
<td>Tenure</td>
<td>0.00364</td>
<td>0.00586</td>
<td>0.00452</td>
<td>0.00518</td>
<td>0.00320</td>
<td>-0.00101</td>
</tr>
<tr>
<td></td>
<td>(0.00285)</td>
<td>(0.00242)</td>
<td>(0.00275)</td>
<td>(0.00311)</td>
<td>(0.00259)</td>
<td>(0.00222)</td>
</tr>
<tr>
<td>Grosewart</td>
<td>-0.0145</td>
<td>-0.0180</td>
<td>-0.0300</td>
<td>-0.0151</td>
<td>-0.0111</td>
<td>-0.00404</td>
</tr>
<tr>
<td></td>
<td>(0.00952)</td>
<td>(0.00973)</td>
<td>(0.00944)</td>
<td>(0.0114)</td>
<td>(0.00960)</td>
<td>(0.00843)</td>
</tr>
<tr>
<td>Ideology $\times$ Rep</td>
<td>0.0754</td>
<td>-0.0189</td>
<td>-0.0706</td>
<td>0.0929</td>
<td>0.127</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>(0.0764)</td>
<td>(0.0699)</td>
<td>(0.0876)</td>
<td>(0.0749)</td>
<td>(0.0783)</td>
<td>(0.0823)</td>
</tr>
<tr>
<td>Tenure $\times$ Rep</td>
<td>0.0110</td>
<td>0.00728</td>
<td>0.00489</td>
<td>0.00776</td>
<td>0.00196</td>
<td>-0.000345</td>
</tr>
<tr>
<td></td>
<td>(0.00329)</td>
<td>(0.00271)</td>
<td>(0.00347)</td>
<td>(0.00302)</td>
<td>(0.00318)</td>
<td>(0.0037)</td>
</tr>
<tr>
<td>Grosewart $\times$ Rep</td>
<td>-0.0177</td>
<td>-0.00613</td>
<td>-0.0205</td>
<td>-0.0239</td>
<td>-0.0326</td>
<td>-0.0341</td>
</tr>
<tr>
<td></td>
<td>(0.0104)</td>
<td>(0.0100)</td>
<td>(0.0115)</td>
<td>(0.0112)</td>
<td>(0.0104)</td>
<td>(0.0124)</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses. The table presents the GMM estimates using the Optimal Weighting Matrix for the parameters of interest, as described in the Estimation section. Standard Errors are computed from estimates of the variance for a GMM estimator with the Optimal Weighting Matrix. Details are in Appendix. The estimate of $\phi$ is its estimated identified set. $Rep$ represents the dummy variable of whether a politician was in the Republican Party. Hence, a variable $Tenure \times Republican$ represents the additional estimate of the $Tenure$ variable for the Republican Party, as compared to the Democratic one.
### Table 2.3: Differences in the Distributions of $\alpha_i$ Across Parties

<table>
<thead>
<tr>
<th>Congress</th>
<th>105</th>
<th>106</th>
<th>107</th>
<th>108</th>
<th>109</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Democrats:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean $\alpha_i$</td>
<td>1.256</td>
<td>1.234</td>
<td>1.329</td>
<td>1.375</td>
<td>1.323</td>
<td>1.196</td>
</tr>
<tr>
<td>Standard Deviation of $\alpha_i$</td>
<td>0.109</td>
<td>0.0946</td>
<td>0.137</td>
<td>0.149</td>
<td>0.125</td>
<td>0.088</td>
</tr>
<tr>
<td><strong>Republicans:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean $\alpha_i$</td>
<td>1.085</td>
<td>1.027</td>
<td>0.985</td>
<td>1.094</td>
<td>1.081</td>
<td>1.073</td>
</tr>
<tr>
<td>Standard Deviation of $\alpha_i$</td>
<td>0.0408</td>
<td>0.0264</td>
<td>0.0277</td>
<td>0.0390</td>
<td>0.0424</td>
<td>0.0465</td>
</tr>
</tbody>
</table>

We show the mean and the standard deviation of the (estimated) distributions of $\alpha_i$ for each party, highlighting the differences in those distributions. The estimates presented are those from Table 2.2

across parties, as Democrats socialize more and (by our social meeting function) more often with other Democrats.

We see that $\phi$ has an estimated set that excludes 0 and is stable over time. Its magnitude ranges from [0.047, 0.076]. To put this into context, note that the marginal utility of an increase in $x_i$ is

$$\alpha_i + \phi s_i \sum_{j\neq i} x_j s_j m_{ij}(s) - cx_i.$$  

The direct benefit $\alpha_i$ ranges from 1 to 1.4, while the social part $\phi s_i \sum_{j\neq i} x_j s_j m_{ij}(s)$ ranges from about 0.15 to 0.32. So, the social incentive is somewhere between a tenth to a third of the direct incentives.

This provides evidence that the returns to socializing are quantitatively sizeable (significantly away from 0 for all Congresses and of a significant magnitude relative to $c$ which ranges in [0.32, 0.37]) and have not changed much over time. In the context of the model, these returns represent the (expected) gain to the politician of the interaction with other politicians and having a bill approved relative to the direct return to legislative effort. $\phi$ being stable also suggests that the returns of having a bill approved have not significantly changed over the period of time studied.

The relative cost of legislative effort to social effort, $c$, is estimated to be increasing over time. This may be consistent with an increase in the complexity of extant statutes, as for example evident from an average number of pages per statute of 3.6 in 1965-66.
to 18.8 in 2015-16, making interaction between politicians more important in drafting and drumming up support for legislations, or in improvements in the technology of social interaction among members of the House, inside and outside the Capitol. We explore how changes in $c$ affect the choices of social and legislative efforts in our estimated equilibrium in the next section.

The estimates of $\zeta$ are also significant and large in magnitude. This indicates that politicians see positive gains from having bills approved. This larger magnitude is needed for the model to be internally consistent. This is because $\zeta$ also operates as a normalizer that guarantees that the probability of bill approval (in equation (2.33)) is between 0 and 1. Using the estimated values of $\zeta$, we can then calculate the probability of bill approvals for each politician. We show these in Figures 2.6 for Democrats and 2.7 for Republicans.

**Figure 2.6:** Estimated Probability of Approval - Democrats, Congresses 105-110

![Graph showing distribution of approval probabilities for different congresses.]

By comparing Figures 2.6 - 2.7 with the average bill passage rates in the summary statistics (Table 2.1), we can see that the model can generate a good match at the mean approval rate (which we observe), while our structural assumptions allow us to represent the whole distribution of expected probabilities of having a bill approved across different politicians. These indicate some variation over time. Later Congresses (108th and 110th) show a higher predicted approval rate for most politicians. For Congress 110, such effects are also driven by an increase in the average $\alpha_i$.

We can also discuss the statistical significance of different observable characteristics in explaining the individual $\alpha_i$. With our baseline specification that uses Ideology, Tenure,

---

Figure 2.7: Estimated Probability of Approval - Republicans, Congresses 105-110

For $z_i$, we see that ideology is statistically significant (especially in later Congresses). The estimates suggest that those on the left of the ideological spectrum (extremist Democrats, moderate Republicans), have higher returns of exerting legislative effort. Meanwhile, the Grosewart variable, capturing the impacts of committee assignments, appears to be noisy.

We also consider another specification where we replace the Grosewart variable by dummy variables of committee assignments to each of Congress main committees. This is shown in Table 2.4. We can see that the results from our main specification are robust. It is noteworthy in that case, though, that the estimate of being in the Rules Committee is positive and significant. The Rules Committee is the committee in charge of determining the rules that allow each bill to come to the floor, fundamental for the progress of legislation. It seems consistent that politicians in that committee are rewarded for effort in it, even conditional on having the same ideology, party, and tenure.

2.5.1 Fit and Discussion

To conclude the section, consider the following out-of-sample validation of our approach.

Although our analysis employs $i$’s total cosponsorship figures to proxy for social effort $s^*_{i,t}$, the more fine-grained data on pairwise cosponsorship information between $i, j$ politicians, directional in nature, is not used in estimation. In this subsection, we predict the cosponsorship $i, j$ links at the level of each Congress member based on what is predicted by our $g_{l,j}(s)$ function separately for each Congress in our sample. The goal is to show that
Table 2.4: Main Results, Specification 2

<table>
<thead>
<tr>
<th></th>
<th>105</th>
<th>106</th>
<th>107</th>
<th>108</th>
<th>109</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma )</td>
<td>0.316</td>
<td>0.336</td>
<td>0.351</td>
<td>0.360</td>
<td>0.346</td>
<td>0.330</td>
</tr>
<tr>
<td>(0.00907)</td>
<td>(0.00839)</td>
<td>(0.0102)</td>
<td>(0.0107)</td>
<td>(0.00935)</td>
<td>(0.00751)</td>
<td></td>
</tr>
<tr>
<td>( \phi )</td>
<td>[0.0473, 0.0594]</td>
<td>[0.0483, 0.0646]</td>
<td>[0.0536, 0.0761]</td>
<td>[0.0517, 0.0715]</td>
<td>[0.0526, 0.0705]</td>
<td>[0.0550, 0.0699]</td>
</tr>
<tr>
<td>(0.907)</td>
<td>(1.839)</td>
<td>(1.346)</td>
<td>(1.025)</td>
<td>(1.278)</td>
<td>(1.018)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>( \Sigma_{Dem} )</td>
<td>0.870</td>
<td>0.998</td>
<td>1.018</td>
<td>0.875</td>
<td>0.918</td>
<td>1.091</td>
</tr>
<tr>
<td>(0.0442)</td>
<td>(0.0346)</td>
<td>(0.0369)</td>
<td>(0.0356)</td>
<td>(0.0312)</td>
<td>(0.0289)</td>
<td>(0.0289)</td>
</tr>
<tr>
<td>( \Sigma_{Rep} )</td>
<td>4.166</td>
<td>4.491</td>
<td>4.522</td>
<td>3.932</td>
<td>4.034</td>
<td>4.450</td>
</tr>
<tr>
<td>(0.217)</td>
<td>(0.211)</td>
<td>(0.259)</td>
<td>(0.224)</td>
<td>(0.249)</td>
<td>(0.293)</td>
<td>(0.293)</td>
</tr>
<tr>
<td>Ideology</td>
<td>-0.568</td>
<td>-0.355</td>
<td>-0.681</td>
<td>-0.769</td>
<td>-0.635</td>
<td>-0.510</td>
</tr>
<tr>
<td>(0.0852)</td>
<td>(0.0801)</td>
<td>(0.0821)</td>
<td>(0.0806)</td>
<td>(0.0837)</td>
<td>(0.0801)</td>
<td>(0.0719)</td>
</tr>
<tr>
<td>Tenure</td>
<td>0.00349</td>
<td>0.00446</td>
<td>0.00188</td>
<td>0.00389</td>
<td>0.00378</td>
<td>0.000119</td>
</tr>
<tr>
<td>(0.00398)</td>
<td>(0.00219)</td>
<td>(0.00237)</td>
<td>(0.00254)</td>
<td>(0.00331)</td>
<td>(0.00299)</td>
<td>(0.00299)</td>
</tr>
<tr>
<td>Appropriations</td>
<td>-0.0480</td>
<td>-0.0158</td>
<td>-0.0814</td>
<td>-0.0990</td>
<td>-0.0900</td>
<td>-0.0507</td>
</tr>
<tr>
<td>(0.0401)</td>
<td>(0.0330)</td>
<td>(0.0356)</td>
<td>(0.0407)</td>
<td>(0.0363)</td>
<td>(0.0289)</td>
<td>(0.0289)</td>
</tr>
<tr>
<td>Energy and Commerce</td>
<td>0.0266</td>
<td>0.0296</td>
<td>0.0178</td>
<td>0.0227</td>
<td>0.0363</td>
<td>0.0312</td>
</tr>
<tr>
<td>(0.0309)</td>
<td>(0.0307)</td>
<td>(0.0346)</td>
<td>(0.0349)</td>
<td>(0.0311)</td>
<td>(0.0223)</td>
<td>(0.0223)</td>
</tr>
<tr>
<td>Oversight</td>
<td>0.0387</td>
<td>0.0420</td>
<td>0.0603</td>
<td>0.0405</td>
<td>0.0590</td>
<td>0.00186</td>
</tr>
<tr>
<td>(0.0266)</td>
<td>(0.0321)</td>
<td>(0.0333)</td>
<td>(0.0401)</td>
<td>(0.0383)</td>
<td>(0.0223)</td>
<td>(0.0223)</td>
</tr>
<tr>
<td>Rules</td>
<td>0.184</td>
<td>0.185</td>
<td>0.146</td>
<td>0.198</td>
<td>0.198</td>
<td>0.142</td>
</tr>
<tr>
<td>(0.0258)</td>
<td>(0.0279)</td>
<td>(0.0346)</td>
<td>(0.0392)</td>
<td>(0.0360)</td>
<td>(0.0247)</td>
<td>(0.0247)</td>
</tr>
<tr>
<td>Leadership</td>
<td>0.0812</td>
<td>0.0182</td>
<td>0.0097</td>
<td>0.0703</td>
<td>0.0580</td>
<td>0.735</td>
</tr>
<tr>
<td>(0.0579)</td>
<td>(0.0510)</td>
<td>(0.0864)</td>
<td>(0.0807)</td>
<td>(0.0522)</td>
<td>(0.233)</td>
<td>(0.233)</td>
</tr>
<tr>
<td>Transportation</td>
<td>-0.0132</td>
<td>0.0444</td>
<td>0.0216</td>
<td>0.00783</td>
<td>0.000136</td>
<td>0.00706</td>
</tr>
<tr>
<td>(0.0310)</td>
<td>(0.0297)</td>
<td>(0.0254)</td>
<td>(0.0294)</td>
<td>(0.0259)</td>
<td>(0.0267)</td>
<td>(0.0267)</td>
</tr>
<tr>
<td>Ways and Means</td>
<td>-0.0674</td>
<td>-0.0521</td>
<td>-0.0306</td>
<td>-0.0388</td>
<td>-0.0313</td>
<td>-0.0718</td>
</tr>
<tr>
<td>(0.0384)</td>
<td>(0.0313)</td>
<td>(0.0385)</td>
<td>(0.0410)</td>
<td>(0.0421)</td>
<td>(0.0345)</td>
<td>(0.0345)</td>
</tr>
<tr>
<td>Ideology \times Rep</td>
<td>0.0453</td>
<td>-0.0527</td>
<td>-0.145</td>
<td>0.0717</td>
<td>0.0839</td>
<td>0.0401</td>
</tr>
<tr>
<td>(0.0774)</td>
<td>(0.0699)</td>
<td>(0.0847)</td>
<td>(0.0750)</td>
<td>(0.0784)</td>
<td>(0.0682)</td>
<td>(0.0682)</td>
</tr>
<tr>
<td>Tenure \times Rep</td>
<td>0.0699</td>
<td>0.0270</td>
<td>0.0098</td>
<td>0.00668</td>
<td>0.00233</td>
<td>0.000134</td>
</tr>
<tr>
<td>(0.0338)</td>
<td>(0.0266)</td>
<td>(0.0311)</td>
<td>(0.0303)</td>
<td>(0.0299)</td>
<td>(0.0333)</td>
<td>(0.0333)</td>
</tr>
<tr>
<td>Appropriations \times Rep</td>
<td>-0.00958</td>
<td>-0.0412</td>
<td>-0.159</td>
<td>-0.113</td>
<td>-0.124</td>
<td>-0.145</td>
</tr>
<tr>
<td>(0.0253)</td>
<td>(0.0271)</td>
<td>(0.0390)</td>
<td>(0.0399)</td>
<td>(0.0398)</td>
<td>(0.0431)</td>
<td>(0.0431)</td>
</tr>
<tr>
<td>Energy and Commerce \times Rep</td>
<td>-0.0794</td>
<td>-0.0109</td>
<td>-0.0423</td>
<td>-0.05403</td>
<td>-0.0295</td>
<td>-0.0438</td>
</tr>
<tr>
<td>(0.0307)</td>
<td>(0.0266)</td>
<td>(0.0309)</td>
<td>(0.0318)</td>
<td>(0.0337)</td>
<td>(0.0373)</td>
<td>(0.0373)</td>
</tr>
<tr>
<td>Oversight \times Rep</td>
<td>0.0780</td>
<td>0.105</td>
<td>0.0892</td>
<td>0.0615</td>
<td>0.0801</td>
<td>0.0363</td>
</tr>
<tr>
<td>(0.0353)</td>
<td>(0.0325)</td>
<td>(0.0312)</td>
<td>(0.0298)</td>
<td>(0.0351)</td>
<td>(0.0414)</td>
<td>(0.0414)</td>
</tr>
<tr>
<td>Rules \times Rep</td>
<td>0.012</td>
<td>0.0961</td>
<td>0.0702</td>
<td>0.0476</td>
<td>0.142</td>
<td>0.117</td>
</tr>
<tr>
<td>(0.0483)</td>
<td>(0.0275)</td>
<td>(0.0291)</td>
<td>(0.0355)</td>
<td>(0.0415)</td>
<td>(0.0497)</td>
<td>(0.0497)</td>
</tr>
<tr>
<td>Leadership \times Rep</td>
<td>-0.0792</td>
<td>-0.0502</td>
<td>-0.158</td>
<td>-0.203</td>
<td>-0.234</td>
<td>-0.0146</td>
</tr>
<tr>
<td>(0.0668)</td>
<td>(0.0229)</td>
<td>(0.163)</td>
<td>(0.137)</td>
<td>(0.139)</td>
<td>(0.0917)</td>
<td>(0.0917)</td>
</tr>
<tr>
<td>Transportation \times Rep</td>
<td>-0.0466</td>
<td>-0.0387</td>
<td>-0.0592</td>
<td>-0.00666</td>
<td>-0.0230</td>
<td>-0.0457</td>
</tr>
<tr>
<td>(0.0324)</td>
<td>(0.0285)</td>
<td>(0.0310)</td>
<td>(0.0311)</td>
<td>(0.0314)</td>
<td>(0.0414)</td>
<td>(0.0414)</td>
</tr>
<tr>
<td>Ways and Means \times Rep</td>
<td>-0.0434</td>
<td>0.000758</td>
<td>-0.0117</td>
<td>-0.0226</td>
<td>-0.0781</td>
<td>-0.150</td>
</tr>
<tr>
<td>(0.0362)</td>
<td>(0.0412)</td>
<td>(0.0411)</td>
<td>(0.0372)</td>
<td>(0.0469)</td>
<td>(0.0486)</td>
<td>(0.0486)</td>
</tr>
</tbody>
</table>

N 436 433 435 437 433 436

Notes: Standard errors in parentheses. The table presents the results from the GMM estimation under the second specification. That is, we replace the Grosewart measure by dummy variables for the most important committees. The variable Leadership represents a dummy of whether the politician was the Speaker, the Majority or Minority Leader, or the Majority or Minority Whip. \( Rep \) is a dummy variable for belonging to the Republican Party. The \( \phi \) is its estimated identified set. The optimal weighting matrix is used, and standard errors are estimated as discussed in Appendix. All other notes follow those in Table 2.2.
our estimated network model fits such proxies for social ties (common in the literature, see [74]), even without explicitly modeling pairwise socialization decisions.

The correlations between the estimated $g_{i,j}(s)$ and any $i, j$ pairwise cosponsorships are reported in Table 2.5. The Table illustrates the correlations for two possible definitions of links based on $i, j$ cosponsorship in the data: in the top panel cosponsorships are considered directed from $i$ to $j$ and in the second panel cosponsorships are considered a-directional. In the two cases the correlations with the model-implied $g_{i,j}(s)$ are 0.25 and 0.31 respectively and statistically significant. The model appears able to fit disaggregated socializing proxies emerging from individual cosponsorships that are not directly targeted in estimation.

We can draw an additional relevant conclusion from this exercise. Results of Fisher’s $z$-transformation tests also suggest that our model with $p_{Dem} > 0, p_{Rep} > 0$ is better at capturing the relationships from the pairwise cosponsorship data than alternative models with full partisanship (at least one of $p_{Dem} = 1$ or $p_{Rep} = 1$) or without any partisanship ($p_{Dem} = p_{Rep} = 0$). These comparisons are possible since different $g_{i,j}(s)$ can be generated using different values for $p_{Dem}, p_{Rep}$.33

We underscore that, although recent political economy research highlights a hollowing out of the moderate middle ground in congressional voting ([128]; [69]), our model with $p_{Dem}, p_{Rep}$ in a range less of 0.1 (values in this range for $p_{Dem}, p_{Rep}$ are obtained in Appendix) produces a substantially better fit of the cosponsorship data than a model with complete polarization $p_{Dem} = p_{Rep} = 1$, which is statistically dominated.

While the exact point estimates of $p_{Dem}, p_{Rep}$ rely on our assumptions on second moments, we believe that the rejection of $p_{Dem} = p_{Rep} = 1$ has to be considered more general. The raw data in Figure 2.2 itself displays a sufficient degree of cross-party cosponsorship to cast doubt on an hypothesis of “full sorting” among party members in the House. Possibly, reconciling a world of more polarized legislators and the thousands of cosponsorships across party lines reported in Figure 2.2 may come from noting that, as ideology may diverge, engagement across party lines becomes more important for getting legislation to the floor and passed. Our model appears to capture such phenomena.

### 2.6 Assessment and Counterfactuals

Our theory allows for multiplicity of equilibria. In this Section, we first discuss what we can say about the equilibrium being played in each Congress. In the second part of this

---

33For these tests, absent an estimate for $p_1$ and $p_2$ from the baseline model where these parameters are not identified, we employ the estimates generated by a specification imposing second moment restrictions, reported in Appendix. This shows that our model captures part of the empirically observed relationship.
Table 2.5: Model Fit: Correlation of Estimated Network of the Model to the Cosponsorship Networks in the Data

<table>
<thead>
<tr>
<th>Congress</th>
<th>Correlation</th>
<th>Fisher’s z-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data from Directed Cosponsorships:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model: $p_{Dem} &gt; 0, p_{Rep} &gt; 0$</td>
<td>0.2544</td>
<td>-</td>
</tr>
<tr>
<td>Model: $p_{Dem} = 0, p_{Rep} = 0$</td>
<td>0.2516</td>
<td>2.219**</td>
</tr>
<tr>
<td>Model: $p_{Dem} = 1$</td>
<td>0.1840</td>
<td>55.615***</td>
</tr>
<tr>
<td><strong>Data from “Combined” Cosponsorships:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model: $p_{Dem} &gt; 0, p_{Rep} &gt; 0$</td>
<td>0.3125</td>
<td>-</td>
</tr>
<tr>
<td>Model: $p_{Dem} = 0, p_{Rep} = 0$</td>
<td>0.3091</td>
<td>2.825***</td>
</tr>
<tr>
<td>Model: $p_{Dem} = 1$</td>
<td>0.2261</td>
<td>70.207***</td>
</tr>
</tbody>
</table>

We compare the performance of the partisan model (with $p_{Dem} > 0, p_{Rep} > 0$) to the performance of the model without partisanship ($p_{Dem} = p_{Rep} = 0$) and complete partisanship ($p_{Dem} = 1$), in explaining the observed cosponsorships in the data. In the first panel, cosponsorships are measured by the directed number: how many times $i$ cosponsors $j$. In the second panel, “combined cosponsorships” are measured by the number of times $i$ cosponsors $j$ and $j$ cosponsors $i$, creating a symmetric undirected graph. To calculate the statistics, we first generate the links using the theoretical definition $g_{ij}(s) = s_i s_j m_{ij}(s)$ under our estimated parameters. That is done for the 3 cases. We then show the correlations of the model links to the values of the cosponsorships in the data. The estimated values for $p_{Dem} > 0, p_{Rep} > 0$ come from the estimates using second moments of $(\tilde{s}_i, \tilde{x}_i)$, from Table A.1 in Appendix. We present Fisher’s z-transformation statistic, for the test that the correlation of the Model with $p_{Dem} > 0, p_{Rep} > 0$ is equal to the correlation of the alternative model (without partisanship/complete partisanship). *** represents that the null hypothesis of equal correlations can be rejected at 1% significance level, * at 5%. Note that, when estimating the model, we did not use cosponsorships at the $ij$ level. We aggregate all Congresses in the analysis above.
Section, we present counterfactual exercises using the estimated model parameters.

### 2.6.1 Stability of Equilibrium and Other Equilibrium Properties

Let us first discuss the stability of the equilibria that we find.

A preliminary technical consideration deals with the fact that we only infer noisy estimates of social and legislative effort \((S_{Dem}, S_{Rep}, X_{Dem}, X_{Rep})\) from the data. Those values, however, do not necessarily solve the original (non-noisy) system defined in Proposition 2.2.1 under our set of estimated parameters, they only approximate its solution. To find the equilibrium values \((S_{Dem}, S_{Rep}, X_{Dem}, X_{Rep})\) that are consistent with our estimated parameters for \((c, \zeta, \beta_{Dem}, \beta_{Rep})\), we solve the system in Proposition 2.2.1 exactly. Those solutions are presented in the upper panel of Table 2.6. They are used to compute the model consistent bill approval rates shown in Figures 2.6 - 2.7, as well as the counterfactuals presented below. This procedure is described in more detail in Appendix.

Based on the results in Figures 2.6 - 2.7 and in the upper panel of Table 2.6 Congress is always at an interior equilibrium of our model. As there can be multiple interior equilibria, we check that the estimated equilibria are stable.\(^{34}\)

To perform this analysis we use the best response dynamics.\(^{35}\) Starting at some vectors \(s^0, x^0\) and iterating through \(t\), we get that the best response dynamics are described by:

\[
\begin{align*}
    s_t^i &= x_t^{i-1} \phi \sum_{j \neq i} m_{ij}(s_t^{j-1})s_{j}^{j-1}x_{j}^{j-1}, \\
    x_t^i &= \frac{\alpha_i}{c} + s_t^{i-1} \phi \sum_{j \neq i} m_{ij}(s_t^{j-1})s_{j}^{j-1}x_{j}^{j-1}.
\end{align*}
\] \(^{(2.41)}\) \(^{(2.42)}\)

We check whether perturbations away from equilibrium socialization and legislative efforts converge back to the estimated equilibrium efforts through the best responses of the players in the network. This is done by starting slightly below or above the values of Table 2.6, and successively iterating the best response dynamics.\(^{36}\) In all Congresses 105th-110th.

\(^{34}\)Generally, when there are multiple interior equilibria, only some are stable. This is in contradiction with Proposition 1 in [36] which claims stability of all interior equilibria. In their model, contrary to the original proof, the largest equilibrium is unstable. In the proof of that proposition the matrix \(\Pi\) cannot be approximated by setting off-diagonal terms to 0. In fact, the eigenvalue can change sign if the off-diagonal terms are included and are on the order of \(1/n\). This reverses their conclusion.

\(^{35}\)For details, we refer the reader to the Appendix. Here \(s_t^i\) takes \(x_t^{i-1}\) as given, but one could also solve for simultaneous best reply dynamics and get the same results.

\(^{36}\)This is not a full check of stability, as we are not verifying all perturbations. However, given the structure of equilibria, all best responses in a party are proportional to the same \(X, S\) and have similar dynamics.
Table 2.6: Estimated and High Effort Equilibria

<table>
<thead>
<tr>
<th>Congress</th>
<th>105</th>
<th>106</th>
<th>107</th>
<th>108</th>
<th>109</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effort Level in the Estimated Equilibrium:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{Dem}$</td>
<td>0.835</td>
<td>0.975</td>
<td>0.957</td>
<td>0.906</td>
<td>0.960</td>
<td>1.052</td>
</tr>
<tr>
<td>$S_{Rep}$</td>
<td>0.822</td>
<td>0.956</td>
<td>0.914</td>
<td>0.873</td>
<td>0.934</td>
<td>1.034</td>
</tr>
<tr>
<td>$X_{Dem}$</td>
<td>3.639</td>
<td>3.851</td>
<td>3.508</td>
<td>3.397</td>
<td>3.570</td>
<td>3.832</td>
</tr>
<tr>
<td>$X_{Rep}$</td>
<td>3.623</td>
<td>3.826</td>
<td>3.455</td>
<td>3.357</td>
<td>3.537</td>
<td>3.808</td>
</tr>
<tr>
<td><strong>Effort Level in the Higher Equilibrium:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{Dem}$</td>
<td>2.893</td>
<td>2.470</td>
<td>2.290</td>
<td>2.396</td>
<td>2.320</td>
<td>2.181</td>
</tr>
<tr>
<td>$S_{Rep}$</td>
<td>2.759</td>
<td>2.362</td>
<td>2.079</td>
<td>2.196</td>
<td>2.183</td>
<td>2.111</td>
</tr>
<tr>
<td>$X_{Dem}$</td>
<td>6.808</td>
<td>6.150</td>
<td>5.456</td>
<td>5.561</td>
<td>5.577</td>
<td>5.530</td>
</tr>
<tr>
<td>$X_{Rep}$</td>
<td>6.584</td>
<td>5.973</td>
<td>5.127</td>
<td>5.250</td>
<td>5.361</td>
<td>5.419</td>
</tr>
</tbody>
</table>

We numerically assess whether the equilibrium we have estimated is the equilibrium with the highest values of social and legislative efforts ($S_{Dem}, S_{Rep}, X_{Dem}, X_{Rep}$). To do so, for each Congress we compute two (possibly) distinct solution to the system of equations in Proposition 2.2.1 under our estimated parameters. First, we find the solutions under our estimated parameters, with starting values for effort levels at those estimated in Table 2.2 (upper panel). Second, we search for a higher effort equilibrium (lower panel). This is done by searching for a solution to equations (2.22) - (2.24), while starting from a vector with large values of effort relative to the estimated ones (namely, (100, 100, 100, 100)). The Table shows that there exists an equilibrium with higher levels of both social and legislative effort in all Congresses. These effort levels are higher than the equilibrium ones we estimated and that we observe in the data.
we find that best-response dynamics converge back to our estimated equilibrium (from the upper panel in Table 2.6) after few iterations.

Second, we compare our estimated equilibrium effort levels to those from a possible equilibrium with higher levels of effort in each Congress. To do so, we take the system defined by Proposition 2.2.1 under the estimated parameters and numerically search for solutions of this system at higher values of \((S_{\text{Dem}}, S_{\text{Rep}}, X_{\text{Dem}}, X_{\text{Rep}})\).

These results are reported in the lower panel of Table 2.6. We find that there exists a higher effort level equilibrium for every Congresses considered in our analysis. These other equilibria are distinct from the ones we estimate, with effort levels that are approximately double in magnitude to the empirically assessed ones. We also verify that these unobserved equilibria are unstable. To do so, we repeat the stability exercise, starting below the values of the higher equilibria. Here, we find that our dynamics diverge away from these higher equilibria.

From these exercises we deduce that Congress is generally in an interior low effort equilibrium and, moreover, all Congresses operate at effort levels lower than at an unobserved, unstable, high effort equilibrium, which Pareto dominates the observed equilibrium.

We conclude this discussion by briefly noting that there also always exists a “semi-corner” equilibrium. In this equilibrium legislative effort is exerted and is chosen at level \(x_i^* = \alpha_i/c\), but there is no socializing, i.e. \(s_i^* = 0\) for all \(i\). This occurs since there is no return to social effort if no other politician is socializing. Each politician acts in autarky. Effort is still provided in the model, because there are direct incentives \(\alpha_i\) for legislative effort, but no law is passed. Such an outcome is not desirable for politicians or voters due to Proposition 2.2.3. Furthermore, this semi-corner equilibrium is unstable, in the sense that, were any politician to deviate to a positive social effort \(s_j\), so would all the other politicians.

The semi-corner equilibrium is not observed, given that we observe positive socialization in Congress. Such a complete shutdown of socialization effort would be unstable, and so should not be observed for any length of time.

### 2.6.2 Counterfactuals

To conclude, we use the model to answer some hypothetical questions. We focus first on changing \(\gamma\) keeping \(\phi\) constant (i.e. changing the bill quality shock). Subsequently, we explore changes in the relative effort cost of legislative effort \(c\). Finally, we conclude with a counterfactual analysis of the Congressional response to the 2008-09 financial crisis by modifying the partisan composition of the House.
Table 2.7: Counterfactual in $\gamma$: Predicted Probability of Bill Approval

<table>
<thead>
<tr>
<th>Congress</th>
<th>105</th>
<th>106</th>
<th>107</th>
<th>108</th>
<th>109</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrats Decrease in 10% in $\hat{\gamma}$</td>
<td>0.0568</td>
<td>0.0495</td>
<td>0.0506</td>
<td>0.0626</td>
<td>0.0514</td>
<td>0.0706</td>
</tr>
<tr>
<td>Republicans Decrease in 10% in $\hat{\gamma}$</td>
<td>0.0570</td>
<td>0.0496</td>
<td>0.0509</td>
<td>0.0629</td>
<td>0.0516</td>
<td>0.0707</td>
</tr>
</tbody>
</table>

We calculate the average probability of bill approval by party, when $\gamma$ is decreased by 10% (keeping $\phi$ constant). This is done by calculating the probability of bill approval, given by $P(y_i = 1) = \frac{1}{\zeta}(s_i^*)^2$, with $s_i^*$ the solution to the non-noisy equilibrium system from Proposition 2.2.1 (further details in Appendix). We then decrease $\gamma$ by 10%.

Change in $\gamma$ with $\phi$ constant

For each of the 105th-110th Congresses, Table 2.7 shows what would happen to equilibrium efforts if $\gamma$ was reduced while keeping $\phi$ constant (noticing that $\zeta$ must adjust inversely).

Let us recall that the shock to the bill passage $\varepsilon$ is assumed to be standard Pareto distributed with scale parameter $\gamma > 0$, hence a lower $\gamma$ determines a lower median draw of the positive shock $\varepsilon$ and a lower chance of legislative success. As the system of equations in Proposition 2.2.1 does not change (the equations depend on $\phi$ directly), only the probability of approval is affected as per equation (2.9). Hence, $\gamma$ only changes the shape of the bill approval function. Quantitatively a decrease in $\gamma$ by 10 percent leads to a sizeable shift of the probability of approval curve to the left (which in Table 2.1 are shown to vary from 9.57 to 12.85 percent). This is shown in Table 2.7, which reports the average probabilities of bill approval under a smaller $\gamma$. As we only change the values of $\gamma$ in equation (2.33), the percentage change in the expected probability of bill passage is linear, dropping by 10 percent as well.

Change in $c$

We now examine some counterfactuals with respect to $c$. By lowering $c$ we lower the cost of legislative effort relative to social effort. Table 2.8 reports what would happen to the equilibrium if $c$ were reduced.

In our example we decrease the estimated value of $c$ by 1 and by 2 percent, and as-
Table 2.8: Counterfactuals in \( c \): Predicted (Proportional) Change in the (Mean) Probability of Bill Approval

<table>
<thead>
<tr>
<th>Congress</th>
<th>105</th>
<th>106</th>
<th>107</th>
<th>108</th>
<th>109</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decrease in 1% in ( \hat{c} )</td>
<td>0.082</td>
<td>0.108</td>
<td>0.116</td>
<td>0.104</td>
<td>0.115</td>
<td>0.140</td>
</tr>
<tr>
<td>Decrease in 2% in ( \hat{c} )</td>
<td>0.177</td>
<td>0.244</td>
<td>0.267</td>
<td>0.233</td>
<td>0.263</td>
<td>0.346</td>
</tr>
</tbody>
</table>

The table presents the change in the average probability of bill approval under the counterfactual, where the estimated cost \( c \), is reduced by either 1% or 2% (of the estimated value \( \hat{c} \)). We do this by calculating the implied optimal \( \{x^*_i, s^*_i\} \) from Proposition 2.2.1 under the appropriate value of \( c \) and calculate the probability of approval defined as \( \frac{1}{2}(s^*_i)^2 \). We then find the percentual change over the predicted values under Table 2.2.

Assess the changes in the probability of bill approval at our estimated equilibrium. We find a positive effect of decreasing \( c \) on the likelihood of bill passage. Moreover, this effect is quantitatively large, substantially more than linear, consistent with the evidence so far pointing towards a substantial role for social complementarities across legislators. There is an increase of bill success probability of approximately 10 percent on average across all Congresses, vis-à-vis a decrease of 1 percent in \( c \). Similarly large effects are present for a 2 percent cut. These magnitudes appear consistent with the overall importance of social interactions in legislative activity reported in Section 5.

The rationale for the sign of this effect is that a decrease in \( c \) leads to an increased choice of legislative effort that has a sufficiently strong complementarity with social effort to also drive up socializing by the individual at this equilibrium (similarly to what happens at the low effort equilibrium of [36]). This increases the returns to socializing by others, who then further increase their own legislative efforts until a new equilibrium is reached. This feeds back and leads to a large positive effect, highlighting the importance of socializing in the approval of bills. These dynamics contrast to an opposite negative feedback effect that can occur, for example, in the higher Pareto equilibrium of [36]. In that case, a decrease in \( c \) may lead to a decrease in bill approval. This is because, at the high equilibrium, a decrease in \( c \) makes socialization more costly, which reduces the incentives for socialization. In turn, this reduces the returns for \( x_i \) due to the complementarity of effort choices.
Counterfactual of the Democratic Party Takeover in the 110th Congress

We conclude this section by proposing a counterfactual of congressional behavior during the 110th cycle.

Elected in November 2006, the House of Representative turned Democratic majority after twelve years of consecutive Republican control. This revealed to be a particularly consequential election, as it was the 110th Congress that voted between the Summer and the Fall of 2008 a host of emergency economic measures in response to the 2008-09 financial crisis (for an analysis of these Congressional votes, see [134]). Some of this legislative activity happened to be extremely momentous, including the vote of the Emergency Economic Stabilization Act of 2008 (EESA, also known as the “TARP” from the Troubled Asset Relief Program), which initially failed passage in the House, inducing one of the largest intra-day losses in NYSE’s history.

An important counterfactual is to assess how relevant the role of congressional networks was in eventually guaranteeing a responsive legislative intervention to the financial crisis. How different would legislative activity have looked absent the Democratic Party takeover in 2006?

Within our framework, this counterfactual corresponds to keeping the composition of the 109th Congress in the 110th Congress. That is, we keep the observed characteristics and estimated $\alpha_i$ for the members of the 109th Congress in an analysis of outcomes in the 110th Congress. Meanwhile, the institutional setting of the 110th Congress remains the same (with the cost and returns to social effort, as well as the other institutional parameters kept at their estimated values for the 110th Congress).

Figure 2.5 shows that the distribution of $\alpha_i$ from the 109th Congress appears mostly mildly to the left of what is estimated for the 110th Congress. However, even small differences in the vector of $\alpha$ could potentially produce substantial effects through the network in terms of bill likelihood of success. In Table 2.9 we inspect the magnitude of the counterfactual reductions in the likelihood of bill passage for some of the most important emergency response legislation during the Fall of 2008. This set includes in addition to the EESA, the Housing and Economic Recovery Act of 2008 (aimed at foreclosure prevention, also studied by [134]), the Economic Stimulus Act of 2008 and the Supplementary Appropriations Act, both large bills precursor of the fiscal intervention of 2009. Table 2.9 reports the relative differences in bill passage probabilities between the counterfactual and the estimated model, as well as the baseline probabilities.\textsuperscript{37} It appears all differences are in the range of a

\textsuperscript{37}Some of these bills’ complex histories align with a low likelihood of success. The EESA of 2008 failed the House. In addition, about H.R. 3221, [134] write: “Roll call 301: “On Agreeing to the Senate Amendment
3 – 5 percent reduction in the likelihood of success, a quantitatively small effect considering the baseline probability of approval. That is, the social network composition of the House would have not changed in a sufficiently different way to substantially affect final voting outcomes. This is counter to the claim that, absent the Democratic takeover of the House in 2006, the financial crisis response would have been substantially different, with a more restrained government intervention under a Republican Congress (obviously, conditionally on the same set of emergency legislation being pushed forth by Treasury Secretary Hank Paulson and Federal Reserve Chairman Ben Bernanke). Of course, our exercise can only speak to the quantity of legislation and not to its content.

2.7 Conclusions

We have developed and estimated a structural model of legislative activity for the U.S. Congress in which endogenous, partisan social interactions play an important role in promoting bill passage. We estimate that social effort matters substantially and significantly for legislative activity.

By endogenizing both legislative and social efforts, we are able to accommodate complementarities in actions that appear to be strong. In particular, we find that the complementarities among politicians are quantitatively substantial (on the order of a tenth to a third of the direct incentives), and are fairly stable across our sample period. We also find that the two parties have different base payoffs from passing legislation, both in terms of the average and variance across party members (both are higher for the Democrats). Overall, we show how the process of informal social interaction among legislators may paint a less extreme, although still partisan, picture of the internal operation of Congress.

Multiple equilibria arise naturally within our theoretical setting (as it is typical of models of endogenous network formation). By careful consideration of the theoretical model and its behavior around the estimated equilibrium, we are able to show that Congress appears in a stable, low-socialization equilibrium, with effort levels lower than in a Pareto superior, but unstable, equilibrium present in all Congresses.

Finally, our estimated model enables us to perform some counterfactuals. We show
Table 2.9: Counterfactuals in $\alpha$: Looking at the Changes in (Ex-Ante) predicted probability of Emergency Crisis bills in the 110th Congress, if the Republicans who lost their seats remained Act Proportional Baseline Probability Change of Success

<table>
<thead>
<tr>
<th>Act</th>
<th>Proportional Change</th>
<th>Baseline Probability of Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergency Economic Stabilization Act of 2008. (H.R. 1424)</td>
<td>-0.0505</td>
<td>0.0848</td>
</tr>
<tr>
<td>Sponsor: Patrick Kennedy, Democrat - RI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing and Economic Recovery Act of 2008 (H.R. 3221)</td>
<td>-0.0391</td>
<td>0.0898</td>
</tr>
<tr>
<td>Sponsor: Nancy Pelosi, Democrat - CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic Stimulus Act of 2008 (H.R. 5140)</td>
<td>-0.0391</td>
<td>0.0898</td>
</tr>
<tr>
<td>Sponsor: Nancy Pelosi, Democrat - CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplementary Appropriations Act, 2008 (H.R. 2642)</td>
<td>-0.0518</td>
<td>0.0706</td>
</tr>
<tr>
<td>Sponsor: Chet Edwards, Democrat - TX</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table presents the proportional change (Counterfactual - Model)/Model of the probability of each bill passing under our counterfactual scenario. The counterfactual scenario is keeping the Republican majority and composition from the 109th Congress, in the 110th Congress. To do so, we keep the characteristics of all politicians from the 109th Congress with the estimated parameters of the 110th Congress. The only difference to characteristics is we add 1 for each politician’s Tenure variable, as those politicians would have stayed 1 extra term in the counterfactual. We do not change the Committee composition or ideology. We then calculate the projected probability of bill approval using the estimated parameters from the 110th Congress. The baseline (model) probability is shown in the second column, computed using $\hat{p}_{Dem}, \hat{p}_{Rep}$ from Table A.1 in Appendix.

substantial impacts from changing the relative cost of social effort in terms of probability of bill passage and estimate that the response to the 2008-09 financial crisis would not have changed in terms of levels of legislation if the Democrats had not taken over the House. With more recent waves of data, further counterfactuals (for example, related to the role of the Tea Party movement) could be performed.
Chapter 3

Estimating Local Interactions
Among Many Agents Who Observe Their Neighbors

3.1 Introduction

Interactions between agents - for example, through personal or business relations - generally lead to their actions being correlated. In fact, such correlated behaviors form the basis of identifying and estimating peer effects, neighborhood effects, or more generally social interactions in the literature. (See [25] and [62] for a review of this literature.)

Empirical modeling becomes nontrivial when one takes seriously the fact that people are often connected directly or indirectly on a large complex network, observing some others’ types, and that the econometrician observes only a small fraction of those on the network. Furthermore, strategic environments are highly heterogeneous across agents as each agent occupies a nearly “unique” position in the network. Information sharing potentially creates a complex form of cross-sectional dependence among the observed actions and yet the econometrician rarely has precise information about the actual network on which people observe other people.

The main contribution of this chapter is to develop a tractable empirical model of linear interactions among agents with the following two major features. First, assuming a large game on a complex exogenous network, the empirical model allows the agents not to

\footnote{This chapter is a joint work with Jacob Schwartz and Kyungchul Song.}
observe the full network, but to observe only part of the types of their neighbors.\footnote{For example, a recent paper by [32] documents that people in a social network has a substantial lack of knowledge on the network, and that the violation of this assumption may have significant implications in the predictions of the model.}

Second, our model explains strategic interdependence among agents through correlated observed behaviors. In this model, the locality of cross-sectional dependence among the observed actions reflects the locality of strategic interdependence among the agents. Most importantly, unlike most incomplete information game models in the literature, our set-up allows for information sharing on unobservables, i.e., each agent is allowed to observe his neighbors’ payoff relevant signals that are not observed by the econometrician.

Third, the econometrician does not need to observe the whole set of players in the game for inference. It suffices that he observes many (potentially) non-random samples of local interactions. The inference procedure that this chapter proposes is asymptotically valid independently of the actual sampling process, as long as the sampling process satisfies certain weak conditions. Accommodating a wide range of sampling processes is useful because random sampling is rarely used for the collection of network data, and a precise formulation of the actual sampling process is often difficult in practice.

A standard approach to model interactions among agents is to model them as a game and use equilibrium strategies from the game to obtain predictions and testable implications. Such an approach is cumbersome in our set-up. Since a particular realization of any agent’s type affects all the other agents’ actions in equilibrium through a chain of information sharing, each agent needs to form a “correct” belief about the entire information graph. Apart from such an assumption being highly unrealistic, it also implies that predictions from an equilibrium that the econometrician uses to form testable implications generally involve all the players in the game, when it is often the case that only part of the players are observed in practice. Thus an empirical analysis which regards the players in the sample as coincident with the actual set of players in the game will suffer from lack of external validity when his target “population” is the original large game involving much more players than those in the sample.

Instead, this chapter adopts an approach of behavioral modeling, where it is assumed that each agent, not knowing fully the information sharing relations, optimizes according to his simple beliefs about other players’ strategies. The crucial part of our behavioral assumption is a primitive form of belief projection which says that each agent, not knowing who his payoff neighbors observe, projects his own beliefs about other players onto his payoff neighbors. More specifically, if agent $i$ gives more weight to agent $j$ than to agent $k$, A standard approach to model interactions among agents is to model them as a game and use equilibrium strategies from the game to obtain predictions and testable implications.
agent \( i \) believes that each of his payoff neighbor \( s \) does the same in comparing agents \( j \) and \( k \).

Belief projection in our chapter is a variant of inter-personal projection studied in behavioral economics. A related behavioral concept is projection bias of [118] which refers to the tendency of a person projecting his own current taste to his future taste. See also [167] who reported experiment results on the interpersonal projection of tastes onto other agents. Since formation of belief is often tied to the information set the agent has, belief projection is closely related to information projection in [119] who focuses on the tendency of a person projecting his information to other agents’ information. The main difference here is that our focus is to formulate the assumption in a way that is useful for inference using observational data on actions on a network.

We show that our primitive form of belief projection yields an explicit form of the best linear response which has intuitive features. For example, the best linear response is such that each agent \( i \) gives more weights to those agents with a higher local centrality to him, where the local centrality of agent \( j \) to agent \( i \) is defined to be high if and only if a high fraction of agents from those whose actions affect agent \( i \)’s payoff have their payoffs affected by agent \( j \)’s action. Also, each agent’s action responds to a change in his own type more sensitively when there are stronger strategic interactions, due to what we call the reflection effect. The reflection effect of player \( i \) captures the way player \( i \)’s type affects his own action through his payoff neighbors whose payoffs are affected by player \( i \)’s types and actions.

One might wonder how close the predictions from our behavioral model is to the predictions from an equilibrium model. For this we consider a simple linear interactions model as a complete information game where one can compute the equilibrium explicitly. The equilibrium strategies are given in a primitive form of a spatial autoregressive model. We compare the network externality from our behavioral model and that from the complete information game model using simulated graphs, one from Erdös-Rényi random graphs and the other from a scale-free random graph generation of Barabási-Albert. In both cases, it is shown that both models have similar predictions when the payoff externality parameter is less than or equal to 0.5. However, when it is close to one, the network externality becomes much higher in the equilibrium model than in the behavioral model. This is because while strong local interactions induce global cross-sectional dependence in the equilibrium model due to extensive information transmission, it does not in our behavioral model. Also, as the network size increases, the network externality from our behavioral model changes more stably than that from the equilibrium strategies from a complete information game.
We investigate the finite sample properties of asymptotic inference through Monte Carlo simulations using various payoff graphs. The results show reasonable performance of the inference procedures. In particular, the size and the power of the test for the strategic interaction parameter work well in finite samples. We also apply our method to an empirical application of decisions of municipalities on state presence revisiting the study by [4]. We consider an incomplete information game model which permits information sharing. The fact that our best linear responses explicitly reveal the local dependence structure means that it is unnecessary to separately correct for spatial correlation following, for example, the procedure of [51].

The literature of social interactions often look for evidence of interactions through correlated behaviors. For example, linear interactions models investigate correlation between $Y_i$ and the average of outcomes over agent $i$’s neighbors. See for example [120], [55], [30] and [26] for identification analysis in linear interactions models, and see [38] for an application in the study of peer effects. [84] considers nonlinear interactions on a social network and discusses endogenous network formation. These models often assume that we observe many independent samples of such interactions, where each independent sample constitutes a game which contains the entire set of the players in the game.

In the context of a complete information game, linear interactions models on a large social network can generally be estimated without assuming independent samples. The outcome equations frequently take the form of spatial autoregressive models which have been actively studied in the literature of spatial econometrics. ([11]) A recent study by [100] consider a model of linear interactions on a large social network which allows for endogenous network formation. Developing inference on a large game model with nonlinear interactions is more challenging. See [132], [172], [160], [173], and [174] for a large game model of nonlinear interactions. This large game approach is suitable when the data set does not have many independent samples of interactions. One of the major issues in the large game approach is that the econometrician often observes only part of the agents in the original game.\footnote{[160], [172], [100], [173] and [174] assume observing all the players in the large game. In contrast, [132] allows for observing i.i.d. samples from the many players, but assumes that each agent’s payoff involves all the other agents’ actions exchangeably.}

Our approach of empirical modeling is also based on a large game model which is closer to the tradition of linear interactions models in the sense that our approach attempts to explain strategic interactions through correlated behaviors among neighbors. In our setup, the cross-sectional dependence of the observed actions is not merely a nuisance that
complicates asymptotic inference; it provides the very piece of information that reveals the strategic interdependence among agents. The correlated behaviors also arise in equilibrium in models of complete information games or games with types that are either privately or commonly observable. (See [30] and [26].) However, as emphasized before, such an approach can be cumbersome in our context of a large game primarily because the testable implications from the model typically involve the entire set of players, when in many applications the econometrician observes only a small subset of the players in the large game. [64] model the interactions as a Bayesian game on a large network with private link information. Like this chapter, they permit the agents not to observe the full network, and show identification of the model primitives adopting a Bayesian Nash equilibrium as a solution concept. One of the major differences of our work from theirs is that we permit information sharing on unobservables, so that the actions of neighboring agents are potentially correlated even after controlling for observables.

A departure from the equilibrium approach in econometrics is not new in the literature. [13] studied the implications of various rationality assumptions for identification of the parameters in a game. Unlike their approach, our focus is on a large game where many agents interact with each other on a single complex network, and, instead of considering all the beliefs which rationalize observed choices, we consider a particular set of beliefs that satisfy a simple rule and yield an explicit form of best linear responses. (See also [81] and [93] for empirical research adopting behavioral modeling for interacting agents.)

This chapter is organized as follows. In Section 3.2, we introduce an incomplete information game of interactions with information sharing. This section derives the crucial result of best linear responses under simple belief rules. In this section, we discuss the issue of external validity of network externality comparing two simple interactions models: a complete information game with equilibrium strategies and our behavioral model. Section 3.3 focuses on econometric inference. This section presents inference procedures, explains a situation where we can measure the role of information sharing on unobservables and compares our approach with a standard linear-in-means model. Section 3.4 investigates the finite sample properties of our inference procedure through a study of Monte Carlo simulations. Section 3.5 presents an empirical application on state capacity among municipalities. Section 3.6 concludes. The technical proofs of the results are found in the Appendix.
3.2 Strategic Interactions with Information Sharing

3.2.1 A Model of Interactions with Information Sharing

Strategic interactions among a large number of information-sharing agents can be modeled as an incomplete information game. Let \( N \) be the set of a finite yet large number of players. Each player \( i \in N \) is endowed with his type vector \((T_i, \eta_i)\), where \( \eta_i \) is a private type and \( T_i \) a sharable type. As we will elaborate later, information \( \eta_i \) is kept private to player \( i \) whereas \( T_i \) is observed by his “neighbors” which we define later. Throughout this chapter, we set \( T_i = (X'_i, \varepsilon_i)' \), where \( X_i \) is the vector of characteristics of player \( i \) that are observed by the econometrician, and \( \varepsilon_i \) the unobserved characteristic of player \( i \). Thus the model permits information sharing on unobservables \( \varepsilon_i \). This feature in fact makes a significant departure from many existing incomplete information interactions models which assume that variables that the econometrician observes are public among the agents whereas the variables that the econometrician does not observe are kept private among themselves. (e.g. [26])

To capture the strategic interactions among players, let us introduce an undirected graph \( G_p = (N, E_p) \), where \( E_p \) denotes the set of edges \( ij, i, j \in N \) with \( i \neq j \) and each edge \( ij \in E_p \) represents that the action of player \( i \) affects player \( j \)’s payoff. We denote \( N_p(j) \) to be the \( G_p \)-neighborhood of player \( j \), i.e., the collection of players whose actions affect the payoff of player \( j \):

\[
N_p(j) = \{ i \in N : ij \in E_p \},
\]

and let \( n_p(j) = |N_p(j)| \). We define \( \overline{N}_p(i) = N_p(i) \cup \{i\} \) and let \( \overline{n}_p(i) = |\overline{N}_p(i)| \).

Player \( i \) choosing action \( y_i \in \mathcal{Y} \) with the other players choosing \( y_{-i} = (y_j)_{j \neq i} \) obtains payoff:

\[
u_i(y_i, y_{-i}, T, \eta_i) = y_i \left( X'_{i,1} \gamma_0 + \bar{X}'_{i,2} \delta_0 + \beta_0 \bar{y}_i + \varepsilon_i + \eta_i \right) - \frac{1}{2} y_i^2,
\]

where \( T = (T_i)_{i \in N}, X_{i,1} \) and \( X_{i,2} \) are subvectors of \( X_i \),

\[
\bar{X}_{i,2} = \frac{1}{n_p(i)} \sum_{k \in N_p(i)} r_{ik} X_{k,2}, \text{ and } \bar{y}_i = \frac{1}{n_p(i)} \sum_{k \in N_p(i)} r_{ik} y_k,
\]

if \( N_p(i) \neq \emptyset \), and \( \bar{X}_{i,2} = 0 \) and \( \bar{y}_i = 0 \) otherwise. The factor \( r_{ik} \) measures the “relative weight”

---

\(^{4}\)A graph \( G = (N, E) \) is undirected if \( ij \in E \) whenever \( ji \in E \) for all \( i, j \in N \).
of individual $k$ in the network from the viewpoint of $i$. Here, we consider two specifications.

Specification A : $r_{ik} = 1$, for all $i, k \in N$. \hspace{1cm} (3.1)

Specification B : $r_{ik} = \frac{\bar{p}_p(k)}{\bar{p}_p(i)}$, for all $i, k \in N$.

The simple choice $r_{ik} = 1$ gives equal weight to every other agent, but the choice of $r_{ik} = \frac{\bar{p}_p(k)}{\bar{p}_p(i)}$ give more weights to those who have more edges with others relative to agent $i$. Thus the payoff depends on other players’ actions and types only through those of his $G_P$-neighbors. We call $G_P$ the payoff graph.

The parameter $\beta_0$ measures the payoff externality among agents. In the terminology of Manski (1993), $\delta_0$ captures the exogenous effect and $\beta_0$ the endogenous effect of social interactions. As for $\beta_0$, we make the following assumption:

Assumption 3.2.1. $0 \leq |\beta_0| < 1$.

This assumption is often used to derive a characterization of a unique pure strategy equilibrium in the literature. (See e.g. [30] and [26] for its use.) When $\beta_0 > 0$, the game is called a game of strategic complements and, when $\beta_0 < 0$, it is called a game of strategic substitutes.

Let us introduce information sharing relations in the form of a directed graph (or a network) $G_I = (N, E_I)$ on $N$ so that each $ij$ in $E_I$ represents the edge from player $i$ to player $j$, where the presence of edge $ij$ joining players $i$ and $j$ indicates that $T_i$ is observed by player $j$. Hence the presence of an edge $ij$ between agents $i$ and $j$ represents information flow from $i$ to $j$. We call graph $G_I$ the information graph. For each $j \in N$, define

$$N_I(j) = \{i \in N : ij \in E_I\},$$

that is, the set of $G_I$-neighbors observed by player $j$. Also write

$$\overline{N}_I(i) = N_I(i) \cup \{i\},$$

i.e., the $G_I$-neighborhood of $i$ including $i$ himself. We define $\overline{n}_I(i) = |\overline{N}_I(i)|$.

We do not assume that each agent knows the whole information graph $G_I$ and the payoff graph $G_P$. To be precise about each agent’s information set, let us introduce some notation.

More precisely, the neighbors in $N_I(j)$ are called in-neighbors and $n_I(i) = |N_I(i)|$ in-degree. Throughout this chapter, we simply use the term neighbors and degrees, unless specified otherwise.
For each $i \in N$, we set $\overline{N}_{P,1}(i) = \overline{N}_P(i)$ and $\overline{N}_{I,1}(i) = \overline{N}_I(i)$, and for $k \geq 2$, define recursively

$$\overline{N}_{P,k}(i) = \bigcup_{j \in \overline{N}_P(i)} \overline{N}_{P,k-1}(j), \quad \text{and} \quad \overline{N}_{I,k}(i) = \bigcup_{j \in \overline{N}_I(i)} \overline{N}_{I,k-1}(j).$$

Thus $\overline{N}_{P,k}(i)$ denotes the set of players which consist of player $i$ and those players who are connected to player $i$ through at most $k$ edges in $G_P$, and similarly with $\overline{N}_{I,k}(i)$. Also, define $N_{P,k}(i) = \overline{N}_{P,k}(i) \setminus \{i\}$ and $N_{I,k}(i) = \overline{N}_{I,k}(i) \setminus \{i\}$. For each $k \geq 1$, let $\mathcal{N}_{i,k-1}$ be the $\sigma$-field generated by $\overline{N}_{P,k+1}(i), \overline{N}_I(i)$ and some additional information $\mathcal{E}_i$ which potentially causes correlation between types across different players. (We will explain $\mathcal{E}_i$ later.) That is, for $k \geq 1$,

$$\mathcal{N}_{i,k-1} = \sigma(\overline{N}_{P,k+1}(i), \overline{N}_{P,k}(i), \ldots, \overline{N}_{P,2}(i), \overline{N}_I(i)) \lor \mathcal{E}_i,$$

where $\lor$ between two $\sigma$-fields is the smallest $\sigma$-field among those which contain the two $\sigma$-fields. Define for each $k \geq 0$,

$$\mathcal{J}_{i,k} = \sigma(T_{\overline{N}_I(i)}, \eta_i) \lor \mathcal{N}_{i,k},$$

where $T_{\overline{N}_I(i)} = (T_j)_{j \in \overline{N}_I(i)}$. We use $\mathcal{J}_{i,k}$ to represent the information set of agent $i$. For example, when agent $i$ has $\mathcal{J}_{i,1}$ as his information set, it means that agent $i$ knows the set of agents whose types he observes (i.e., $N_I(i)$), the set of agents $j$ whose actions affect his payoff (i.e., $N_{P,1}(i)$) and the set of agents whose actions affect the payoff of his $G_P$-neighbors $j$ (i.e., $N_{P,2}(i)$), and the sharable types of his $G_I$-neighbors (i.e., $T_{\overline{N}_I(i)}$) and his own private signal $\eta_i$.

Throughout the chapter, it is not assumed that any agent $i$ knows $N_I(k)$ for any of his $G_P$-neighbors $k$. In other words, there might be some $G_P$-neighbor $k$ who may observe other agents that agent $i$ does not observe, and agent $i$ does not know who such $G_P$-neighbor $k$ is or who those other agents player $k$ observes are.

Regarding the joint distribution of the profile of sharable types $T$, we make the following assumption:

**Assumption 3.2.2.** For each $i \in N$, $T_{N \setminus \overline{N}_I(i)}$ and $T_{\overline{N}_I(i)}$ are conditionally independent given $(G_P, \overline{N}_I(i))$ and $\mathcal{E}$, where

$$\mathcal{E} = \lor_{i \in N} \mathcal{E}_i.$$

This assumption allows the individual types to be correlated unconditionally. Each
player $i$ has information $C_i$ which can cause correlation between his type and other agents’ types. For example, any two types $T_i$ and $T_j$ may contain a common signal which comes from a common observation by the two agents $i$ and $j$.

Assumption 3.2.2 says that the sharable types between two non-neighbors in $G_I$ are independent conditional on all such pieces of information $C_i$.

The assumption permits the situation where the payoff network $G_P$ is exogenously formed, for example, as a dyadic regression model degree heterogeneity, $a_i$, with errors $u_{ij}$’s that are independent of $\varepsilon_i$’s, $\eta_j$’s, $X_i$’s and $a_i$’s. (See e.g. [89].) In this case, if we set $C_i = \sigma(X_i, a_i)$, Assumption 3.2.2 is reduced to that for each $i \in N$, $\varepsilon_{N\setminus I_i(i)}$ and $\varepsilon_{N_j(i)}$ are conditionally independent given $(G_P, N_I(i), X, a)$, where $X = (X_i)_{i \in N}$ and $a = (a_i)_{i \in N}$.

### 3.2.2 Predictions from Rationality

Each player chooses a strategy that maximizes his expected payoff according to his beliefs. This provides predictions for their actions given their beliefs. To characterize predictions from rationality, we introduce some notation. For $i, j, k \in N$, let $w^i_{kj}$ denote the weight that player $i$ believes that player $k$ gives to player $j$. Suppose that the strategy of player $k$ as believed by player $i$ is given as follows:

$$s^i_k(I_k) = \sum_{j \in N^i_k} T^j w^i_{kj} + \eta_k, \quad (3.2)$$

where $N^i_k(k)$ denotes the set of players (including player $k$) who player $i$ believes that player $k$ observes. Given player $i$’s strategy and his expected strategy of other players $s^i_{-i} = (s^i_k)_{k \in N \setminus \{i\}}$, the (interim) expected payoff of player $i$ is defined as

$$U_i(s_i, s^i_{-i}; I_i) = E[u_i(s_i(I_i), s^i_{-i}(I_{-i}), T, \eta_i)|I_i],$$

where $s^i_{-i}(I) = (s^i_k(I))_{k \in N \setminus \{i\}}$, $I_{-i} = \bigvee_{k \neq i} I_k$ and $T = (T_i)_{i \in N}$. A best linear response $s^i_{BR}$ of player $i$ corresponding to the strategies $s^i_{-i}$ of the other players as expected by player $i$ is a linear strategy such that for any linear strategy $s_i$,

$$U_i(s^i_{BR}, s^i_{-i}; I_i) \geq U_i(s_i, s^i_{-i}; I_i), \text{ a.e.}$$

Under the assumptions of the model, the best linear responses can be shown to produce
a map from beliefs to actions. To see this, first let

\[ w^B = (w^1, ..., w^n) \]

be the belief profile of all the agents, where \( w^i = (w^i_{kj})_{k, j \in N} \). Then the rationality of agents (i.e., their choosing a best linear response given their beliefs) gives the following relation:

\[ w = \mathcal{M} w^B, \]

where \( w = (w_{ij})_{i, j \in N} \) corresponds to best responses and \( \mathcal{M} \) is the best response operator which assigns a strategy profile (in terms of weights \( w_{ij} \)) to a given belief profile \( w^B \). (The explicit form of the best response operator is found in the Appendix.)

In order to generate predictions, one needs to deal with the beliefs \( w^B \). There are three approaches to model these beliefs. The first approach is an equilibrium approach where the beliefs \( w^B \) coincide with the actual weights implemented by the agents in equilibrium. The second approach uses rationalizability where all the linear strategies that are rationalizable given some belief \( w^B \) are in consideration. The third approach is a behavioral approach where one considers a set of simple behavioral assumptions on the beliefs \( w^B \) and focuses on the best linear responses to corresponding to these beliefs.

There are pros and cons among the three approaches. One of the main differences between the equilibrium approach and the behavioral approach is that the former approach requires the beliefs \( w^B \) to be “correct” for all players \( i \) in equilibrium. However, since each player \( i \) generally does not know who each of his \( G_P \)-neighbors observes, a Bayesian player in a standard model with rational expectations would need to know the distribution of the entire information graph \( G_i \) (or at least have a common prior on the information graph commonly agreed upon by all the players) to form a “correct” belief given his information. Given a potentially complex form of \( G_P \) (partially observed in data) and that the econometrician rarely observes \( G_I \) with precision, producing a testable implication from this equilibrium model appears far from a trivial task.

The rationalizability approach can be used to relax this rational expectations assumption by eliminating the requirement that the beliefs be correct. The approach considers all the predictions that are rationalizable given some beliefs. However, in our context, the best response operator \( \mathcal{M} \) depends on unknown parameters in general, and hence the set of predictions from rationalizability can potentially be large and may fail to produce sharp predictions that would be useful in practice.

As we explain later in detail, this chapter takes the third approach. We adopt a set of
simple behavioral assumptions on players’ beliefs which can be incorrect from the viewpoint of a person with full knowledge on the distribution of the information graph, yet useful as a rule-of-thumb guidance for an agent in a complex decision-making environment such as one in our model. As we shall see later, this approach can give a sharp prediction that is intuitive and analytically tractable.

3.2.3 Belief Projection and Best Linear Responses

In this chapter, we consider the following set of behavioral assumptions on the beliefs.

Condition BP (Belief Projection): (i) For each \( i \in N \) and \( k \in N_P(i) \),
(a) \( w^i_{kk} = w_{ii} \),
(b) \( w^i_{kj} = \tau^i_{kj} w_{ij} \) for all \( j \in N_I(i) \cap N^i_I(k) \) for some positive number \( \tau^i_{kj} \), where
\[
\tau^i_{ki} = 1 / (r_{ik} \bar{n}_P(k)), \quad \text{and} \quad (3.3)
\]
\[
\tau^i_{kj} = 1 / r_{ik}, \quad \text{for all} \quad j \in N_I(i) \cap N^i_I(k), \quad \text{and}
\]
(ii) \( w^i_{kj} = 0 \) for all \( j \notin \bar{N}_P(k) \).

As mentioned before, each player \( i \) does not know who his \( G_P \)-neighbors observe, and Condition BP describes a simple rule of belief formation in this environment. The main premise of Condition BP is that each agent projects his own beliefs about himself and other players onto his \( G_P \) neighbors. Condition BP (i)(a) says that each player \( i \) believes that the self-weight his \( G_P \)-neighbor \( k \) gives to himself is the same as the self-weight of player \( i \) himself. Condition BP (i)(b) says that player \( i \)'s belief on his \( G_P \) neighbor \( k \)'s weight to player \( j \) is formed in reference to his own weight to player \( j \). This assumption says that each agent believes that his \( G_P \) neighbors follow the same ranking of other agents as he does. The belief projection is taken as a rule of thumb for each agent \( i \) who needs to form an expectation about his \( G_P \)-neighbors’ actions when he does not know who his \( G_P \)-neighbors observe.

The specification of \( \tau^i_{ki} \) in (3.3) reflects that player \( i \) believes that player \( k \) does not care much about player \( i \)'s type in choosing an action if the player \( k \) has many \( G_P \)-neighbors. The specification of \( \tau^i_{kj} \) in (3.3) says that each player \( i \) believes that the weight of each of his \( G_P \)-neighbors given to a \( G_P \)-neighbor \( j \) is \( (1/r_{ik}) w_{ij} \). For example, if \( r_{ik} = \bar{n}_P(k)/\bar{n}_P(i) \), we have
\[
w^i_{kj} = \frac{\bar{n}_P(i)}{\bar{n}_P(k)} w_{ij}.
\]

52
Therefore, player $i$ believes that when player $k$ has more $G_P$-neighbors than he does, player $k$ gives less weight to player $j$ than he does. Not knowing who player $k$ observes, player $i$ employs this rule-of-thumb belief regarding player $k$’s weights given to other players.

Condition BP(ii) is concerned with player $i$’s belief about the players that his $G_P$-neighbors observe. A standard approach in an incomplete information game with Bayesian players assumes that the players agree on a common prior on the entire information graph $G_I$. From this, each agent $i$ derives his posterior on the $G_I$-neighbors of each of his $G_P$-neighbors. Instead, Condition BP(ii) states that player $i$ simply considers only those players in $N_P(k)$ when he deliberates on those players whose action affects the payoff of player $k$. This is because while player $i$ knows player $k$’s $G_P$-neighborhood, he does not know player $k$’s $G_I$-neighborhood.

Let us distinguish between different environments with different information structures of the game.

**Definition 3.2.1.** (i) Each agent $i \in N$ is said to be of simple type if (a) she has beliefs about the other players’ strategies as in Condition BP, (b) believes that other players play a linear strategy of the form in (3.2), and (c) has information set $\mathcal{I}_i = \mathcal{I}_{i,0}$ with $N_{P_2}(i) \subseteq N_I(i)$.

(ii) Each agent $i \in N$ is said to be of first order sophisticated type, if she believes that the other players are of simple type and has information set $\mathcal{I}_i = \mathcal{I}_{i,1}$ with $N_{P_3}(i) \subseteq N_I(i)$.\[6\]

The difference between the simple type and a sophisticated type lies not only in the difference in the rationality type but also in the information set. A first order sophisticated type agent knows who the neighbors of the neighbors of their neighbors in $G_P$ (i.e., $N_{P_3}(i)$) are, whereas a simple type agent knows only who the neighbors of their neighbors in $G_P$ (i.e., $N_{P_2}(i)$) are.

Regarding the sophistication of agents, we make explicit the following basic assumption which we assume throughout.

**Assumption 3.2.3.** The game is populated by agents with the same order of sophistication.

Different levels of reasoning for agents of the game are assumed in level $k$ models which have received a great deal of attention as a behavioral model in the experiment literature. (See Chapter 5 of [39] for a review.) In these experiments, a simple type agent is often much simpler than those in our set-up, where the agent chooses an action without considering any

---

\[6\] One can also define a higher order sophisticated type, although this chapter does not fully elaborate on such a case. More specifically, for $k \geq 2$, each agent $i \in N$ who believes that the other players are of the $(k-1)$-th order sophisticated type and has information set $\mathcal{I}_i = \mathcal{I}_{i,k}$ with $N_{P_{k+2}}(i) \subseteq N_I(i)$ is said to be of the $k$-th order sophisticated type.
strategic interdependence. In contrast, our simple type agent already considers strategic interdependence and forms a best linear response. On the other hand, the experiment literature of level-$k$ models allows the agents to be of different rationality type in the same game. In our set-up which focuses on observational data, identification of the unknown proportion of each rationality type appears far from trivial. Hence, we consider a game where all the agents have the same order of sophistication.

Our focus on linear strategies in combination with other assumptions gives an explicit form of best linear responses. For the expression, let us introduce some notation: for each $i \in N$ and $j \in \mathcal{N}_i(i)$,

$$c_{ij} \equiv \frac{1}{n_p(i)} \sum_{k \in N_p(i)} 1\{j \in N_p(k)\}, \text{ if } i \neq j, \quad \text{and} \quad (3.4)$$

$$c_{ii} \equiv \frac{1}{n_p(i)} \sum_{k \in N_p(i)} 1\{i \in N_p(k)\} = \frac{1}{n_p(i)} \sum_{k \in N_p(i)} \frac{1}{n_p(k)}$$

where the last equality follows due to $G_P$ being undirected. Note that $c_{ij}$ is the proportion of player $i$’s $G_P$-neighbors whose payoffs are influenced by the type and action of player $j$. Hence $c_{ij}$ represents the local centrality of player $j$ to player $i$ in terms of player $j$’s influence on player $i$’s $G_P$-neighbors. On the other hand, $c_{ii}$ is the average of $1/n_p(k)$ among player $i$’s $G_P$-neighbors $k$ whose payoffs are affected by player $i$’s sharable type and action.

Using the explicit form of the best response operator $M$ and Condition BP, we can derive the explicit form of best linear responses. The following theorem gives the form in the case where all the players are of simple type.

**Theorem 3.2.1.** Suppose that Assumptions 3.2.1 - 3.2.3 hold and all the players are of simple type. Suppose further that for each $i \in N$ and $k \neq i$, $E[\eta_k|\mathcal{F}_i] = 0$. Then each player $i$’s best linear response $s_{i}^{BR}$ takes the following form:

$$s_{i}^{BR}(\mathcal{F}_i) = \lambda_{ii} \left( \gamma X_{i,1} + \epsilon_i + \frac{\beta_0}{n_p(i)} \sum_{j \in N_p(i)} \lambda_{ij} (\gamma X_{j,1} + \epsilon_j) \right) + \frac{1}{n_p(i)} \sum_{j \in N_p(i)} \lambda_{ij} \delta X_{j,2} + \eta_i,$$

where $\lambda_{ij} \equiv r_{ij}/(1 - \beta_0 c_{ij})$.

The result in Theorem 3.2.1 shows multiple intuitive features. First, it shows that each player $i$’s best linear response does not depend on the types of payoff-irrelevant agents
whose types player \(i\) observes but whose actions do not affect player \(i\)'s payoff. Note that agents indirectly connected to agent \(i\) in \(G_P\) can still shape the player’s strategies through the local centralities \(c_{ij}\). (Later, we also consider the case of sophisticated type, where the types of indirectly connected agents are permitted to influence the agent \(i\)'s actions.)\(^7\)

Furthermore, observe that for \(j \in N_P(i)\),

\[
\frac{\partial s_i^{BR}(\mathcal{X}_i)}{\partial x_{j,1}} = \frac{\beta_0 r_{ij}}{n_P(i)(1 - \beta_0 c_{ii})(1 - \beta_0 c_{ij})^\gamma} \text{ and } \frac{\partial s_i^{BR}(\mathcal{X}_i)}{\partial x_{j,2}} = \frac{\delta_0 r_{ij}}{n_P(i)(1 - \beta_0 c_{ij})},
\]

both of which measure the response of actions of agent \(i\) to a change in the observed type change of his \(G_P\)-neighbors. Hence, these quantities capture the network externality in the strategic interactions.

First, note that the network externality for agent \(i\) from a particular agent \(j\) decreases in the neighborhood size \(n_P(i)\) of agent \(i\). More importantly, the network externality for each agent \(i\) is different across \(i\)'s and across their \(G_P\) neighbors \(j\) depending on their “importance” to agent \(i\) in the payoff graph. This is seen from the network externality (3.5) being an increasing function of agent \(j\)'s local centrality to agent \(i\), i.e., \(c_{ij}\), when the game is that of strategic complements (i.e., \(\beta_0 > 0\)). In other words, the larger the fraction of agent \(i\)'s \(G_P\)-neighbors whose payoff is affected by agent \(j\)'s action, the higher the network externality of agent \(i\) from agent \(j\)'s type change becomes. Therefore, in our model network externality is heterogeneous across agents, depending on the local feature of the payoff graph around each agent.

It is interesting to note that the network externality for agent \(i\) with respect to his own type \(X_{i,1}\) has a factor \(\lambda_{ii} \equiv r_{ii}/(1 - \beta_0 c_{ii}) = 1/(1 - \beta_0 c_{ii})\) which is increasing in \(c_{ii}\) when \(\beta_0 > 0\). We call

\[
\frac{1}{1 - \beta_0 c_{ii}} - 1
\]

the reflection effect which captures the way player \(i\)'s type affects his own action through his \(G_P\) neighbors whose payoffs are affected by player \(i\)'s types and actions. The reflection effect arises because each agent, in decision making, considers the fact that his type affects other \(G_P\)-neighbors’ decision making. When there is no payoff externality (i.e., \(\beta_0 = 0\)),

\(^7\)The local dependence of actions from best linear responses regardless of what values \(\beta_0\) take in \((-1, 1)\) is in contrast with the complete information version of the game, where a high value of \(\beta_0\) makes the dependence close to be global.
the reflection effect is zero. However, when there are strong strategic interactions or when a majority of player $i$'s $G_P$-neighbors have a small $G_P$-neighborhood (i.e., small $n(k)$ in the definition of $c_{ii}$ in (3.4)), the reflection effect is large. Note that for those agents whose $c_{ii}$ the econometrician observes, the reflection effect is easily recovered once one estimates the payoff externality $\beta_0$.

Now let us turn to the case where the game is played among the first-order sophisticated players.

**Theorem 3.2.2.** Suppose that Assumptions 3.2.1 - 3.2.3 hold and that all the players are of first-order sophisticated type. Suppose further that for each $i \in N$ and $k \neq i$, $E[\eta_k | J_i] = 0$. Then each player $i$'s best linear response $s_i^{BR,FS}$ takes the following form:

\[
s_i^{BR,FS}(J_i) = \gamma_0 X_{i,1} + \varepsilon_i + \frac{\beta_0}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{jj}(\gamma_0 X_{j,1} + \varepsilon_j) + \beta_0^2 \sum_{j \in N_P(i)} \tilde{\lambda}_{ij}(\gamma_0 X_{j,1} + \varepsilon_j)
+ \delta_0 \tilde{X}_{i,2} + \delta_0' \beta_0 \sum_{j \in N_P(i)} \tilde{\lambda}_{ij} X_{j,2} + \eta_i.
\]

where,

\[
\tilde{\lambda}_{ij} = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} \hat{\lambda}_{kj} \mathbb{1}\{j \in N_P(k)\}
\]

and

\[
\hat{\lambda}_{ij} = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} \frac{r_{ik} \hat{\lambda}_{kj} \mathbb{1}\{j \in N_P(k)\}}{n_P(k)(1 - \beta_0 c_{kk})}.
\]

Note that as compared to the case of the game with agents of simple type, the game with agents of the first order sophisticated type predicts outcomes with broader network externality. For example, in contrast to the case of simple type agents, the types of neighbors whose actions do not affect player $i$’s payoff can affect his best response. More specifically, note that for $j \in N_{P2}(i) \setminus N_P(i)$,

\[
\frac{\partial s_i^{BR,FS}(J_i)}{\partial x_{j,1}} = \beta_0^2 \gamma_0 \tilde{\lambda}_{ij} \quad \text{and} \quad \frac{\partial s_i^{BR,FS}(J_i)}{\partial x_{j,2}} = \beta_0 \delta_0 \tilde{\lambda}_{ij}.
\]

This externality from player $j$ on player $i$ is strong when $c_{kj}$’s are large for many $k \in N_P(i)$.
i.e., when player $j$ has a high local centrality to a large fraction of player $i$’s $G_P$-neighbors.

### 3.2.4 The External Validity of Network Externality

Through a simple model of linear interactions, we explore two issues of external validity. The first issue is about generalizing the results that come from a model with a smaller graph to the population with a larger graph. We see how sensitively the network externality changes as the network grows. If the sensitivity is not high, this supports the external validity of a model toward a larger graph. The second issue is about misspecification of behavioral assumptions. Here we set the benchmark (true) model to be a complete information model with equilibrium strategies, but assume that the econometrician adopts our behavioral model to make the analysis tractable. Then we explore how close the network externality from the behavioral model is to the true model of complete information game. Both models assume the same payoff function and the same payoff graph. For simplicity, we remove $X_i$’s and $\eta_i$’s. The main focus here is on the stability of the prediction of the network externalities as we progressively move from a small payoff graph to a large payoff graph. Let $Y_i$ be the observed outcome of player $i$ as predicted from either of the two game models.

The complete information game model assumes that every agent observes all the types $\varepsilon_i$’s of other agents. This model yields the following equilibrium equation:

$$Y_i = \frac{\beta_0}{n_{P(i)}} \sum_{j \in N_{P(i)}} Y_j + \varepsilon_i.$$

Then the reduced form for $Y_i$’s can be written as

$$y = (I - \beta_0 A)^{-1} \varepsilon,$$

where $y = (Y_1, ..., Y_n)'$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$, and $A$ is a row-normalized adjacency matrix of the payoff graph $G_P$, i.e., the $(i,j)$-th entry of $A$ is $1/n_{P(i)}$ if $j \in N_P(i)$ and zero otherwise. Thus in the complete information equilibrium model, each $Y_i$ is a linear combination of all $\varepsilon_i$’s. The model implies that when $\beta_0$ is close to one (i.e., the local interaction becomes strong), the equilibrium outcome can exhibit extensive cross-sectional dependence.

On the other hand, our behavioral model (with specification A: $r_{ik} = 1$ in (3.1) and with

---

8 Using the explicit form of the best response operator $M$ and Condition BP, we can derive best linear responses in a game populated by agents of a higher order sophisticated type. As the sophistication of agents becomes of higher order, the network externality of each agent broadens to a wider set of agents.
Table 3.1: The Characteristics of the Payoff Graphs

<table>
<thead>
<tr>
<th></th>
<th>Erdös-Rényi</th>
<th></th>
<th>Barabási-Albert</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Network A</td>
<td>Network B</td>
<td>Network C</td>
<td>Network A</td>
</tr>
<tr>
<td>n</td>
<td>162.0</td>
<td>766.4</td>
<td>3067.4</td>
<td>432.2</td>
</tr>
<tr>
<td>$d_{mx}$</td>
<td>10.72</td>
<td>12.50</td>
<td>14.14</td>
<td>76.62</td>
</tr>
<tr>
<td>$d_{av}$</td>
<td>2.043</td>
<td>2.296</td>
<td>3.186</td>
<td>1.437</td>
</tr>
</tbody>
</table>

Notes: This table gives average characteristics of the payoff graphs, $G_P$, used in the simulation study, where the average was over 50 simulations. $d_{av}$ and $d_{mx}$ denote the average and maximum degrees of the payoff graphs.

The assumption that all the players are of simple type) predicts the outcomes in the following simple reduced form:

$$Y_i = \lambda_{ii} \left( \varepsilon_i + \frac{\beta_0}{n_{P(i)}} \sum_{j \in N_{P(i)}} \lambda_{ij} \varepsilon_j \right),$$

which comes from Theorem 3.2.1 without $X_i$’s and $\eta_i$’s. It is important to note that the two models have the same payoff with the same payoff externality parameter $\beta_0$. The only difference is the information set assumptions and the solution concepts of the game.

The parameter of interest is the average network externality:

$$\frac{1}{n} \sum_{i \in N} \frac{1}{n_{P(i)}} \sum_{j \in N_{P(i)}} \frac{\partial s_{BR}^P(i)}{\partial \varepsilon_j} = \begin{cases} 
\frac{1}{n} \sum_{i \in N} \frac{1}{n_{P(i)}} \sum_{j \in N_{P(i)}} [(I - \beta_0 A)^{-1}]_{ij}, & \text{from the equilibrium model} \\
\frac{\beta_0 \lambda_{ii}}{n} \sum_{i \in N} \frac{1}{n_{P(i)}} \sum_{j \in N_{P(i)}} \lambda_{ij}, & \text{from the behavioral model,}
\end{cases}$$

where $[(I - \beta_0 A)^{-1}]_{ij}$ denotes the $(i, j)$-th entry of the matrix $(I - \beta_0 A)^{-1}$.

Note that the network externalities depend only on $\beta_0$ and the payoff graph $G_P$. For the payoff graph $G_P$, we considered two different models for random graph generation. The first kind of random graphs are Erdős-Rényi (ER) random graph with the probability equal to $5/n$ and the second kind of random graphs are Barabási-Albert (BA) random graph such that beginning with an Erdős-Rényi random graph of size 20 with each link forming with equal probability $1/19$ and grows by including each new node with two links formed with the existing nodes with probability proportional to the degree of the nodes.
For each random graph, we first generate a random graph of size 10,000, and then construct three subgraphs $A, B, C$ such that network $A$ is a subgraph of network $B$ and the network $B$ is a subgraph of network $C$. We generate these subgraphs as follows. First, we take a subgraph $A$ to be one that consists of agents within distance $k$ from agent $i = 1$. Then network $B$ is constructed to be one that consists of the neighbors of the agents in network $A$ and network $C$ is constructed to be one that consists of the neighbors of the agents in network $B$. For an ER random graph, we took $k = 3$ and for a BA random graph, we took $k = 2$. We repeated the process 50 times to construct an average behavior of network externality as we increase the network. Table 3.1 shows the average network sizes and degree characteristics as we move from Networks A, B to C.

First, we would like to see how sensitive the predicted average network externality becomes as we move across three networks of increasing sizes. The results are in Figures 3.1 and 3.2. Figure 3.1 captures the relation between $\beta_0$ and the average network externality for the case of ER graphs and Figure 3.2 captures that for the case of BA graphs. The left panel depicts the relation from the complete information equilibrium model and the right panel depicts the relation from the behavioral model.

As shown in Figures 3.1,3.2, the predicted network externality from the behavioral model is less sensitive to the change of the networks than that from the equilibrium model. In particular, this contrast is stark when $\beta_0$ is close to 1. The main reason behind this contrast is that in the case of the equilibrium model, stronger local strategic interactions induce extensive cross-sectional dependence. This extensiveness will sensitively depend on the size and the shape of the network. On the other hand, the behavioral model limits the extent of the cross-sectional dependence even when $\beta_0$ is high. Hence the predicted network externality does not vary as much as the equilibrium model as we change the network. The result illustrates the point that our behavioral model translates local strategic interactions to local stochastic dependence of observed actions gives a better property of external validity than the complete information equilibrium model.

Suppose that the econometrician believes the true model is an equilibrium model, but uses our behavioral model as a proxy for the equilibrium model. If these two models generate “similar” predictions, using our behavioral model as a proxy will not be a bad idea. The results in Figures 3.1 and 3.2 again show that the answer depends on the payoff externality $\beta_0$. Unless the parameter $\beta_0$ is very high (say larger than or equal to 0.5), both the equilibrium approach and the behavioral approach give similar network externality. However, the discrepancy widens when $\beta_0$ is high. Hence in this set-up, using our behavioral approach as a proxy for an equilibrium approach makes sense only when strategic interdependence is
Figure 3.1: Network Externality Comparison Between Equilibrium and Behavioral Models: Erdős-Rényi Graphs

Notes: Each line gives the average network externality as a function of $\beta_0$, where the network is generated through an ER graph. The complete information game shows how the relationship between the network externality and $\beta_0$ changes as we expand the graph from a subgraph of agents within distance $k$ from the agent 1. (Networks A, B, and C correspond to networks with $k = 3, 4, 5$ from a small graph to a large one.) The figures show that the average network externality from the behavioral model behaves more stably across different networks than that from the equilibrium model in particular when $\beta_0$ (local interaction parameter) is high.

The comparison here uses a set-up where the econometrician observes all the players in the game. However, it should be kept in mind that as we shall see later when we propose inference, the behavioral model naturally accommodates the case where one observes only part of the players whereas the complete information game model does not in general. Hence when the local strategic interactions are not very high, the behavioral model can be a good proxy for a complete information game model with predictions from an equilibrium when only part of the players are observed in the sample.
Figure 3.2: Network Externality Comparison Between Equilibrium and Behavioral Models: Barabási-Albert Graphs

Notes: The figure is similar to the previous one except that the graph is now BA. The complete information game shows the relation changes as we expand the graph from a subgraph of agents within distance $k$ from the agent 1. Again, the behavioral model gives a prediction of the relation that tends to be more stable than the complete information game in this network generation.

3.3 Econometric Inference

3.3.1 General Overview

Partial Observation of Interactions

A large network data set is often obtained through a non-random sampling process. (See e.g. [110].) The main difficulty in practice is that the actual sampling process by which the network data are gathered is hard to formulate formally with accuracy. Our approach of empirical modeling can be useful in such a situation where interactions are observed only partially through a certain non-random sampling scheme that is not precisely known. In this section, we make explicit the data requirements for the econometrician and propose inference procedures. We mainly focus on the game where all the players in the game are of simple type.

Suppose that the original game of interactions consists of a large number of agents whose set we denote by $N$. Let the set of players $N$ be on a payoff graph $G_P$ and an information graph $G_I$, facing the strategic environment as described in the preceding section.
Denote the best response as an observed dependent variable \( Y_i \): for \( i \in N \),  
\[
Y_i = s^{BR}_i(\mathcal{I}_i).
\]

Let us make the following additional assumption on this original large game. Let us first define  
\[
\mathcal{F} = \sigma(X, G_P, G_I) \vee \mathcal{C},
\]
i.e., the \( \sigma \)-field generated by \( X = (X_i)_{i \in N}, G_P, G_I \) and \( \mathcal{C} \).

**Assumption 3.3.1.** (i) \( \varepsilon_i \)'s and \( \eta_i \)'s are conditionally i.i.d. across \( i \)'s given \( \mathcal{F} \).

(ii) \( \{\varepsilon_i\}_{i=1}^n \) and \( \{\eta_i\}_{i=1}^n \) are conditionally independent given \( \mathcal{F} \).

(iii) For each \( i \in N \), \( \mathbb{E}[\varepsilon_i | \mathcal{F}] = 0 \) and \( \mathbb{E}[\eta_i | \mathcal{F}] = 0 \).

The last condition (iii) excludes endogenous formation of \( G_P \) or \( G_I \), because the condition requires that the unobserved type components \( \varepsilon_i \) and \( \eta_i \) be conditionally mean independent of these graphs, given \( X = (X_i)_{i \in N} \) and \( \mathcal{C} \). However, the condition does not exclude the possibility that \( G_P \) and \( G_I \) are formed based on \( (X, \mathcal{C}) \). Hence the formation of networks by agents using information in \( X \) or \( \mathcal{C} \) is permitted.

The econometrician observes only a subset \( N^* \subset N \) of agents and part of \( G_P \) through a potentially stochastic sampling process of unknown form. We assume for simplicity that \( n^* \equiv |N^*| \) is nonstochastic. This assumption is satisfied, for example, if one collects the data for agents with predetermined sample size \( n^* \). We assume that though being a small fraction of \( N \), the set \( N^* \) is still a large set justifying our asymptotic framework that sends \( n^* \) to infinity. Most importantly, constituting only a small fraction of \( N \), the observed sample \( N^* \) of agents induces a payoff subgraph which one has no reason to view as “approximating” or “similar to” the original payoff graph \( G_P \). Let us make precise the data requirements.

**Condition A:** The stochastic elements of the sampling process are conditionally independent of \( \{(T_i', \eta_i')\}_{i \in N} \) given \( \mathcal{F} \).

**Condition B:** For each \( i \in N^* \), the econometrician observes \( N_P(i) \) and \( (Y_i, X_i) \), and for each \( j \in N_P(i) \), the econometrician observes \( |N_P(i) \cap N_P(j)|, n_P(j) \) and \( X_j \).

**Condition C:** Either of the following two conditions is satisfied:

(a) For \( i, j \in N^* \) such that \( i \neq j \), \( N_P(i) \cap N_P(j) = \emptyset \).

(b) For each agent \( i \in N^* \), and for any agent \( j \in N^* \) such that \( N_P(i) \cap N_P(j) \neq \emptyset \), the econometrician observes \( Y_j, |N_P(j) \cap N_P(k)|, n_P(k) \) and \( X_k \) for all \( k \in N_P(j) \).
Before we discuss the conditions, it is worth noting that these conditions are trivially satisfied when we observe the full payoff graph $G_P$ and $N^* = N$. Condition A is satisfied, for example, if the sampling process is based on observed characteristics $X$ and some characteristics of the strategic environment that is commonly observed by all the players. This condition is violated if the sampling is based on the outcomes $Y_i$'s or unobserved payoff-relevant signals such as $\varepsilon_i$ or $\eta_i$. Condition B essentially requires that in the data set, we observe $(Y_i, X_i)$ of many agents $i$, and for each $G_P$-neighbor $j$ of agent $i$, observe the number of the agents who are common $G_P$-neighbors of $i$ and $j$ and the size of $G_P$-neighborhood of $j$ along with the observed characteristics $X_j$.

As for a $G_P$-neighbor $j$ of agent $i \in N^*$, this condition does not require that the agent $j$'s action $Y_j$ or the full set of his $G_P$-neighbors are observed. Condition C(a) is typically satisfied when the sample of agents $N^*$ is randomly selected from a much larger set of agents so that no two agents have overlapping $G_P$-neighbors in the sample. In practice for use in inference, one can take the set $N^*$ to include only those agents that satisfy Conditions A-C as long as $N^*$ thereof is still large and the selection is based only on $(X, G_P)$. One can simply use only those agents whose $G_P$-neighborhoods are not overlapping, as long as there are many such agents in the data.

**Estimating Payoff Parameters and the Average Network Externality**

In order to introduce inference procedures for $\beta_0$ and other payoff parameters, let us define for $i \in N$,

$$Z_{i,1} = \lambda_{ii}X_{i,1} + \frac{\beta_0\lambda_{ii}}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij}X_{j,1}, \quad \text{and}$$

$$Z_{i,2} = \frac{1}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij}X_{j,2}.$$  

(Note that $Z_{i,1}$ and $Z_{i,2}$ rely on $\beta_0$ although it is suppressed from notation for simplicity as we do frequently below for other quantities.) By Theorem 3.2.1, we can write

$$Y_i = Z_{i,1}'\gamma_0 + Z_{i,2}'\delta_0 + v_i,$$

Note that this condition is violated when the neighborhoods are top-coded in practice. For example, the maximum number of friends in the survey for a peer effects study can be set to be lower than the actual number of friends for many students. The impact of this top-coding upon the inference procedure is an interesting question on its own.

This random selection does not need to be a random sampling from the population of agents. Note that the random sampling is extremely hard to implement in practice in this situation, because one needs to use the equal probability for selecting each agent into the collection $N^*$, but this equal probability will be feasible only when one has at least the catalog of the entire population $N$.  

63
where

\[ v_i = \lambda_{ii} \varepsilon_i + \beta_0 \lambda_{ii} \sum_{j \in N_P(i)} \lambda_{ij} \varepsilon_j + \eta_i. \]

Note that the observed actions \( Y_i \) are cross-sectionally dependent (conditional on \( X_i \)'s) due to information sharing on unobservables \( \varepsilon_i \). However, since only the types of \( G_P \)-neighbors turn out to be relevant in the best linear response, the correlation between \( Y_i \) and \( Y_j \) is non-zero only when agents \( i \) and \( j \) are \( G_P \)-neighbors.

We define \( Z_i = [Z'_i, 1] \in \mathbb{R}^{d_1 + d_2} \) and \( \rho_0 = [\gamma_0', \delta_0']' \in \mathbb{R}^{d_1 + d_2} \), where \( X_{i,1} \in \mathbb{R}^{d_1} \) and \( X_{i,2} \in \mathbb{R}^{d_2} \), so that we can rewrite the linear model as

\[ Y_i = Z'_i \rho_0 + v_i. \]

Suppose that \( \varphi_i \) is an \( M \times 1 \) vector of instrumental variables (which potentially depend on \( \beta_0 \)) with \( M > d \equiv d_{x_1} + d_{x_2} \) such that for all \( i \in N \),

\[ \mathbb{E}[v_i \varphi_i] = 0. \]

Note that the orthogonality condition above holds for any \( \varphi_i \) as long as for each \( i \in N \), \( \varphi_i \) is \( \mathcal{F} \)-measurable, i.e., once \( \mathcal{F} \) is realized, there is no extra randomness in \( \varphi_i \). This is the case, for example, when \( \varphi_i \) is a function of \( X = (X_i)_{i \in N} \). We also allow that each \( \varphi_i \) depends on \( \beta_0 \).

While the asymptotic validity of our inference procedure admits a wide range of choices for \( \varphi_i \)'s, one needs to choose them with care to obtain sharp inference on the payoff parameters. Especially, it is important to consider instrumental variables which involve the characteristics of \( G_P \)-neighbors to obtain a sharp inference on payoff externality parameter \( \beta_0 \). This is because the cross-sectional dependence of observations carries substantial information for estimating strategic interdependence among agents.

The moment function is nonlinear in the payoff externality \( \beta_0 \) and it is not easy to ensure that these moment conditions uniquely determine the true parameter vector even in the limit as \( n^* \) goes to infinity.\(^{11}\) In this chapter, we adopt a Bonferroni procedure in which we first obtain a confidence interval for \( \beta_0 \) and, using this, we perform inference on \( \rho_0 \). This approach works well even when \( \beta_0 \) is not consistently estimable.

\(^{11}\)One might consider following the nonlinear iterated least squares approach of [27]. However, it is not clear in our context whether the parameter \( \beta_0 \) is consistently estimable across various payoff graph configurations as \( n^* \) diverges to infinity. Thus, we consider a Bonferroni approach.
We proceed first to estimate $\rho_0$ assuming knowledge of $\beta_0$. Define

$$S_{\phi\phi} = \phi' \phi/n^*, \text{ and } \hat{\phi} = \phi S_{\phi\phi}^{-1/2}$$

where $\phi$ is an $n^* \times M$ matrix whose $i$-th row is given by $\phi_i', i \in N^*$. Define

$$\Lambda = \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N^*} \mathbb{E}[v_iv_j|\mathcal{F}] \hat{\phi}_i \hat{\phi}_j', \quad (3.7)$$

and let $\hat{\Lambda}$ be a consistent estimator of $\Lambda$. (We will explain how we construct this estimator later.) Define

$$\Lambda = \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N^*} \mathbb{E}[v_iv_j|\mathcal{F}] \hat{\phi}_i \hat{\phi}_j', \quad (3.7)$$

and let $\hat{\Lambda}$ be a consistent estimator of $\Lambda$. (We will explain how we construct this estimator later.) Define

$$S_{Z\phi} = Z' \hat{\phi}/n^*, \text{ and } S_{\phi y} = \phi' y/n^*,$$

where $Z$ is an $n^* \times d$ matrix whose $i$-th row is given by $Z_i'$ and $y$ is an $n^* \times 1$ vector whose $i$-th entry is given by $Y_i, i \in N^*$. Since (from the fact that $G_P$ is undirected)

$$c_{ij} = \frac{|N_P(i) \cap N_P(j)|}{n_P(i)},$$

we can construct $Z_i$ for each $i \in N^*$ from the data satisfying Conditions A-C. Then we estimate

$$\hat{\rho} = \left[ S_{Z\phi} \hat{\Lambda}^{-1} S_{Z\phi}' \right]^{-1} S_{Z\phi} \hat{\Lambda}^{-1} S_{\phi y}. \quad (3.8)$$

Using this estimator, we construct a vector of residuals $\hat{v} = [\hat{v}_i]_{i \in N^*}$, where

$$\hat{v}_i = Y_i - Z_i' \hat{\rho}. \quad (3.9)$$

Finally, we form a test statistic as follows:

$$T(\beta_0) = \frac{\hat{v}' \hat{\phi} \hat{\Lambda}^{-1} \phi' \hat{v}}{n^*}, \quad (3.10)$$

making it explicit that the test statistic depends on $\beta_0$. Later we show that

$$T(\beta_0) \rightarrow_d \chi^2_{M-d}, \text{ as } n^* \rightarrow \infty,$$

where $\chi^2_{M-d}$ denotes the $\chi^2$ distribution with degree of freedom $M - d$. Let $C^\beta_{1-(\alpha/2)}$ be the
\[(1 - \alpha/2)100\% \text{ confidence set for } \beta_0 \text{ defined as} \]
\[C^{\beta}_{1-(\alpha/2)} \equiv \{ \beta \in (-1,1) : T(\beta) \leq c_{1-(\alpha/2)} \},\]
where \(T(\beta)\) is computed as \(T(\beta_0)\) with \(\beta_0\) replaced by \(\beta\) and the critical value \(c_{1-(\alpha/2)}\) is the \((1 - \alpha/2)\)-quantile of \(\chi^2_{M-d}\).

Then we establish that under regularity conditions,
\[\sqrt{n^*} \hat{\nu}^{-1/2}(\hat{\rho} - \rho_0) \to_d N(0, I),\]
as \(n^* \to \infty\), where
\[\hat{\nu} = \left[ S \hat{\rho} \Lambda^{-1} S' \right]^{-1}.\]
(See Section 3.3.2 below for conditions and formal results.) Using this estimator \(\hat{\rho}\), we can construct a \((1 - \alpha)100\%\) confidence interval for \(a' \rho_0\) for any non-zero vector \(a\). For this define
\[\hat{\sigma}^2(a) = a' \hat{\nu} a.\]

Let \(c_{1-(\alpha/4)}^a\) be the \((1 - \alpha/4)\)-percentile of \(N(0, 1)\). Define for a vector \(a\) with the same dimension as \(\rho\),
\[C^{\rho}_{1-(\alpha/2)}(\beta_0, a) = \left[ a' \hat{\rho} - \frac{c_{1-(\alpha/4)}^a \hat{\sigma}(a)}{\sqrt{n}}, a' \hat{\rho} + \frac{c_{1-(\alpha/4)}^a \hat{\sigma}(a)}{\sqrt{n}} \right].\]

Then the confidence set for \(a' \rho\) is given by
\[C^{\rho}_{1-\alpha}(a) = \bigcup_{\beta \in C^{\beta}_{1-(\alpha/2)}} C^{\rho}_{1-(\alpha/2)}(\beta, a).\]

Notice that since \(\beta\) runs in \((-1, 1)\) and the estimator \(\hat{\rho}\) has an explicit form, the confidence interval is not computationally costly to construct in general.

Often the eventual parameter of interest is one that captures how strongly the agents’ decisions are inter-dependent through the network. Here let us introduce parameters representing the sensitivity. Let \(s^{BR}(\mathcal{I}_i)\) be the best linear response of agent \(i\) having information set \(\mathcal{I}_i\). Let us define the average network externality with respect to variable \(X_{i,1,r}\) (where
\(X_{i,1,r}\) represents the \(r\)-th entry of \(X_{i,1}\) to be
\[
\theta_1(\beta_0, \gamma_{0,r}) = \frac{1}{n^*} \sum_{i \in N^*} \frac{1}{n_P(i)} \sum_{j \in n_P(i)} \frac{\partial s_i^{BR}(x_i)}{\partial x_{j,1,r}} 
\]
\[
= \frac{1}{n^*} \sum_{i \in N^*} \frac{1}{n_P(i)} \sum_{j \in n_P(i)} \frac{\beta_0 r_{ij}}{n_P(i)(1 - \beta_0 c_{ii})(1 - \beta_0 c_{ij})} \gamma_{0,r},
\]
where \(\gamma_{0,r}\) denotes the \(r\)-th entry of \(\gamma_0\). See (3.5). Thus the confidence interval for \(\theta_1(\beta_0, \gamma_0)\) can be constructed from the confidence interval for \(\beta_0\) and \(\gamma_0\) as follows:
\[
C^\theta_{1-\alpha} = \left\{ \theta_1(\beta, \gamma_r) : \beta \in C^\beta_{1-\alpha}, \text{ and } \gamma_r \in C^{\gamma_r}_{1-\alpha} \right\},
\]
where \(C^\beta_{1-\alpha}\) denotes the confidence interval for \(\gamma_{0,r}\). We can define similarly the average network externality with respect to an entry of \(X_{i,2}\) and construct a confidence interval for it. Details are omitted.

**Downweighting Players with High Degree Centrality**

When there are players who are linked to many other players in \(G_P\), the graph \(G_P\) tends to be denser, making it difficult to obtain good variance estimators that perform stably in finite samples. To remedy this situation, this chapter proposes a downweighting of those players with high degree centrality in \(G_P\). More specifically, in choosing an instrument vector \(\varphi_i\), we may consider the following:
\[
\varphi_i(X) = \frac{1}{\sqrt{n_P(i)}} g_i(X), \quad (3.11)
\]
where \(g_i(X)\) is a function of \(X\). This choice of \(\varphi_i\) downweights players \(i\) who have a large \(G_P\)-neighborhood. Thus we rely less on the variations of the characteristics of those players who have many neighbors in \(G_P\).

Taking downweighting agents too heavily may hurt the power of the inference because the actions of agents with high centrality contain information about the parameter of interest through the moment restrictions. On the other hand downweighting them too lightly will hurt the finite sample stability of the inference due to strong cross-sectional dependence they cause to the observations. Since a model with agents of higher order sophisticated type results in observations with more extensive cross-sectional dependence, the role of downweighting can be prominent in securing finite sample stability in such a model.
Comparison with Linear-in-Means Models

One of the most frequently used interaction models in the econometrics literature is a linear-in-means model specified as follows:

\[ Y_i = X_i^' \gamma_0 + \bar{X}_i^' \delta_0 + \beta_0 \mu_i^e(\bar{Y}_i) + v_i, \]  

(3.12)

where \( \mu_i^e(\bar{Y}_i) \) denotes the player \( i \)'s expectation of \( Y_i \), and

\[ \bar{Y}_i = \frac{1}{nP(i)} \sum_{i \in N_P(i)} Y_i \text{ and } \bar{X}_i = \frac{1}{nP(i)} \sum_{i \in N_P(i)} X_i. \]

The literature assumes rational expectations by equating \( \mu_i^e(\bar{Y}_i) \) to \( \mathbb{E}[Y_i|I_i] \), and then proceeds to identification analysis of parameters \( \gamma_0, \delta_0 \) and \( \beta_0 \). For actual inference, one needs to use an estimated version of \( \mathbb{E}[Y_i|I_i] \). One standard way in the literature is to replace it by \( \bar{Y}_i \) so that we have

\[ Y_i = X_i^' \gamma_0 + \bar{X}_i^' \delta_0 + \beta_0 \bar{Y}_i + \bar{v}_i, \]

where \( \bar{v}_i \) is an error term defined as \( \bar{v}_i = \beta_0 (\mathbb{E}[\bar{Y}_i|I_i] - \bar{Y}_i) + v_i \). The complexity arises due to the presence of \( \bar{Y}_i \) which is an endogeneous variable that is involved in the error term \( \bar{v}_i \).

One of the frequently used approaches is to use instrumental variables. There are two types of instrumental variables. The first kind is a peers-of-peers type instrumental variable which is based on the observed characteristics of the neighbors of the neighbors. This strategy was proposed by [102], [30] and [55]. The second kind of an instrumental variable is based on observed characteristics excluded from the group characteristics as instrumental variables. (See [33] and [63].) However, finding such an instrumental variable in practice is not always a straightforward task in empirical research.

Our approach of empirical modeling is different in several aspects. Our modeling uses behavioral assumptions instead of rational expectations, and produces a reduced form for observed actions \( Y_i \) from using best linear responses. This reduced form gives a rich set of testable implications and makes explicit the source of cross-sectional dependence in relation to the payoff graph. Our approach permits any nontrivial functions of \( \mathcal{F} \) as instrumental variables at least for the validity of the inference. Furthermore, one does not need to observe many independent interactions for inference.

\[ \text{[12] A similar observation applies in the case of a complete information version of the model, where one directly uses } \bar{Y}_i \text{ in place of } \mu_i^e(\bar{Y}_i) \text{ in (3.12). Still due to simultaneity of the equations, } \bar{Y}_i \text{ necessarily involve error terms } v_i \text{ not only of agent } i \text{'s own but other agents' as well.} \]
Estimation of Asymptotic Covariance Matrix

The inference requires an estimator of $\hat{V}$. First, let us find the population version of $\hat{V}$. After some algebra, it is not hard to see that the population version (conditional on $\mathcal{F}$) of $\hat{V}$ is given by

$$ V = \left[ S_{\hat{Z}\Phi} \Lambda^{-1} S_{\hat{Z}\phi} \right]^{-1}. $$

(3.13)

For estimation, it suffices to estimate $\Lambda$ defined in (3.7). For this, we need to incorporate the cross-sectional dependence of the residuals $v_i$ properly. From the definition of $v_i$, it turns out that $v_i$ and $v_j$ can be correlated if $i$ and $j$ are connected indirectly through two edges in $G_P$. However, constructing an estimator of $\Lambda$ simply by imposing this dependence structure and replacing $v_i$ by $\hat{v}_i$ can result in a conservative estimator with unstable finite sample properties, especially when each player has many players connected through two edges. Instead, we propose an alternative estimator of $\Lambda$ as follows. This estimator is found to work well in our simulation studies. We present the explanation and construction of this estimator in the Appendix.

To obtain an estimator $\hat{\Lambda}$ of $\Lambda$ (up to $\beta_0$), we first obtain a first-step estimator of $\rho$ as follows:

$$ \tilde{\rho} = \left[ S_{\hat{Z}\Phi} S_{\hat{Z}\phi} \right]^{-1} S_{\hat{Z}\Phi} S_{\Phi} S_{\Phi}. $$

(3.14)

Using this estimator, we construct a vector of residuals $\tilde{v} = [\tilde{v}_i]_{i \in N^*}$, where

$$ \tilde{v}_i = Y_i - Z_i' \tilde{\rho}. $$

(3.15)

Then we estimate$^{13}$

$$ \hat{\Lambda} = \hat{\Lambda}_1 + \hat{\Lambda}_2, $$

where

$$ \hat{\Lambda}_1 = \frac{1}{n^*} \sum_{i \in N^*} \tilde{v}_i \tilde{\phi}_i \tilde{\phi}_i', $$

and

$$ \hat{\Lambda}_2 = \frac{s_{\epsilon}}{n^*} \sum_{i \in N^*} \sum_{j \in N^* \cap N_P(i)} \sum_{j \in N_P(j) \cap N_P(i)} q_{i,j} \tilde{\phi}_i \tilde{\phi}_j', $$

$^{13}$Under Condition C(a) for sample $N^*$, we have $\Lambda_2 = 0$ because the second sum in the expression for $\lambda_2$ is empty. Hence in this case, we can simply set $\Lambda_2 = 0.$
and

\[ \hat{s}_e = \frac{\sum_{i \in N} \sum_{j \in N \cap N^*} \bar{v}_i \bar{v}_j}{\sum_{i \in N} \sum_{j \in N \cap N^*} q_{e,ij}}, \]

(3.16)

\[ q_{e,ij} = \frac{\lambda_{ii} \lambda_{jj}}{n_p(j)} + \frac{\lambda_{ij} \lambda_{jj}}{n_p(i)} + \frac{\beta_0 \lambda_{ij} \lambda_{jj}}{n_p(i) n_p(j)} \sum_{k \in N_p(i) \cap N_p(j)} \lambda_{ik} \lambda_{jk}. \]

(Note that the quantity \( q_{e,ij} \) can be evaluated once \( \beta_0 \) is fixed.) Using \( \hat{\Lambda} \), we construct the estimator for the covariance matrix \( V \), i.e.,

\[ \hat{V} = [S Z \hat{\phi}^{-1} S Z \hat{\phi}]^{-1}. \]

(3.17)

Later we provide conditions for the estimator to be consistent for \( V \).

### 3.3.2 Asymptotic Theory

In this section, we present the assumptions and formal results of asymptotic inference. We introduce some technical conditions.

**Assumption 3.3.2.** There exists \( c > 0 \) such that for all \( n^* \geq 1 \), \( \lambda_{\min}(S_{\phi \phi}) \geq c \), \( \lambda_{\min}(S_{\theta \phi} S_{\theta \phi}') \geq c \), \( \lambda_{\min}(S_{\theta \theta} \hat{\Lambda}^{-1} S_{\theta \theta}') \geq c \), \( \lambda_{\min}(\Lambda) \geq c \), and

\[ \frac{1}{n^*} \sum_{i \in N^*} \frac{\lambda_{ii}}{n_p(i)} \sum_{j \in N \cap N^*} \lambda_{ij} > c, \]

where \( \lambda_{\min}(A) \) for a symmetric matrix \( A \) denotes the minimum eigenvalue of \( A \).

**Assumption 3.3.3.** There exists a constant \( C > 0 \) such that for all \( n^* \geq 1 \),

\[ \max_{i \in N^*} ||X_i|| + \max_{i \in N} ||\phi_i|| \leq C \]

\[ 14 \] In finite samples, \( \hat{V} \) is not guaranteed to be positive definite. We can modify the estimator by using spectral decomposition similarly as in [40]. More specifically, we first take a spectral decomposition \( \hat{V} = \hat{\Lambda} \hat{B} \hat{B}' \), where \( \hat{\Lambda} \) is a diagonal matrix of eigenvalues \( \hat{\lambda}_j \) of \( \hat{V} \). We replace each \( \hat{\lambda}_j \) by the maximum between \( \hat{\lambda}_j \) and some small number \( c > 0 \) to construct \( \hat{\Lambda}_c \). Then the modified version \( \hat{V} \equiv \hat{B} \hat{\Lambda}_c \hat{B}' \) is positive definite. For \( c > 0 \), taking \( c = 0.005 \) seems to work well in the simulation studies.
and $E[\varepsilon_i^2|\mathcal{F}] + E[\eta_i^2|\mathcal{F}] < C$, where $n^0 = |N^0|$ and

$$N^0 = \bigcup_{i \in N^*} N_P(i).$$

Assumption 3.3.2 is used to ensure that the asymptotic distribution is nondegenerate. This regularity condition is reasonable, because an asymptotic scheme that gives a degenerate distribution would not be adequate to derive a finite sample, nondegenerate distribution of an estimator. Assumption 3.3.3 can be weakened at the expense of complexity in the conditions and the proofs.

We introduce an assumption which requires the payoff graph to have a bounded degree over $i$ in the observed sample $N^*$.

**Assumption 3.3.4.** There exists $C > 0$ such that for all $n^* \geq 1$,

$$\max_{i \in N^*} |N_P(i)| \leq C.$$

We may relax the assumption to a weaker, yet more complex condition at the expense of longer proofs, but in our view, this relaxation does not give additional insights. When $N^*$ is large, one can remove very high-degree nodes to obtain a stable inference. As such removal is solely based on the payoff graph $G_P$, the removal does not lead to any violation of the conditions given in this chapter.

The following theorem establishes the asymptotic validity of the inference based on the best linear responses in Theorem 3.2.1. The proof is found in the Appendix.

**Theorem 3.3.1.** Suppose that the conditions of Theorem 3.2.1 and Assumptions 3.3.1 - 3.3.4 hold. Then,

$$T(\beta_0) \rightsquigarrow \chi^2_{M-d}, \text{ and } \hat{V}^{-1/2} \sqrt{n^*} (\hat{\rho} - \rho_0) \rightsquigarrow N(0, I),$$

as $n^* \to \infty$.

### 3.4 A Monte Carlo Simulation Study

In this section, we investigate the finite sample properties of the asymptotic inference across various configurations of the payoff graph, $G_P$. The payoff graphs are generated according to two models of random graph formation, which we call Specifications 1 and 2. Specification 1 uses the Barabási-Albert model of preferential attachment, with $m$ representing the
number of edges each new node forms with existing nodes. The number $m$ is chosen from $\{1, 2, 3\}$. Specification 2 is the Erdős-Rényi random graph with probability $p = \lambda/n$, where $\lambda$ is also chosen from $\{1, 2, 3\}$\textsuperscript{15}.

In the first table, we report degree characteristics of the payoff graphs used in the simulation study.

For the simulations, we also set the following: $\rho_0 = (\gamma_0', \delta_0')'$, with $\gamma_0 = (2, 4, 1)'$, and $\delta_0 = (3, 4)'$. We choose $a$ to be a vector of ones so that $a'\rho_0 = 14$. The variables $\varepsilon$ and $\eta$ are drawn i.i.d. from $N(0, 1)$. The first column of $X_{i,1}$ is a column of ones, while the remaining columns of $X_{i,1}$ are drawn independently from $N(1, 1)$. The columns of $X_{i,2}$ are drawn independently from $N(3, 1)$.

For instruments, we consider the following nonlinear transformations of $X_1$ and $X_2$:

$$\varphi_i = [\tilde{Z}_{i,1}, X_{i,1}^2, X_{i,2}^2, X_{i,2}^3]'$$

where we define

$$\tilde{Z}_{i,1} = \frac{1}{n\rho(i)} \sum_{j \in N(i)} \lambda_{ij} \lambda_{jj} X_{j,1}.$$  

We generate $Y_i$ from the best response function in Theorem 3.2.1. While the instruments $X_{i,1}^2, X_{i,2}^2, X_{i,2}^3$ capture the nonlinear impact of $X_i$'s, the instrument $\tilde{Z}_{i,1}$ captures the cross-sectional dependence along the payoff graph. The use of this instrumental variable is crucial in obtaining a sharp inference for $\beta_0$. Note that since we have already concentrated out $\rho$ in forming the moment conditions, we cannot use linear combinations of $X_{i,1}$ and $X_{i,2}$ as our instrumental variables. The nominal size in all the experiments is set at $\alpha = 0.05$.

Overall, the simulation results illustrate the good power and size properties for the asymptotic inference on $\beta_0$ and $a'\rho_0$. As expected, the average length of the confidence intervals for both $\beta_0$ and $a'\rho_0$ become shorter as the sample size increases. We find that the confidence interval for $\beta_0$ exhibits empirical coverage close to the 95% nominal level, while the confidence interval for $a'\rho_0$ is somewhat conservative. This conservativeness is expected, given the fact that the interval is constructed using a Bonferroni approach.

\textsuperscript{15}Note that in Specification 1, the Barabási-Albert graph is generated with an Erdős-Rényi seed graph, where the number of nodes in the seed is set to equal the smallest integer above $5\sqrt{n}$. All graphs in the simulation study are undirected.
Table 3.2: The Degree Characteristics of the Graphs Used in the Simulation Study

<table>
<thead>
<tr>
<th>n</th>
<th>Specification 1</th>
<th>Specification 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m = 1</td>
<td>m = 2</td>
</tr>
<tr>
<td>500</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>1.7600</td>
<td>3.2980</td>
</tr>
<tr>
<td>1000</td>
<td>18</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>1.8460</td>
<td>3.5240</td>
</tr>
<tr>
<td>5000</td>
<td>32</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>1.9308</td>
<td>3.7884</td>
</tr>
</tbody>
</table>

Notes: This table gives characteristics of the payoff graphs, $G_P$, used in the simulation study. $d_{av}$ and $d_{mx}$ denote the average and maximum degrees of the payoff graphs.

3.5 Empirical Application: State Presence across Municipalities

3.5.1 Motivation and Background

State capacity (i.e., the capability of a country to provide public goods, basic services, and the rule of law) can be limited for various reasons. (See e.g. [24] and [76]). A “weak state” may arise due to political corruption and clientelism, and result in spending inadequately on public goods ([3]), accommodating armed opponents of the government ([151]), and war ([125]). Empirical evidence has shown how these weak states can persist from precolonial times. Higher state capacities seem related to the current level prosperity at the ethnic and national levels ([75] and [136]).

Our empirical application is based on a recent study by [4] who investigate the local choices of state capacity in Colombia, using a model of a complete information game on an exogenously formed network. In their set-up, municipalities choose a level of spending on public goods and state presence (as measured by either the number of state employees or state agencies). There is network externality in a municipality’s choice because municipalities that are adjacent to each other can benefit from their neighbors’ choices of public goods provisions, such as increased security, infrastructure and bureaucratic connections. Thus, a municipality’s choice of state capacity can be thought of as a strategic decision on a geographic network.

It is not obvious that public good provision in one municipality leads to higher spending on public goods in neighboring municipalities. Some neighbors may free-ride and under-invest in state presence if they anticipate others will invest highly. Rent-seeking by munici-
Table 3.3: The Empirical Coverage Probability and Average Length of Confidence Intervals for $\beta_0$ at 95% Nominal Level.

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>Specification 1</th>
<th>Specification 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 1$</td>
<td>$m = 2$</td>
</tr>
<tr>
<td>$-0.5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.9760</td>
<td>0.9715</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.9665</td>
<td>0.9720</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>0.9745</td>
<td>0.9705</td>
</tr>
<tr>
<td>$-0.3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.9730</td>
<td>0.9690</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.9665</td>
<td>0.9710</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>0.9700</td>
<td>0.9670</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.9610</td>
<td>0.9660</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.9640</td>
<td>0.9690</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>0.9685</td>
<td>0.9670</td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.9640</td>
<td>0.9665</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.9670</td>
<td>0.9695</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>0.9705</td>
<td>0.9660</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.9725</td>
<td>0.9735</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.9690</td>
<td>0.9705</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>0.9720</td>
<td>0.9695</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>Specification 1</th>
<th>Specification 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 1$</td>
<td>$m = 2$</td>
</tr>
<tr>
<td>$-0.5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.2735</td>
<td>0.3069</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.2264</td>
<td>0.2513</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>0.1771</td>
<td>0.1783</td>
</tr>
<tr>
<td>$-0.3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.2541</td>
<td>0.2971</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.2132</td>
<td>0.2445</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>0.1710</td>
<td>0.1739</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.2434</td>
<td>0.2881</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.2051</td>
<td>0.2377</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>0.1639</td>
<td>0.1715</td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.2481</td>
<td>0.2891</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.2114</td>
<td>0.2377</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>0.1657</td>
<td>0.1729</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.2613</td>
<td>0.2974</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.2193</td>
<td>0.2429</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>0.1703</td>
<td>0.1770</td>
</tr>
</tbody>
</table>

Notes: The first half of the table reports the empirical coverage probability of the asymptotic confidence interval for $\beta_0$ and the second half reports its average length. The simulated rejection probability at the true parameter is close to the nominal size of $\alpha = 0.05$ and the average lengths decrease with $n$. The simulation number is $R = 2000$. 

74
Table 3.4: The Empirical Coverage Probability and Average Length of Confidence Intervals for $d' \rho_0$ at 95% Nominal Level.

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>Specification 1</th>
<th>Specification 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 1$</td>
<td>$m = 2$</td>
</tr>
<tr>
<td>$-0.5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.9945</td>
<td>0.9955</td>
</tr>
<tr>
<td></td>
<td>0.9925</td>
<td>0.9930</td>
</tr>
<tr>
<td></td>
<td>0.9850</td>
<td>0.9875</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9935</td>
<td>0.9960</td>
</tr>
<tr>
<td></td>
<td>0.9905</td>
<td>0.9935</td>
</tr>
<tr>
<td></td>
<td>0.9795</td>
<td>0.9880</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9935</td>
<td>0.9965</td>
</tr>
<tr>
<td></td>
<td>0.9925</td>
<td>0.9935</td>
</tr>
<tr>
<td></td>
<td>0.9785</td>
<td>0.9875</td>
</tr>
</tbody>
</table>

| $0$      |         |         |         |         |         |         |
| $n = 500$| 0.9950  | 0.9935  | 0.9935  | 0.9940  | 0.9915  | 0.9925  |
|          | 0.9925  | 0.9935  | 0.9945  | 0.9960  | 0.9970  | 0.9920  |
|          | 0.9790  | 0.9895  | 0.9925  | 0.9970  | 0.9860  | 0.9850  |
| $n = 1000$|         |         |         |         |         |         |
|          | 0.9940  | 0.9940  | 0.9940  | 0.9920  | 0.9965  | 0.9925  |
|          | 0.9790  | 0.9895  | 0.9925  | 0.9970  | 0.9860  | 0.9850  |
| $n = 5000$|         |         |         |         |         |         |
|          | 0.9935  | 0.9925  | 0.9935  | 0.9940  | 0.9915  | 0.9920  |
|          | 0.9945  | 0.9940  | 0.9945  | 0.9980  | 0.9930  | 0.9930  |
|          | 0.9835  | 0.9900  | 0.9910  | 0.9645  | 0.9850  | 0.9860  |

| $0.3$    |         |         |         |         |         |         |
| $n = 500$| 1.5337  | 1.6361  | 1.7225  | 1.4037  | 1.4196  | 1.7567  |
|          | 1.3263  | 1.4140  | 1.4466  | 1.2028  | 1.1714  | 1.3966  |
|          | 0.9896  | 0.9749  | 1.0145  | 0.9656  | 0.8244  | 0.8650  |
| $n = 1000$|         |         |         |         |         |         |
|          | 1.5068  | 1.6007  | 1.6761  | 1.5646  | 1.4632  | 1.6976  |
|          | 1.3060  | 1.3607  | 1.4031  | 1.3721  | 1.2259  | 1.3798  |
|          | 0.9840  | 0.9486  | 0.9873  | 1.1527  | 0.9154  | 0.8922  |
| $n = 5000$|         |         |         |         |         |         |
|          | 1.5516  | 1.6019  | 1.6501  | 1.8290  | 1.5653  | 1.6869  |
|          | 1.3416  | 1.3257  | 1.3754  | 1.6133  | 1.3204  | 1.3959  |
|          | 1.0066  | 0.9412  | 0.9690  | 1.3938  | 1.0199  | 0.9213  |
| $0.5$    |         |         |         |         |         |         |
| $n = 500$| 1.6553  | 1.6353  | 1.6567  | 2.1022  | 1.6772  | 1.7101  |
|          | 1.4069  | 1.3146  | 1.3731  | 1.8362  | 1.4272  | 1.4376  |
|          | 1.0552  | 0.9420  | 0.9542  | 1.5865  | 1.1029  | 0.9593  |
| $n = 1000$|         |         |         |         |         |         |
|          | 1.5516  | 1.6019  | 1.6501  | 1.8290  | 1.5653  | 1.6869  |
|          | 1.3416  | 1.3257  | 1.3754  | 1.6133  | 1.3204  | 1.3959  |
|          | 1.0066  | 0.9412  | 0.9690  | 1.3938  | 1.0199  | 0.9213  |
| $n = 5000$|         |         |         |         |         |         |
|          | 1.6553  | 1.6353  | 1.6567  | 2.1022  | 1.6772  | 1.7101  |
|          | 1.4069  | 1.3146  | 1.3731  | 1.8362  | 1.4272  | 1.4376  |
|          | 1.0552  | 0.9420  | 0.9542  | 1.5865  | 1.1029  | 0.9593  |

Notes: The true $d' \rho_0$ is equal to 14. The first half of the table reports the empirical coverage probability of the asymptotic confidence interval and the second half its average length for $d' \rho_0$. The empirical coverage probability of the confidence interval for $d' \rho_0$ is generally conservative which is expected from the use of the Bonferroni approach. Nevertheless, the length of the confidence interval is reasonably small. The simulation number, $R$, is 2000.
Figure 3.3: Degree Distribution of $G_P$

Notes: The figure presents the degree distribution of the graph $G_P$ used in the empirical specification. The average degree is 5.48, the maximum degree is 20, and the minimum degree is 1.

Pal politicians would also limit the provision of public goods. On the other hand, economies of scale could yield complementarities in state presence across neighboring municipalities.

In our study, we extend the model in [4] to an incomplete information game where information may be shared across municipalities. In particular, we do not assume that all municipalities know and observe all characteristics and decisions of the others. It seems reasonable that the decisions made across the country may not be observed or well known by those municipalities that are geographically remote.

3.5.2 Empirical Set-up

Let $y_i$ denote the state capacity in municipality $i$ (as measured by the (log) number of public employees in municipality $i$) and $G_P$ denote the geographic network, where an edge is defined on two municipalities that are geographically adjacent. We assume that $G_P$ is exogenously formed. The degree distribution of $G_P$ is shown in Figure 3.3. We study the optimal choice of $y_i$, where $y_i$ leads to a larger prosperity $p_i$. Prosperity in municipality $i$ is modeled as:

$$p_i = (\beta \bar{y} + x_{1i} \gamma + \eta_i + e_i + \xi_i^D) y_i,$$

(3.18)

This corresponds to the case in of $\delta_1 = \delta_2 = 0$ in [4].
where $\zeta^D$ is a district specific dummy variable, $\epsilon_i$ and $\eta_i$ are our sharable and non-sharable private information, and $\bar{y} = \frac{1}{n_{e(i)}} \sum_{j \in N_{e(i)}} y_j$. The term $x_{1,i}$ represents municipality characteristics. These include geographic characteristics, such as land quality, altitude, latitude, rainfall; and municipal characteristics, such as distance to highways, distance to royal roads and Colonial State Presence.\footnote{Note that, here we take $x_{2,i} = 0$. This is done for a closer correspondence to the specification in [4]. Finally, note that $p_i$ is only a function of terms are multiplied by $y_i$. This is a simplification from their specification. We do so because we will focus on the best response equation. The best response equation, derived from the first order condition to this problem, would not include any term that is not a function of $y_i$ itself.}

The welfare of a municipality is given by

$$u_i(y_i, y_{-i}, T, \eta_i) = p_i(y_i, \bar{y}, T, \eta_i) - \frac{1}{2} y_i^2,$$  

(3.19)

where the second term refers to the cost of higher state presence, and the first term is the prosperity $p_i$.

We can rewrite the welfare of the municipality by substituting (3.18) into (3.19):

$$u_i(y_i, y_{-i}, T, \eta_i) = (\beta \bar{y} + x_{1,i} + \eta_i + \epsilon_i + \zeta^D) y_i - \frac{1}{2} y_i^2,$$  

(3.20)

which is our model from the previous section. We assume that municipalities (or the mayor in charge), wishes to maximize welfare by choosing state presence, given their beliefs about the types of the other municipalities.

In our specification, we allow for incomplete information. This is reflected in the terms $\epsilon_i$, $\eta_i$, which will be present in the best response function. The municipality, when choosing state presence $y_i$, will be able to observe $\epsilon_i$ of its neighbors and will use its beliefs over the types of the others to generate its best response. The best response will follow the results from Theorem 3.2.1.

### 3.5.3 Model Specification

We follow closely Table 3 in [4] for the choice of specifications and variables. Throughout the specifications, we include longitude, latitude, surface area, elevation, annual rainfall, department fixed effects and a department capital dummy (all in $X_1$). We further consider the effect of variables distance to current highways, land quality and presence of rivers in the municipality.

For the choice of instruments, we consider two separate types of instruments. The first is
the sum of neighbor values (across \( G_P \)) of the historical variables (denoted as \( C_i \)).\(^\text{18}\) The historical variables used are Total Crown Employees (also called Colonial State Officials), Distance to Royal Roads, Colonial State Agencies and Historical Population, as well as Colonial State Presence Index squared and Distance to Royal Roads squared. The later two add additional power to the inference. We also use the variable \( \tilde{Z}_i = n_P(i)^{-1} \sum_{j \in N_P(i)} \lambda_{ij} X_{j,1} \) as part of the instrumental variable, which was shown to perform well in the Monte Carlo Simulations in Section 3.4. This variable captures cross sectional dependence as a crucial source of variation for inference on the strategic interactions. We use downweighting of our instruments as explained in a preceding section.

### 3.5.4 Results

The results across a range of specifications are presented in Table 3.5. In these results, we see that the effect is statistically different than 0 and stable across specifications. It indicates that there is complementarity in the provision of public goods and state presence (\( \beta > 0 \)).

Let us compare our results to those in [4]. There, the authors report the average marginal effects over their weighted graph. The (weighted) average degree is 0.0329, so our results can be compared in an approximation, by considering 0.0329 \( \hat{\beta} \).

In general, our estimates have the same sign and significance as those of [4]. Our estimates are in the range of [0.002, 0.013], after reweighting as mentioned before, somewhat comparable to theirs of [0.016, 0.022] (in the case of the outcome of the number of public employees, in Table 3). Hence, we find similar qualitative effects, although a smaller magnitude. Recall that our confidence set is built without assuming that \( \beta_0 \) is consistently estimable.

In Figure 3.4, we show the results of our estimated network externalities for the estimates from Table 3.5 for the importance of being a department capital. The average network externality is computed from

\[
\frac{1}{N} \sum_{i \in \mathbf{N}} \frac{1}{n_P(i)} \sum_{j \in N_P(i)} \frac{\beta_0 \hat{\gamma}_{dc}}{n_P(i)(1 - \beta_0 c_{ii})(1 - \beta_0 c_{ij})},
\]

where \( \hat{\gamma}_{dc} \) is the estimated parameter of the \( X_1 \) variable department capital, and we vary \( \beta_0 \) within its confidence set. The parameter is defined in Section 3.3, and captures the average

\(^{18}\)For this, we assume the exclusion restriction in [4], namely that historical variables only affect prosperity in the same municipality. This means that although one’s historical variables (Total Crown Employees, Distance to Royal Roads, Colonial State Agencies and Historical Population, as well as functions thereof) can affect the same municipality’s prosperity, it can only affect those of the neighbors by impacting the choice of state capacity in the first, which then impacts the choice of the state capacity in the neighbors.
Table 3.5: State Presence and Networks Effects across Colombian Municipalities

<table>
<thead>
<tr>
<th>Outcome: The Number of State Employees</th>
<th>Baseline</th>
<th>Distance to Highway</th>
<th>Land Quality</th>
<th>Rivers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>[0.16, 0.31]</td>
<td>[0.15, 0.32]</td>
<td>[0.17, 0.39]</td>
<td>[0.09, 0.38]</td>
</tr>
<tr>
<td>(dy_i/d) (colonial state officials)</td>
<td>[-0.051, 0.001]</td>
<td>[-0.045, 0.001]</td>
<td>[-0.043, 0.000]</td>
<td>[-0.024, 0.003]</td>
</tr>
<tr>
<td>Average (dy_i/d) (colonial state agencies)</td>
<td>[-1.138, 3.760]</td>
<td>[-1.335, 2.742]</td>
<td>[-0.609, 3.388]</td>
<td>[-1.775, 1.987]</td>
</tr>
<tr>
<td>Average (dy_i/d) (distance to Royal Roads)</td>
<td>[-0.010, 0.009]</td>
<td>[-0.008, 0.010]</td>
<td>[-0.007, 0.015]</td>
<td>[-0.005, 0.012]</td>
</tr>
<tr>
<td>(n)</td>
<td>1018</td>
<td>1018</td>
<td>1003</td>
<td>1003</td>
</tr>
</tbody>
</table>

Notes: Confidence sets for \(\beta\) are presented in the table, obtained from inverting the test statistic \(T(\beta)\) from Section 3.3, with confidence level of 95%. The critical values in the first row come from the asymptotic statistic. Downweighting is used. The average marginal effects for historical variables upon state capacity are also shown. This is computed from finding the confidence set for the appropriate \(\gamma\) estimate. For Colonial State Agencies and Distance to Royal Roads, since they enter in quadratic form in \(X_1\), we show the average marginal effect. All specifications include controls of latitude, longitude, surface area, elevation, rainfall, as well as Department and Department capital dummies. Instruments are constructed from payoff neighbors’ sum of the \(G_P\) neighbors values of the historical variables Total Crown Employees, Colonial State Agencies, Colonial State Agencies squared, population in 1843, distance to Royal Roads, distance to Royal Roads squared, together with the non-linear function \(Z_i = n_P(i)^{-1} \sum_{j \in X_{P}(i)} \lambda_{ij} \lambda_{ij} X_{j,t}\). Column (2) includes distance to current highway in \(X_1\), Column (3) expands the specification of Column (2) by also including controls for land quality (share in each quality level). Column (4) controls for rivers in the municipality and land quality, in addition to those controls from Column (1). One can see that the results are very stable across specifications.

The figure shows that there is a strong and increasing network externality from being a department capital over the range of the confidence set of \(\beta\). This indicates that the effect of being a capital has spillovers on other municipalities: since \(\beta > 0\), and one expects that department capitals have more state presence and resources, being a department capital yields increasing returns the stronger the complementarity.
**Figure 3.4: Average Network Externality from being a Department Capital**

Notes: The figure presents the average network externalities from being a department capital. We use the estimated results from Column (3) in Table 3.5. This captures the externality for a municipality from being a department capital, which involves higher state presence and centralization of resources. This effect is not only the direct effect, but it also quantifies a reflection effect: neighbors of department capitals also benefit from it. The grey shaded area represents the 95% confidence interval for $\beta_0$.

### 3.6 Conclusion

This chapter proposes a new approach of empirical modeling for interactions among many agents when the agents observe the types of their neighbors. The main challenge arises from the fact that the information sharing relations are typically connected among a large number of players whereas the econometrician observes only a fraction of those agents. Using a behavioral model of belief formation, this chapter produces an explicit form of best linear responses from which an asymptotic inference procedure for the payoff parameters is developed. As we showed, this explicit form gives a reduced form for the observed actions, and exhibits various intuitive features. For example, the best linear responses show that network externality is heterogeneous across agents depending on the relations of their payoff neighbors.

The advantage of our chapter’s approach is two-fold. First, the empirical modeling according to our approach accommodates a wide range of sampling processes. Such a feature is crucial because the econometrician rarely has precise knowledge about the actual sampling process through which data are generated. Second, the model can be used when only part of the players are observed from a large connected network of agents.
Chapter 4

Unbundling Polarization

4.1 Introduction

This chapter focuses on a set of open questions in the literature on political polarization, a phenomenon that has taken a sharply increasing tack since the mid-1970s [126, 129]. A first, main question concerns the exact assessment of the role played by parties in legislative activity. This typically takes (at least) two perspectives. The first perspective concerns itself with the role of political parties in setting extreme agendas, selecting the policy issues under deliberation [21, 53]. A second perspective focuses on party discipline, how much parties have been a source of voting divergence by inducing more adversarial vote choices of their members. Both perspectives appear germane to the debate on polarization, as the power of party leaders vis-à-vis rank and file has been increasing over time.2

A related and equally important open question extends to the primitive problem of assessing the ideological stance of politicians in the first place - that is, absent any equilibrium disciplining by parties on floor votes [113]. Polarization appears driven at least in part from the replacement of more moderate with more extreme legislators [164] and to a much lower extent by changes in individual preferences (consistently with [54, 116]). To answer who these politicians are - or more precisely where their policy ideal points are located - requires recovering an unbiased distribution of within-party individual ideologies before legislators are persuaded by the leadership (we will refer to this latter action as “whipping”). Such distributions are of great interest to the political economy and political science scholarship.

---

1This chapter is a joint work with Chad Kendall and Francesco Trebbi
2See [16] for a discussion of whether political polarization is the result of better internal enforcement by party leaders. Also [139] for a recent discussion.
focused on the behavior of national legislators [117, 129, 133, 150].

The goal of this chapter is to present a credibly identified approach for unbundling these different components - agenda setting, party discipline, and ideology - in driving polarization. How much of political polarization is the result of more ideologically polarized members of Congress relative to more polarized party leaderships setting divisive agendas and forcing their members to toe the party line? Rather than seeking a mono-causal explanation, we try to gauge the magnitude of such drivers.

A first contribution of this chapter is to provide an economic model of legislative activity for a two-party legislature. It is designed to capture strategic considerations on multiple nested dimensions. The first dimension is what issues and, given an issue, what policy alternatives are selected by parties. Policies that are not valuable vis-à-vis a specific status quo, or too hard to pass given the chamber composition, may not even be pursued. The second dimension is whether, once a certain alternative to a status quo is proposed, the leadership decides to invest in acquiring extra information to ascertain the prospects of that specific policy alternative (i.e. “to whip count” a bill) or not. Certain policy alternatives may be worth the extra informational effort, others not. Again, policies that appear not promising once more information is acquired may not be further pursued by the leadership.

A third dimension for consideration by the party is, if a bill is eventually brought to the floor for a vote, which legislators to discipline (i.e. “to whip”), and by exactly how much, in order maximize the likelihood of legislative success, given the opposing party may whip as well. As the model formalizes, member voting decisions on legislation are ultimately endogenous to all phases of this process.³

³Seminal work from [52], [53] and [7] emphasizes the importance of parties for the functioning of Congress. It focuses on how parties use available institutions to coordinate and set policies to their benefit, as well as how party leaders work for that goal with their party members. Cox and McCubbins emphasize institutional mechanisms by which parties get their policies on the floor, and block the minority’s policies. They discuss the incentives to do so, which included the “brand” value of a party, increasing reelection chances for politicians, increasing the coordination of policies that the politicians may be unsure of, setting policy positions, as well as helping the enforcement and coordination of policies and votes. Evidence such as in [71] has shown that these mechanisms of policy positioning and agenda setting are present, as measured by the attendance rates and transcripts from party caucuses, and affect legislative roll call voting. [7] and his Conditional Party Government theory proposes that parties play an important role in pushing policies of interest to the rank and file. Economists such as [37] have also taken a similar stance to party organization, emphasizing internal control issues, but with a focus on electoral success.
sources by [65] for the 95th to 99th Congress (1977 to 1986), at the inflection point of contemporary party polarization. These counts, which are run by the party leadership in order to ascertain the floor prospects of specific bills, allow us to introduce information on the actual position of members before a floor vote occurs. Whip counts provide information needed to pin down ideological positions of legislators at a point before party control is exerted and, in fact, these counts are precisely used to identify members on the fence on a bill. An informational argument at the base of our use of whip counts relies on two arguments. The information revelation value of whip counts resides in the repeated interaction between members and the leadership, which limits the ability of rank-and-file politicians to systematically lie or deceive powerful observers, their own party leaders. Their interactions are frequent and stakes are generally high. In addition, by a revealed preference argument, the fact that costly whip counts are systematically employed by the leadership on crucial bills bear witness to their usefulness and informational value - why would the leaders spend time on these counts otherwise. In fact, the Majority and Minority Whips who organize these counts are high ranking position in the party hierarchy.

From whip counts information one also learns about party discipline. Switching behavior in Yea/Nay between the whip count stage and the roll call stage provides the identifying variation useful to pin down the extent of whipping - who is the target of party influence and how much legislators close to the marginal voter for a bill need to be persuaded.

Finally, the new data provides identifying variation for assessing agenda setting. This arises from the fact that not all bills that are voted on the floor are previously whip counted and that certain bills that are whip counted are subsequently dropped without a subsequent vote. From flexible assumptions on the distribution of latent status quo policies and from theoretically identified thresholds determining which bills are roll called and/or whip counted, we can identify policies that are never proposed and never voted.

To conduct such an analysis, we must first recover legislator ideal points. Standard

---

4The data structure of whip counts has been explored before, such as in the work of [152] and [61], for example, but with different scope or objective. In both papers, the data was collected as the authors worked within the Whip Offices (as American Political Science Association Congressional Fellows). Our final data provides a comprehensive set-up: for many bills over different Congresses, we can track the voting intentions of politicians, how those changed in the final vote, and the whips who were responsible in making these changes happen. Two works, in particular, have looked at whip counts in the context of parties and party discipline. [35] look at 16 whip counts and their roll calls and find that most of the switching of votes has gone in the direction of party leaders. They argue that even if this undermines the true impacts of whips (as many of the votes are guaranteed by leaders in equilibrium, without having it actually changed), it still presents evidence of high effectiveness of this measure. [66] also use whip counts, and provide an extensive survey of whipping in the House of Representatives and the Senate, drawing attention to some historical examples.

5For a recent instance, consider the early 2017 efforts to repeal the Affordable Care Act by the Republican leadership in the House, which were repeatedly whip counted, but not voted.

83
approaches to the estimation of ideal points based on random utility models or optimal classification approaches to roll call votes alone, such as the popular DW-Nominate [150], miss important density in the middle of the support of the ideological distribution. These methods, which confound party discipline and agenda setting in the estimation of individual ideology [158], show a polarization level of ideal points much larger than the actual one based on our unbiased estimates. This rationale finds exact quantification in our environment. Across our Congresses the distance between party medians is on average one third of the corresponding distance based on standard DW-Nominate estimates. According to our estimates the share of total polarization attributed to party discipline as opposed to ideological drift over time varies from 64.7 percent in the 96th Congress to 71.2 percent in the 99th Congress.

The chapter tackles several paradoxes in the literature on the political economy of legislatures, including the observation by [111, 112] that party unity in floor voting may not necessarily be conclusive evidence of discipline. It is at its core an identification critique. Politicians from the same party are likely to share a similar ideology, hence they could be voting the same way, regardless of any role for party leadership.\(^6\)

With our structural estimates at hand, we show that, quantitatively, parties matter substantially. To assess the importance of party discipline, we show counterfactuals that shut down the whipping phase of floor voting and produce alternative outcomes in roll calls. Eliminating party discipline in the form of whipping is precisely rejected relative to a model with party discipline using model selection tests. The extent of party discipline is statistically different from zero, quantitatively sizable, and growing between 1977 and 1986. Given the specific time period over which our whip count data is available, we are also able to assess the role of parties in steering particularly salient bills in the early 1980s, including the National Energy Act of 1977 (H.R. 8444), the Foreign Intelligence Surveillance Act of 1978 (H.R. 7308), the Contra affair in Nicaragua: prohibiting covert paramilitary activity in Nicaragua (H.R. 5399) in 1984, the lifting the arms embargo to Turkey in 1978 (H.R. 8444).

---

\(^6\)The main difficulty lies in being able to compare outcomes with parties, to outcomes with none. In a series of works, Keith Krehbiel ([111], [112], [113]) argued that the previous literature failed to address the confounding issues of whether parties are effective, or whether they are only a grouping of similar minded politicians. This identification problem comes from using outcomes such as roll call votes, party cohesion, or party unity scores. These measures are a combination of politicians' preferences and of party effects. Politicians from the same party are likely to share similar ideologies, so they could be voting in the same way regardless of party discipline. The paradox, as stated by [112], is that this would make it seem that parties are strongest when they are most homogeneous ideologically (and hence, when they are needed the least). That, in turn, leads to an empirically difficult problem: how does one separate individual ideology measurements from party effects? In particular, how does one estimate party effects, when ideology measures confound both parties and individual ideologies?
the implementation of the Panama Canal Treaty (H.R. 111) in 1979, and several key tax bills.

This chapter touches several strands of literature. It is primarily concerned with polarization in the political elite. The empirical literature on political polarization has a rich history [148] and has experienced a recent resurgence in interest due to glaring increases in partisanship ([126], but also media reports\(^7\)). Exemplifying one of the most popular existing procedures to estimate legislator ideology from the political economy and political science perspectives\(^8\), [129] offers a broad discussion of this research area and links it to parallel relevant phenomena, such as the co-determined evolution of U.S. income inequality [146]. Raising political polarization has been detected not only in legislator ideology assessments based on roll calls but in candidate survey responses [139], congressional speech scores [79], and campaign contributions measures [28]. Considerations on polarization from the economic perspective and related to the seemingly increasing policy gridlock after the 2008 financial crisis are offered in [135].

We contribute to this discussion from the empirical perspective, in an effort to parse quantitatively some of the deep determinants of polarization. In this respect our work complements other recent attempts, such as [139]. It differs in terms of theory, identification strategy, and in the use of a structural approach.

These decomposition efforts are rooted in an older related literature that seeks ways to separate a politician’s true policy preferences from that of the party, by focusing on situations one or the other factor would not be present. [158] propose one such method of separating party effects from politician ideology, which has been widely used and adapted (e.g. [127, 138]). The argument is that parties concentrate their efforts on results they can impact, such as close legislative votes. Seemingly, expected lopsided votes would not attract nor need party intervention. Absent party effects on lopsided votes, [158] argue in favor of estimating individual ideologies from a first stage on lopsided roll calls alone. After recovering estimates of individual preferences, in a second stage they study close votes to recover party effects, given the previously estimated legislator true preferences. There are two main methodological obstacles to this this approach. First, which vote is lopsided and which is a contested is endogenous to the choice of policy alternative by the agenda setter (see the discussion in [21]). This selection mechanism will be explicit in our framework. Secondly, [127] note that this method provides poor identification, due to minimal varia-

---

\(^7\)See, for instance, Philip Bump, December 21, 2016, “Farewell to the most polarized Congress in more than 100 years!” *Washington Post.*

\(^8\)Among the standard approaches to estimation are [49, 92, 149].
tion of vote choices within a party in lopsided votes and offer a nonparametric alternative. Evidence of party effects appears quantitatively small according to these studies.

Our chapter attempts to address both methodological issues and does not rely on arbitrary selection of votes where parties are active or not. Other closely related papers such as [49], which use Bayesian methods to estimate ideal points, also employ lopsided bills to recover party discipline. [12]) use a survey directly targeted at candidate ideology (NPAT, also used in [139]) to estimate ideal points, hence moving away from roll calls.[9]

The rest of our work is organized as follows. Section 4.2 presents our model and Section 4.3 our main analytical results. Section 4.4 describes our data, with emphasis on our application of whip count information. Section 4.5 focuses on the identification of the model and Section 4.6 presents our estimator. Section 4.7 discusses our estimation and Section 4.8 our counterfactual exercises and benchmarks our analysis to extant metrics of polarization. Section 4.9 concludes. For convenience, all Figures and Tables referred to in the Results are shown after the Conclusion. The Appendix contains all proofs and additional empirical supporting material.

4.2 Model

Two parties compete for votes on a series of issues that make up a congressional term. To discipline members, each party employs whips who serve two purposes: they aggregate information, and, at a personal cost, can persuade members to vote along party lines. For a given status quo policy, a party choose the alternative policy (the bill) to be voted upon. It does so accounting for both its own ability to discipline (whip) its members, as well as that of the other party, and on the value and likelihood of passage of the alternative policy. Because floor votes are costly, not all status quo’s are contested. In addition, the proposing party can employ a formal whip count which allows it to obtain additional information about a bill’s probability of success before a floor vote, and to drop bills that are unlikely to pass conditional on these counts[10] Whether the proposing party chooses to conduct a

---

[9] Another approach looks at politicians who change party and see how their voting behavior changes. As [142] finds, congressmembers who switch party do change voting patterns, suggesting that ideology is not their sole decision factor. Our model results in this change in behavior. Nonetheless, other methods such as DWNominate, rely on distinct ideology estimates for the same legislator depending on their party affiliation.

One interesting historical approach is presented by [99]. By looking at congressmembers who served the U.S. House and then went serving in the Confederate House during the American Civil War, he finds striking differences in the estimated ideologies for the same politician from voting behavior in the different Houses. Since legislators were the same, under very similar institutional settings, he concludes (with further evidence) that differences were due to agenda setting and party discipline rather than mere ideology.

[10] The party not setting the agenda may also conduct a whip count, but this occurs less frequently in our data so we do not model its reason for doing so.
formal whip count depends upon its option value relative to the fixed cost of undertaking this process. Not all bills are therefore brought to the floor and not all status quo policies are even contested to begin with.

4.2.1 Preliminaries

Party members vote on a series of policies at times $t = 1, 2, \ldots, T$ with the majority vote determining the winning policy. We work in a single-dimensional ideological space. Each party, $p \in \{D, R\}$, has a mass of $N_p$ members whose underlying ideologies, $\theta$, are continuously distributed with cumulative distribution functions (CDFs), $F_p(\theta)$. We assume that the corresponding probability distribution functions (PDFs), $f_p(\theta)$, have unbounded support. The median member(s) of a party are identified by $\theta_{m,p}$ and represent the preference of the party overall. We assume without loss that $\theta_{m,D} < \theta_{m,R}$. In each period, with probability $\gamma$ party $D$ is randomly recognized to set the policy alternative, $x_t$, to be put to a vote. With the remaining probability $1 - \gamma$, party $R$ is recognized. The recognized party draws a status quo policy, $q_t$, from the party-specific continuous CDF, $W_p(q)$, with corresponding PDF, $w_p(q)$, which is assumed to have unbounded support.\[^{11}\]

4.2.2 Preferences

There are three sets of actors for each party: non-whip members, whip members, and the party itself.

Whips are a ‘technology’ that a party uses to discipline its members. We take the mass and ideologies of whips as exogenous and assume an exogenous matching of whips to members for which they are responsible, such that each member is controlled by exactly one whip. Whips acquire information from members and are eventually rewarded for obtaining votes that the party desires.

All party members (whips and non-whips) derive expressive utility from the policy, $k_t \in \{q_t, x_t\}$, that they vote for. This utility is given by $u(k_t, \omega_{jt})$, where $\omega_{jt} = \theta_i + \delta_{jt} + \delta_{jt}^2 + \eta_i^1 + \eta_i^2$ determines their position on a particular bill. We assume a symmetric, strictly concave utility function: $u(k_t, \omega_{jt}) = u(|k_t - \omega_{jt}|)$ with $u(\omega_{jt}, \omega_{jt}) = u_k(\omega_{jt}, \omega_{jt}) = 0$, $u_{kk}(k_t, \omega_{jt}) < 0$.

$\theta_i$ is a member’s fundamental ideology, which we assume a constant trait of $i$.\[^{12}\]

\[^{11}\]In our application $D$ is the majority party and one can assume $\gamma \gg 1/2$. We do not model how the frequency of recognition can be affected by the leadership of both parties, but to allow more flexibility in the structure of issue selection we allow the distributions of primitive status quo’s to vary by party $p$.

\[^{12}\]In this we follow the discussion and evidence from \([116],[139]\).
member’s position on any particular bill is determined by this ideology, two idiosyncratic shocks, \( \delta_1^i, \delta_2^i \), and two aggregate shocks, \( \eta_1^i \) and \( \eta_2^i \). The need for the accrual of multiple shocks will become clearer below, where we model the information acquisition problem for the proposing party. The aggregate shocks are independent draws from normal distributions with mean zero and standard deviations, \( \sigma_1 \) and \( \sigma_2 \), respectively. The aggregate shocks \( \eta_1^i \) and \( \eta_2^i \) are common across all members of both parties. The idiosyncratic shocks \( \delta_1^i, \delta_2^i \) are identically and independently distributed across \( i \) and \( t \) according to the continuous, unbounded, and mean zero CDF, \( G(\delta) \) with corresponding PDF, \( g(\delta) \).

Whip members, in addition to their utility from voting, receive a payment of \( r_p \) (which may differ across parties) for each member \( i \) for which the whip is responsible who votes with the party. \( r_p \) may represent, for example, improved future career opportunities within the party hierarchy. We model whip influence over the members for which she is responsible as an ability to persuade a member to change his position on a particular bill. To influence a member’s position by an amount, \( y_i, t \) (i.e. to move his bliss point to \( \omega_i, t + y_i, t \)), a whip bears an increasing cost, \( c(y_i, t) \) (\( c'(t) > 0 \)), which can be thought of most simply as an effort cost. We assume \( c(0) < r_p \) so that the whip always exerts a non-zero amount of influence. The contribution to a whip’s utility from whipping is therefore given by \( \sum_i (r_p I(i\text{ votes with party}) - c(y_i, t)) \), where \( I(.) \) is the indicator function and the summation is over the members for whom he is responsible.

Each party derives utility from that of its median member, \( u(k_t, \theta_{m,p}) \) where \( k_t \in \{q_t, x_t\} \) is the winning policy. For simplicity, we assume that the party’s position, represented by their median member is not subject to idiosyncratic or aggregate shocks. Because the party does not directly bear the cost of whipping members, whipping is costless to the party (and thus both parties always whip votes to the maximum extent possible).

4.2.3 Information and Timing

The timing of the model is as follows (see Figure 4.1). At each time \( t \):

1. Party \( D \) is randomly recognized to be the proposing party with probability \( \gamma \). With

---

13 Rewarding the whip only if he switches a member’s vote does not change the results.

14 Having the shocks and influence operate on the ideological bliss point rather than as changes in utility (i.e. \( u(k_t, \theta) + \delta_1^i + \delta_2^i + \eta_1^i + \eta_2^i + y_i, t \)) simplifies the model in two ways. First, it ensures that the maximum influence exerted by a whip (see Section 4.3.2) is a constant, independent of the locations of the policies and the distance between them. Second, it ensures the expected number of votes monotonically decreases in the extremeness of the alternative policy, \( x_t \) (see the proof of Proposition 4.3.4), which need not be the case for utility shocks.

15 This assumption simplifies the policy setting decision because aggregate shock realizations never cause the proposing party to prefer the status quo over the proposed alternative.
the remaining probability, $1 - \gamma$, the proposing party is $R$. The status quo policy, $q_t$, is drawn and observed by all.

2. The proposing party chooses the policy $x_t$ as an alternative to the status quo $q_t$ and decides whether or not to conduct a whip count at a cost, $C_w > 0$.

3. The first aggregate and idiosyncratic shocks, $\eta^1_t$ and $\delta^1_{i,t}$, are realized and observed noisily: each member observes his idiosyncratic shock, $\delta^1_{i,t}$, and the policy he prefers, $u(x_t, \theta_i + \delta^1_{i,t} + \eta^1_t) \leq u(q_t, \theta_i + \delta^1_{i,t} + \eta^1_t)$, but not the realization of $\eta^1_t$.

4. If a whip count is undertaken, each member makes a report, $m_{i,t} \in \{\text{yes, no}\}$, to his whip, answering the question, ‘Given your position, will you vote with the party?’. The outcome of the whip count is common knowledge. In the aggregate, the whip count reveals the realization of $\eta^1_t$ (see Section 4.3.3).

5. The proposing party (conditional on the whip count, if taken) decides whether or not to proceed with the bill, taking it to a roll call vote at a cost, $C_b > 0$.

6. The second aggregate and idiosyncratic utility shocks, $\eta^2_t$ and $\delta^2_{i,t}$, are realized and observed as in the case of the first shocks. Each member observes his idiosyncratic shock, $\delta^2_{i,t}$, and the policy he prefers $u(x_t, \omega_{i,t}) \geq u(q_t, \omega_{i,t})$, but not the realization of $\eta^2_t$.

7. Whips communicate with their members in order to learn the sum of the aggregate shock, $\eta^1_t + \eta^2_t$.

8. Whips learn the sum of the idiosyncratic shocks, $\delta^1_{i,t} + \delta^2_{i,t}$ of the members for whom they are responsible and then choose the amount of influence to exert, $y_{i,t}$, with each member.

9. The roll call vote occurs.

The information structure (who knows what and when) is a formalization of the two main duties of a whip. First, whips aggregate information - no single member is likely to know
the outcome of the bill, so information must be aggregated across all members in order to (i) decide whether or not to continue with a bill (whip count) and (ii) decide how much to try to influence members. Second, whips, by maintaining close relationships with the rank-and-file members they are responsible for, obtain information about individual positions, and use this information to decide which members can be most easily persuaded to toe the party line.

4.3 Analysis

We solve the model via backwards induction. In Sections 4.3.1 and 4.3.2, we determine the decisions of members and whips. These decisions are the same in each party, so we drop the party label for convenience. In Sections 4.3.3 through 4.3.5 we turn to the decisions unique to the proposing party: which, if any, alternative policy to pursue, and whether or not to conduct a floor vote and a whip count.

4.3.1 Roll Call Votes

Prior to the roll call vote, whips communicate with the members for whom they are responsible in order to learn the value of $\eta_1 + \eta_2$, which is necessary for deciding how much influence to exert (see Section 4.3.2). To do so, each whip asks each member for which he is responsible the policy he intends to vote for. In the aggregate, this process reveals the aggregate shocks as in the case of a whip count (see Section 4.3.3). Whips then communicate the values of the aggregate shocks to all members, so that they have full information at the time of their vote.

A member votes for $x_t$ if and only if
\[ u(x_t, \omega_{i,t} + y_{i,t}) \geq u(q_t, \omega_{i,t} + y_{i,t}) \]
where $\omega_{i,t} + y_{i,t}$ is the member’s ideological bill point after whip influence.\(^1^6\) It is convenient to define the marginal voter as the ideological position of the voter who is indifferent between the two policies. Given symmetric utility functions, this voter is located at $\omega_{i,t} = MV_t = \frac{x_t + q_t}{2}$, absent any party discipline.

4.3.2 Whip Decisions

At the time of the whipping decision (just prior to roll call), each whip has full information about the aggregate shocks and the idiosyncratic shocks of his members. He therefore knows whether or not a given (conditional) transfer induces a vote for a party’s preferred

\(^{1^6}\)Ties have measure zero due to the continuous nature of the shocks and therefore the vote tie-breaking rule is immaterial.
policy or not, and so either exerts the minimal influence necessary to make the member indifferent between policies, or exerts no influence at all. The maximum influence, $y_p^{\text{max}}$, he is willing to exert is defined by $r_p = c(y_p^{\text{max}})$, or $y_p^{\text{max}} = c^{-1}(r_p)$. $y_p^{\text{max}}$ is strictly greater than zero by assumption ($c(0) < r_p$).

Given $y_p^{\text{max}}$, Lemma 4.3.1 establishes that only members who wouldn’t otherwise vote for the party’s preferred policy, and are within a fixed distance of the marginal voter are influenced (whipped).

**Lemma 4.3.1.** Assume a party strictly prefers policy $k_t$ over policy $k_t'$. Then, the whips of the party whip only members, $i$, whose realized ideologies are on the opposite side of $MV_t$ from $k_t$ and such that $|\omega_{i,t} - MV_t| \leq y_p^{\text{max}}$.

### 4.3.3 The Whip Count

The proposing party may conduct a whip count in order to learn about the first aggregate shock, $\eta_{1,t}^1$. Whips receive reports, $m_{i,t} \in \{\text{yes}, \text{no}\}$, from each member for whom they are responsible and subsequently make these reports public. If each member reports truthfully, he reports $m_{i,t} = \text{yes}$ if $u(x_t, \theta_i + \delta_{i,t} + \eta_{1,t}^1) \geq u(q_t, \theta_i + \delta_{i,t} + \eta_{1,t}^1)$ and $m_{i,t} = \text{no}$ otherwise. Given the continuum of reports, $\{m_{i,t}\}$, by the law of large numbers, $E[\eta_{1,t}^1 | \{m_{i,t}\}] = \hat{\eta}_{1,t}^1$, where $\hat{\eta}_{1,t}^1$ is the realized value of $\eta_{1,t}^1$.

Furthermore, all members reporting truthfully forms part of an equilibrium strategy of the overall game because no single member can influence beliefs about $\hat{\eta}_{1,t}^1$, and hence cannot influence the eventual policy outcome by misreporting.\(^\text{[17]}\) We therefore assume in what follows that members play a truth-telling strategy.\(^\text{[18]}\)

We formalize these claims in Lemma 4.3.2.

**Lemma 4.3.2.** Truth-telling at the whip count stage forms part of an equilibrium strategy. Under truth-telling, the realization of the first aggregate shock, $\hat{\eta}_{1,t}^1$, is known with probability one.

---

\(^{[17]}\) In addition, misreporting does not change the amount of influence a member’s whip exerts because the whip learns the member’s true position before exerting influence.

\(^{[18]}\) As usual, there also exists an equilibrium of the whip count subgame in which each member babbles, so that nothing is learned about $\hat{\eta}_{1,t}^1$. 

91
4.3.4 Optimal Policy Choices

After observing \( q_t \), the proposing party can choose to do one of three things. One, it can decide not to pursue any alternative policy. Two, it can choose an alternative policy to pursue, \( x_t \), without conducting a whip count. In this case, the party pays the cost, \( C_b \), of pursuing the bill to the roll call stage. Three, the party can choose an alternative policy to pursue and conduct a whip count at a cost, \( C_w \). In this case, after observing the results of the whip count, the party can decide whether or not to continue with the bill at a cost of \( C_b \).

Choosing to undertake the whip count is analogous to purchasing an option: the option to save the cost of pursuing the bill should the initial aggregate shock \( \eta^1 \) turn out unfavorably.

For status quo policies to the left of the proposing party’s ideal point, \( \theta_{m,p} \), the alternative policy pursued (if any) must lie to the right of the status quo: any policy to the left of \( q_t \) is less preferred than \( q_t \) and \( q_t \) can be obtained at no cost. Similarly, for status quo policies to the right of \( \theta_{m,p} \), the proposed alternative policy must lie to the left of the status quo. In choosing how far from the status quo to set the alternative policy, the proposing party faces an intuitive trade-off: policies closer to its ideal point are more valuable, should they be successfully voted in, but are less likely to obtain the necessary votes to pass.

Lemma 4.3.3 formalizes this intuition.

**Lemma 4.3.3.** The expected number of votes that the alternative policy, \( x_t \), obtains strictly decreases with the distance between \( x_t \) and the proposing party’s ideal point.

The result of Lemma 4.3.3 guarantees that the alternative policy proposed must lie between the party’s ideal point and the status quo policy. An alternative policy on the opposite side of the ideal point from the status quo is dominated by \( x_t = \theta_{m,p} \), which is both more preferred and obtains more votes in expectation.

For ease of exposition, for the remainder of the analysis we present the case in which party \( D \) is the proposer - the case of party \( R \) is symmetric. Given the whipping technologies available to each party defined by the maximum influence their whips are willing to exert, \( y_{R_{\text{max}}} \) and \( y_{D_{\text{max}}} \), we can define the position of the marginal voter when the alternative policy is such that it obtains exactly half of the votes. Denote this position, \( \hat{MV}_{i,j} \), where the subscripts \( i,j \in \{L,R\} \) indicate the directions of the policy that parties \( D \) and \( R \) whip for, respectively.\(^9\) As shown in the proof of Lemma 4.3.3, for a given realized marginal voter,

\(^9\)Each \( MV_{i,j} \) is a function of many parameters of the model, so we suppress their dependencies for convenience. Note, however, that each is independent of \( q_t \) and \( x_t \).
\[ \tilde{MV}_i = MV_i - \eta_i^1 - \eta_i^2, \]

the number of votes for \( x_i \) is known with probability one due to the continuum of members in each party. Denoting the number of votes for \( x_i \) as a function of the realized marginal voter, \( Y(\tilde{MV}_i) \), each \( \tilde{M}\hat{V}_{i,j} \) is then given by \( Y(\tilde{M}\hat{V}_{i,j}) = \frac{N_R + N_D}{2} \).

In the absence of a whip count, if party \( D \) pursues an alternative policy, the alternative policy \( x_i \) must maximize

\[
EU_D^{\text{no count}}(q_i, x_i) = Pr(x_i \text{ wins})u(x_i, \theta_{m,D}) + Pr(x_i \text{ loses})u(q_i, \theta_{m,D}) - C_b
\]

where the cost of of proceeding with the bill, \( C_b \), is paid with certainty.

For status quo policies to the left of \( \theta_{m,D} \), since \( x_i \in (q_i, \theta_{m,D}] \), both parties prefer and whip for \( x_i \), the rightmost policy. Because \( Y(MV_i) \) is monotonically decreasing in \( x_i \), and therefore \( \tilde{MV}_i, x_i \) wins if and only if \( \tilde{MV}_i < \tilde{MV}_{R,R} \) so that \( Pr(x_i \text{ wins}) = Pr(\tilde{MV}_i < \tilde{MV}_{R,R}) \).

The sum of the aggregate shocks, \( \eta_i^1 + \eta_i^2 \), is normally distributed with a variance of \( \sigma^2 = \sigma_1^2 + \sigma_2^2 \) so that we can write \( Pr(x_i \text{ wins}|x_i > q_i) = 1 - \Phi\left(\frac{\tilde{MV}_i - \tilde{MV}_{R,R}}{\sigma}\right) \)

where \( \Phi \) denotes the CDF of the standard normal distribution.

For status quo policies to the right of \( \theta_{m,D} \), we have \( x_i \in [\theta_{m,D}, q_i) \). Party \( D \) therefore whips for the leftmost policy, \( x_i \), but party \( R \) may whip for either policy depending on where \( q_i \) and \( x_i \) lie with respect to \( \theta_{m,R} \). As a simplification, we assume party \( R \) always whips for \( q_i \) in this case. Under this assumption, \( x_i \) wins if and only if \( \tilde{MV}_i > \tilde{MV}_{L,R} \), so that \( Pr(x_i \text{ wins}|x_i < q_i) = \Phi\left(\frac{\tilde{MV}_i - \tilde{MV}_{L,R}}{\sigma}\right) \).

Conducting a whip count provides the option value of dropping the bill and avoiding the cost, \( C_b \), if the first aggregate shock makes it unlikely the bill will pass. After conducting the whip count to learn the first aggregate shock, party \( D \) continues to pursue the bill if and only if \( Pr(x_i \text{ wins}|\eta_i^1 = \tilde{\eta}_i^1) (u(x_i, \theta_{m,D}) - u(q_i, \theta_{m,D}) + u(q_i, \theta_{m,D}) - C_b \geq u(q_i, \theta_{m,D}) \),

where \( \tilde{\eta}_i^1 \) is the realized value of \( \eta_i^1 \) and \( u(q_i, \theta_{m,D}) \) is the party’s utility from the outside option of dropping the bill. \( Pr(x_i \text{ wins}|\eta_i^1 = \tilde{\eta}_i^1) \) is easily shown to be strictly monotonic in \( \tilde{\eta}_i^1 \), so that we can define cutoff values of \( \eta_i^1, \eta_i^1, \) and \( \tilde{\eta}_i^1 \), such that party \( D \) continues to pursue the bill if and only if \( \eta_i^1 > \eta_i^1 \) (for status quo’s to the left of \( \theta_{m,D} \)) or \( \eta_i^1 < \tilde{\eta}_i^1 \) (for status quo’s to the right).

---

20Ties occur with measure zero so any tie-breaking rule suffices.

21Similarly, if party \( R \) proposes an alternative to a status quo policy, \( q_i < \theta_{m,R} \), we assume party \( D \) always whips for the status quo. We can solve the model without these assumptions, and the results are qualitatively similar. The only difference is that the proposing party may choose to set the alternative policy such that the other party is exactly indifferent between policies in order to gain its support, rather than pushing for an alternative policy closer to the proposing party’s ideal point. Thus, the model predicts a mass of bills for which the the marginal voter is at exactly the opposing party’s ideal point. In reality, uncertainty about party positions is likely to prevent such fine-tuning of policies.

93
Given these continuation policies, prior to the whip count, party D chooses $x_t$ to maximize

$$EU_D^{count} (q_t, x_t) = Pr(\eta_t^1 > \eta_t^1) \left[ Pr(x_t \text{ wins} | \eta_t^1 > \eta_t^1) (u(x_t, \theta_{m,D}) - C_b) + \left(1 - Pr(x_t \text{ wins} | \eta_t^1 > \eta_t^1) \right) (u(q_t, \theta_{m,D}) - C_b) \right] + Pr(\eta_t^1 < \eta_t^1)u(q_t)$$

for status quo policies to the left of $\theta_{m,D}$ and

$$EU_D^{count} (q_t, x_t) = Pr(\eta_t^1 < \eta_t^1) \left[ Pr(x_t \text{ wins} | \eta_t^1 < \eta_t^1) (u(x_t, \theta_{m,D}) - C_b) + \left(1 - Pr(x_t \text{ wins} | \eta_t^1 < \eta_t^1) \right) (u(q_t, \theta_{m,D}) - C_b) \right] + Pr(\eta_t^1 > \eta_t^1)u(q_t)$$

for status quo policies to the right of $\theta_{m,D}$.

We define $x_t^{count}$ and $x_t^{no \, count}$ to be the optimal alternative policies pursued (if any alternative is pursued) when a whip count is conducted and when it is not, respectively. Proposition 4.3.4 shows that, provided that the cost of pursuing a bill, $C_b$, is not too large, these optimal policies are unique and bounded away from the party’s ideal point.

Furthermore, the alternative policy pursued with a whip count is closer to the party’s ideal policy. Intuitively, the fact that a whip count allows the party to drop bills that are unlikely to pass after observing the first aggregate shock allows it to pursue policies that are more difficult to pass.

**Proposition 4.3.4.** There exists a strictly positive cutoff cost of pursuing a bill, $\hat{C}_b > 0$, such that for all $C_b < \hat{C}_b$, the optimal alternative policies, $x_t^{count}$ and $x_t^{no \, count}$, are unique and contained in $(q_t, \theta_{m,D})$ for $q_t < \theta_{m,D}$, contained in $(\theta_{m,D}, q_t)$ for $q_t > \theta_{m,D}$, and equal to $\theta_{m,D}$ for $q_t = \theta_{m,D}$.

The requirement in Proposition 4.3.4 that $C_b$ be sufficiently small is for analytical purposes only for the whip counted case. Numerically, we have been unable to find a counterexample in which the proposition does not hold.

### 4.3.5 The Whip Count and Bill Pursuit Decisions

To complete the analysis, we determine for which status quo policies alternative policies are pursued and, when they are pursued, whether or not a whip count is conducted.
Define the value functions, \( V_{\text{count}}^D (q_t) = EU_{\text{count}}^D (q_t, x_{\text{count}}^t) - u(q_t, \theta_{m.D}) \) and \( V_{\text{no count}}^D (q_t) = EU_{\text{no count}}^D (q_t, x_{\text{no count}}^t) - u(q_t, \theta_{m.D}) \), as the gains from pursuing an alternative policy with and without conducting a whip count, respectively (note that these definitions account for the cost of pursuing a bill, \( C_b \), but ignore the cost of the whip count, \( C_w \)).

Lemma 4.3.5 characterizes the value functions as a function of the status quo policy.

**Lemma 4.3.5.** Fix \( C_b < \hat{C}_b \) such that the optimal alternative policies, \( x_{\text{count}}^t \) and \( x_{\text{no count}}^t \), are unique. Then, for all \( q_t \neq \theta_{m.D} \), the value of pursuing an alternative policy with a whip count, \( V_{\text{count}}^D (q_t) \), strictly exceeds that without, \( V_{\text{no count}}^D (q_t) \). Furthermore, both value functions strictly decrease with \( |q_t - \theta_{m.D}| \), but the difference between them, \( V_{\text{count}}^D (q_t) - V_{\text{no count}}^D (q_t) \) strictly increases.

Intuitively, both value functions decrease as the status quo approaches the proposing party’s ideal point because there is less to gain from an alternative policy. More interestingly, the reason the difference between the value functions increases as the status quo approaches the party’s ideal point is because the whip count is an option that allows the proposing party to initially pursue a bill, but drop it if the initial aggregate shock turns out to be unfavorable (thus avoiding the cost, \( C_b \)). This option value is always positive because the party could always ignore the result of the whip count. It increases as the status quo nears the party’s ideal point because passing an alternative policy becomes more difficult (fixing \( x_t \), as \( q_t \) approaches \( \theta_{m.D} \), the marginal voter approaches \( \theta_{m.D} \), resulting in a lower probability of passing). Therefore, exercising the option becomes more likely, and thus more valuable.

Using the nature of the value functions, Proposition 4.3.6 shows which bills are pursued with and without a whip count, accounting for the fact that whipping is costly.

**Proposition 4.3.6.** Fix \( C_b < \hat{C}_b \) such that the optimal alternative policies, \( x_{\text{count}}^t \) and \( x_{\text{no count}}^t \), are unique and fix the cost of a whip count, \( C_w > 0 \). Then, we can define a set of cutoff status quo policies, \( q_l, \tilde{q}_l, q_r, \) and \( \tilde{q}_r \), with \( q_l \leq \tilde{q}_l < \theta_{m.D} < q_r \leq \tilde{q}_r \) such that:

1. for \( q_t \in [\tilde{q}_l, q_r] \), the optimal alternative policy, \( x_{\text{no count}}^t \), is pursued without conducting a whip count.

2. for \( q_t \in (q_l, \tilde{q}_l] \cup [q_r, \tilde{q}_r) \), the optimal alternative policy, \( x_{\text{count}}^t \), is pursued and a whip count is conducted. 

95
Figure 4.2: Example of Value Functions

Note: Value functions of pursuing an alternative policy with and without a whip count. The x-axis shows $q$.Party $D$ is the proposing party. The value functions are simulated using $\theta_{m,D} = -0.5$, $\theta_{m,R} = 0.5$, $MV_{R,R} = MV_{L,R} = -0.5$, $\sigma_1 = \sigma_2 = 1$, $C_b = 0.5$, $C_w = 0.025$, and quadratic utility.

3. for $q_t \in (q_l, q_r)$, no alternative policy is pursued.

We illustrate Proposition 4.3.6 by example in Figure 4.2.

For status quo policies nearest to party $D$’s ideal policy, alternative policies are never pursued. There are two reasons for this. First, the optimal policy alternative that could be proposed would be very much opposed by $R$ and have a low chance of success. Secondly, because $q_t$ is close to $\theta_{m,D}$ to begin with, any additional policy change is not very valuable. For status quo policies farther away, alternative policies may be pursued with or without a whip count, but when both are possible (as in the empirically relevant case illustrated), it is always policies farthest from the party’s ideal policy that are pursued without a whip count, as they are easier to pass in roll call.

4.4 Data

The data used in this chapter come from two main sources. We use data on whip counts, compiled from historical sources by Professor Lawrence Evans (College of William and

96
Mary) and roll call voting data from VoteView (e.g. [149,150]).

By merging the data on whip counts to the roll call voting data, we can see the variation of outcomes from a Congress member due to whipping.

The data collected by Professor Evans is a comprehensive set of whip counts retrieved from a variety of historical sources. These are mostly from historical archives holding former whip and party leaders’ papers. The data collection procedure is described in depth in [65], and involved visits to the archives, collection of the data, and treatment by his team of researchers. We focus on his data from 1977-1986. Our specific choice of time period is driven by the fact that whip count data as comprehensive and complete as the one for Congresses 95th-99th is not systematically available for other Congresses. This is mostly due to idiosyncratic differences in the effort and diligence in record keeping by the Whips. It is nonetheless the case that the period under analysis in the chapter is interesting, sitting at the inflection point of modern political polarization in American politics. Our data hence captures a crucial turning point.

For the Republican Party, between 1977-1980, the data originally comes from the Robert H. Michel Collection, in the Dirksen Congressional Center, Pekin, Illinois, Leadership Files, 1963-1996. This part of the data “appears to be nearly comprehensive about whip activities on that side of the partisan aisle, 1975-1980”.

For the Democratic Party from 1977 to 1986, data comes from the Congressional Papers of Thomas S. Foley, Manuscripts, Archives and Special Collections Department, Holland Library, Washington State University, Boxes 197-203. Although John Brademas was the Majority whip from 1977 to 1980, his papers are collected within the Thomas Foley Collection (his successor). According to [65], “the Brademas records are extensive and very well organized, and I am confident that they are nearly comprehensive. For that matter, I also have a similar sense of the archival file from Foley’s time in the position”. The data also allows us to merge with Roll Call data, since Professor Evans associates it to bill that was voted on the floor (if the latter was sufficiently close to the one that had a whip count).

In total, we have 340 bills with whip counts covering the period of 1977 to 1986. 70 of the bills are Republican Whip Counts, from the years 1977-1980. The remaining 270 are Democratic Whip Counts, from 1977 to 1986. Within the file for each bill, we have data on how each Congress member responded “Yea” or “Nay” to the party’s question at the whip count stage. Some bills include further whip counts (i.e. a second, third whip count), which follow the same structure. Table 4.1 shows aggregate statistics for the progression of bills in our time frame. These include the number of bills whip counted (dropped and pursued
in roll calls), as well as those roll called.\footnote{\textsuperscript{22}}

The bills included in this data address a variety of questions about foreign aid, domestic policy, economic policy, among others. Some bills that are covered include the National Energy Act of 1977 (H.R. 8444), the Foreign Intelligence Surveillance Act of 1978 (H.R. 7308), Healthcare for the Unemployed Act of 1983 (H.R. 3021), the Dr. Martin Luther King Jr. National Holiday Bill of 1983 (H.R. 3345), the Contra affair in Nicaragua: prohibiting covert paramilitary activity in Nicaragua (H.R. 5399) in 1984 (as well as other bills regarding funds for those), the lifting the arms embargo to Turkey in 1978 (H.R. 12514), the implementation of the Panama Canal Treaty (H.R. 111) in 1979, and successive votes for increasing the debt limit.

We merge the data from whip counts and whip identities to the final votes for those bills on the floor. The roll call data comes from VoteView.org (\cite{149}), a standard reference. From the total of 340 whip counts, we obtain 238 cases which can be directly associated with a subsequent floor vote in House.

### 4.5 Transition to Estimation and Identification

For convenience to the reader, we omit time subscripts where not necessary and remind that the CDF of \( \delta_1 \) and \( \delta_2 \) is denoted as \( G \). We will assume that it is the same across parties and whips/non-whips and denote the CDF of the convolution \( \delta_1 + \delta_2 \) as \( G_{1+2} \). Finally, we recall that \( \eta_1 \) and \( \eta_2 \) are i.i.d., with Normal distributions \( \mathcal{N}(0, \sigma_1^2) \) and \( \mathcal{N}(0, \sigma_2^2) \), respectively.

From the model, a politician (from party \( D \)) will say “Yea” at the whip count if:\footnote{\textsuperscript{23}}

\[
\delta_{1,i,t} + \eta_{1,i,t} + \theta_{i,t} \leq MV_i,
\]

where \( MV_i = \frac{y_i + q_i}{2} \) is the marginal voter. Let us define two auxiliary variables:

\[
\begin{align*}
\gamma_{1,i,t} &= MV_i - \eta_{1,i,t} \\
\gamma_{2,i,t} &= MV_i - \eta_{1,i,t} - \eta_{2,i,t}.
\end{align*}
\]

\footnote{\textsuperscript{22}We also obtained data on the identity of whips (including the regional and assistant whips that compose part of the party ranks in addition to the main party whip - Majority Whip or Minority Whip) for each party since the 1970s, originally compiled by \cite{131}. The data covers the periods between the 95th and the 106th Congresses (1977 to 2000). This was originally collected from the editions of the Congressional Quarterly Almanac and Congressional Quarterly’s Politics in America and provides lists for Democratic and Republican whip membership. The data on the number of such whips by party and Congress is reported in Table 4.2. The Table shows intuitively how large the apparatus for the enforcement of party discipline within each party is.}

\footnote{\textsuperscript{23}The case with party \( R \) can be found in Appendix.}
These represent the realized Marginal Voters at the whip counts and at the roll call stages. Hence, the probability of a “Yea” at the whip count stage is given by:

\[
P(Y_{i,t}^{wc} = 1) = P(\delta_{1,i,t} + \theta_i \leq MV_t - \eta_{1,t})
\]
\[
= P(\delta_{1,i,t} \leq \gamma_{1,t} - \theta_i)
\]
\[
= G(\gamma_{1,t} - \theta_i). \tag{4.4}
\]

Now, for the roll call vote, we have a “Yea” (given that the party may whip) if:

\[
\delta_{1,i,t} + \delta_{2,i,t} + \eta_{1,t} + \theta_i \leq MV_t + y_{max}^D, \tag{4.5}
\]

Hence, the probability of a “Yea” at the roll call stage is given by:

\[
P(Y_{i,t}^{rc} = 1) = P(\delta_{1,i,t} + \delta_{2,i,t} \leq MV_t - \eta_{1,t} - \eta_{2,t} - \theta_i + y_{max}^D)
\]
\[
= P(\delta_{1,i,t} + \delta_{2,i,t} \leq \gamma_{2,t} - \theta_i + y_{max}^D)
\]
\[
= G_{1+2}(\gamma_{2,t} - \theta_i + y_{max}^D). \tag{4.6}
\]

We now proceed with some parametrization assumptions, and prove identification of the parameters of the model.

Consider the following Assumptions:

**Assumption 1 (Normalization):** We normalize one politician (without loss of generality, politician “0”) such that \(\theta_0 = 0\).

**Assumption 2 (Distributions):** (i) \(G\) is the CDF of a standard Normal distribution, with CDF denoted by \(\Phi(\cdot)\). It is the same for whips and non-whips, and across both parties\(^{24}\).

Furthermore, (ii) \(q\) follows a \(\text{Normal}(\mu_q, \sigma_q^2)\). We will allow this distribution to be party specific as well.

Assumption 1 normalizes the ideology of one politician to 0. Without this normalization, we cannot identify the individual ideologies \(\theta_i\) (just as in fixed effects regressions). We would only recover the difference of ideologies across legislators.

\(^{24}\)It is not essential for it to be a Normal distribution, we will only need it to be a standard distribution (no parameters to be estimated). The Normal distribution is convenient as it has a simple closed form for the convolution \(G_{1+2}\), which also becomes a Normal distribution.
Assumption 2 (i) implies that the variance of the distribution $G_{1+2}$ is equal to 2. We need to standardize the distribution of $\delta_1$, because the decision at the whip count stage is analogous to a discrete choice model (e.g. Probit or Logit model). The variance and mean of the errors are not identified in this class of models. We had already assumed that $\delta_1, \delta_2$ were mean zero.

Meanwhile, Assumption 2 (ii) describes $q_t$ with a flexible (parametric) distribution that satisfies the main assumptions in the theoretical model. A parametric distribution is needed, since we can only recover $q_t$ from bills that are pursued (see Proposition 4.3.6). This parametrization allows us to infer the distribution of the status quo over the bills that are not pursued as well, once we have estimated its conditional version. Note that, although we use a Normal distribution for $q_t$, the resulting distribution of marginal voters, $MV_t$, can be very different than that of a Normal distribution.

We now prove that we can identify the set of parameters, $\Theta = \{\theta_i, \gamma_{1,t}, \gamma_{2,t}, \sigma_1^2, \sigma_2^2, \mu_q, \sigma_q^2, q_l, q_r, q_{max, D}, q_{max, R}, y_{max, R}, y_{max, D}, \mu_q, \sigma_q^2, q_l, q_r, \}$, as well as the mass of bills that are whip counted.\[^{25}\]

### 4.5.1 Identification of the Model

We focus of identifying the party-specific parameters for the Democratic Party. The argument is analogous for Republicans.

From equation (4.4) under Assumption 2(i), we have that, for every $i$ and $t$:

$$\Phi^{-1}(P(Y_{wc}^{i,t} = 1)) = \gamma_{1,t} - \theta_i.$$  \(^{(4.7)}\)

The left hand side of this equation is (a transformation of) the probability of “Yea” at the whip count stage.

The difference of equation (4.7) across politicians $i$ and 0 in period $t$:

$$\Phi^{-1}(P(Y_{wc}^{i,t} = 1)) - \Phi^{-1}(P(Y_{wc}^{0,t} = 1)) = \theta_i,$$  \(^{(4.8)}\)

where we have used that $\theta_0 = 0$ (Assumption 1). Intuitively, $\theta_i$ is identified by the differences in the probability of saying “Yea” at the whip count stage for different politicians relative to the normalizer. The normalization of $\theta_0 = 0$ allows us to pin down exactly the distribution of $\theta_i$ and not of the differences. The whip count stage serves as a baseline, giving us how ideology affects the probability of saying “Yea” or “Yea” before party discipline.

\[^{25}\]That is, the mass of $q$ within the theoretical set from Proposition 4.3.6, given by $(q_l, q_r)$.
takes place.

Since \( \theta \) is known, we have that \( \gamma_{t,i} \) is known for an arbitrary \( t \) from equation (4.7). Therefore, the realized marginal voter at the whip count, \( \gamma_{t,i} \) is identified from the probability of saying “Nay” for a known politician at the whip count stage.

Moving on to the roll call period. Using Assumption 2, we rewrite equation (4.6) as:

\[
P(Y_{rc}^{t,i} = 1) = G_{1+2}(\gamma_{t,i} - \theta_i + \gamma_{t,i}') = G_{1+2}(\gamma_{t,i} - \theta_i + \gamma_{t,i}' + \eta_{t,i}) = G_{1+2}(\gamma_{t,i} - \theta_i + \gamma_{t,i}' + \eta_{t,i}) = \Phi\left(\frac{\gamma_{t,i} - \theta_i + \gamma_{t,i}' + \eta_{t,i}}{\sqrt{2}}\right), \tag{4.9}
\]

where \( G_{1+2} \) is a Normal distribution CDF with variance 2 by Assumption 2(i).

Equation (4.9) implies that:

\[
\Phi^{-1}(P(Y_{rc}^{t,i} = 1)) = \frac{\gamma_{t,i} - \theta_i + \gamma_{t,i}'}{\sqrt{2}}, \tag{4.10}
\]

for every \( i, t \).

Note that, by their definition:

\[
\gamma_{t,i} - \gamma_{t,i}' = \eta_{t,i}. \tag{4.11}
\]

Therefore, using equations (4.7), (4.10) and (4.11), we have that for an arbitrary bill \( t \):

\[
\Phi^{-1}(P(Y_{wc}^{t,i} = 1)) - \sqrt{2}\Phi^{-1}(P(Y_{rc}^{t,i} = 1)) = \gamma_{t,i} - \theta_i - (\gamma_{t,i}' - \theta_i + \gamma_{t,i}'' + \eta_{t,i}') = \eta_{t,i} - \gamma_{t,i}' = \eta_{t,i} - \gamma_{t,i}'' + \eta_{t,i}' = \gamma_{t,i}' - \gamma_{t,i}'' = \gamma_{t,i}' - \gamma_{t,i}'' + \eta_{t,i}', \tag{4.12}
\]

Taking expectations (over \( t \)) on both sides implies that:

\[
\mathbb{E}_t\left(\Phi^{-1}(P(Y_{wc}^{t,i} = 1)) - \sqrt{2}\Phi^{-1}(P(Y_{rc}^{t,i} = 1))\right) = -\gamma_{t,i}' = -\gamma_{t,i}' + \gamma_{t,i}'' + \eta_{t,i}', \tag{4.13}
\]

since \( \eta_{t,i} \) is mean zero. The intuition is that the average change of voting behavior from the whip count stage to the roll call stage for party \( D \) is given by the whipping parameter, \( \gamma_{t,i}' \). Using an average is important: there are idiosyncratic ideology shocks with every bill between both stages, but their average is zero. The changes that are not mean zero are those originating from party discipline. This argument can be repeated for every Congress and
allows us to estimate \( y_{\text{max}}^D \) for every congressional cycle\(^{26}\).

Since \( y_{\text{max}}^D \) is identified, we recover the individual values of \( \gamma_2,t \) from equation (4.10). The set of \( \gamma_2,t \) that is recovered includes bills with only roll calls (which, for convenience, we will denote as \( \gamma_{\text{rc only}}^2,t \)), as well as those that have both roll calls and whip counts.

Since \( \gamma_1,t, \gamma_2,t \) have been identified, equation (4.11) implies that the distribution of \( \eta_2,t \) is semiparametrically identified. It follows that we can recover its variance, \( \sigma^2 \).

To bound the variance of \( \eta_1,t \), denoted by \( \sigma^2_1 \), we then use the following information. We know that for bills with whip counts:

\[
\text{Var}(\gamma_1,t) = \text{Var}(MV_{i}^{\text{count}}) + \sigma^2_1. \tag{4.14}
\]

The left hand side is known, and we know that \( \text{Var}(MV_i^{\text{count}}) \geq 0 \). Hence, we have an upper bound for \( \sigma^2_1 \) given by \( \text{Var}(\gamma_1,t) \), which must be satisfied. This bound will be sufficient for empirical purposes, although we can construct an improved pointwise value through a more involved recursive argument\(^{27}\).

Now, let us consider bills that only have roll calls\(^{28}\). By Proposition 4.3.6, these bills with only roll calls have status quo’s that satisfy \( q \in (-\infty, q_l] \cup [q_r, \infty) \). For these bills, we know the distribution of:

\[
MV_{i}^{\text{rc only}} = \eta_1,t + \eta_2,t + \gamma_{\text{rc only}}^2,t. \tag{4.15}
\]

This is because \( \eta_1,t + \eta_2,t \sim \mathcal{N}(0, \sigma^2_1 + \sigma^2_2) \) and the distribution of \( \gamma_{\text{rc only}}^2,t \) is identified based on estimated bill fixed effects in the roll call votes.

The left hand side of (4.15) is given by \( \frac{x_{\text{no count},*}^{\text{count}}(q) + q}{2} \), which is a known invertible function of \( q \) (Lemma C.1.1 (2) in Appendix). Hence, the distribution of \( \{q : q \in (-\infty, q_l] \cup [q_r, \infty)\} \) is identified. This includes the truncation points, \( q_l, q_r \), and the parameters \( \{\mu_q, \sigma^2_q\} \).

\(^{26}\)For the Congresses where we do not have Republican whip counts, we can still recover its \( y_{\text{max}}^R \). This is because we have data for the Democrats for each Congress, and hence, we can recover \( y_{\text{max}}^D \) for each Congress by the argument above. Finally, note that with only roll calls, we can always identify the sum \( y_{\text{max}}^R + y_{\text{max}}^D \). This is done by taking the difference of equation (4.10) and its Republican counterpart for members of opposing parties.

\(^{27}\)As we show in the next paragraphs, given an initial value of \( \sigma^2_1 \), including its upper bound, we can recover the distribution of \( q_t \) and the remaining parameters consistent with this initial value. We can then check whether equation (4.14) holds with our initial value. If it doesn’t, we can generate a new estimate of \( \sigma^2_1 \) using the observed \( \text{Var}(\gamma_1,t) \) minus the \( \text{Var}(MV_i^{\text{count}}) \) from our previous estimate. This algorithm can work recursively until convergence of a \( \sigma^2_1 \) estimate that exactly satisfies equation (4.14).

\(^{28}\)Bills with both whip counts and roll calls are a selected subsample. A truncation on \( \eta_1 \) after the whip count, denoted by the threshold \( \overline{\eta}_1 \) in our model indicates the draw of the first aggregate shock below which the bill is not brought to the floor after a whip count (because there is insufficient support). Looking at bills with only roll calls avoids this specific selection issue for the identification of the distribution of \( q \).
as they uniquely define this truncated distribution. Since we observe the proportion of bills that only have roll calls relative to those that also have whip counts in the data, it follows that we know the mass of bills that have been whip counted. This completes the identification of the model.

4.5.2 Krehbiel’s critique: Lack of identification of $\theta_i$ and of party effects without whip counts

Without whip count data, the whipping parameter $y_{\text{max}}^D$ is not identified. To note this, we can look at equation (4.6).

If we did not know $\theta_i$ and had to estimate it from roll call data, we could redefine $\tilde{\theta}_i = \theta_i - y_{\text{max}}^D$ and have that:

$$P(Y_{i,t} = 1) = G_1 + 2 \gamma_{i,t} \theta_i + y_{\text{max}}^D$$

$$= G_1 + 2 \gamma_{i,t} \tilde{\theta}_i.$$

Hence, with roll call data alone, we cannot separate a shift in everyone’s (true) ideology from party discipline effect due to whipping (the basis of the critique in [111]). It is further important to note that a correct estimation of the $y_{\text{max}}^D$ “shift” is crucial to correctly position legislator ideology distributions and therefore to assess the extent of polarization - typically measured in inter-party distance between median ideologies.

4.6 Estimation

Given that we have identified the set of parameters of interest, we can now proceed to estimation. We observe the outcome of votes at both the whip count stage for both parties $p \in \{D, R\}$ (denoted $Y_{i,t}^{p,\text{wc}}$) and at the roll call stage (denoted $Y_{i,t}^{p,\text{rc}}$), for each politician $i \in \{1, \ldots, N\}$ and bills $t \in \{1, \ldots, T\}$.

To estimate the model, we find the distribution of $\{\theta_i, \gamma_{i,t}, \gamma_{\text{max}}^D, \gamma_{\text{max}}^R, \sigma_1^2, \sigma_2^2\}$ for politicians and bills that have whip counts and/or roll calls. This is done by Maximum Likelihood.

\[^{29}\text{Although } \hat{M}\hat{V}\text{ is still unknown, we can recover it from its definition: as we know the distribution of } \theta \text{ for each party, we know } y_{\text{max}}^p \text{ for each party and the number of politicians from each party, } N_R, N_D.\]

\[^{30}\text{We do not know, however, how that mass is divided to the left and to the right of the party median.}\]

\[^{31}\text{Although the parameters of the agenda setting part of the model are identified, we do not pursue their estimation in this chapter due to finite sample limitations.}\]
By replacing the conditional probability of voting “Yea” on the roll call given a “Yea” on the whip count by the unconditional one, we can define a pseudo-likelihood for the first step given by:

$$L(\Theta; Y_{wc}, Y_{rc}) = \prod_{p \in \{D, R\}} \prod_{t=1}^{T} \prod_{n=1}^{N} P(Y_{i,t}^{p,wc} = 1)^{Y_{i,t}^{p,wc}} P(Y_{i,t}^{p,wc} = 0)^{1-Y_{i,t}^{p,wc}} P(Y_{i,t}^{p,rc} = 1)^{Y_{i,t}^{p,rc}} P(Y_{i,t}^{p,rc} = 0)^{1-Y_{i,t}^{p,rc}},$$

(4.17)

Operating with the pseudo-likelihood as opposed to the more cumbersome original likelihood has no effect on the consistency of the estimation ([88], [171]). This is because our model is identified despite the nuisance of the dependency between the roll call and the whip count stages.

Focusing on the Democratic Party, we can use equations (4.4) and (4.6), together with our parametrization to reexpress (4.17) as:

$$L_D(\Theta; Y_{wc}, Y_{rc}) = \prod_{t=1}^{T} \prod_{n=1}^{N_D} \Phi(\gamma_{1,t} - \theta_i) Y_{i,t}^{wc} (1 - \Phi(\gamma_{1,t} - \theta_i))^{1-Y_{i,t}^{wc}} \times$$

$$\times \Phi \left( \frac{\gamma_{2,t} - \theta_i + \gamma_{max}^D}{\sqrt{2}} \right)^{Y_{i,t}^{rc}} \Phi \left( \frac{\gamma_{2,t} - \theta_i}{\sqrt{2}} \right)^{1-Y_{i,t}^{rc}},$$

(4.18)

where we use that $G$ is a standard Normal distribution CDF, $G_{1+2}$ is a Normal distribution CDF with variance 2, $N_D$ denotes the number of politicians in $D$, and $P(Y_{i,t}^{m} = 1) = 1 - P(Y_{i,t}^{m} = 0)$, for $m = wc, rc$.

Analogously, for the Republican Party we get:

$$L_R(\Theta; Y_{wc}^{R}, Y_{rc}^{R}) = \prod_{t=1}^{T} \prod_{n=1}^{N_R} (1 - \Phi(\gamma_{1,t} - \theta_i)) Y_{i,t}^{wc} \Phi(\gamma_{1,t} - \theta_i)^{1-Y_{i,t}^{wc}} \times$$

$$\times \left( 1 - \Phi \left( \frac{\gamma_{2,t} - \theta_i - \gamma_{max}^R}{\sqrt{2}} \right) \right)^{Y_{i,t}^{rc}} \Phi \left( \frac{\gamma_{2,t} - \theta_i}{\sqrt{2}} \right)^{1-Y_{i,t}^{rc}},$$

(4.19)

Our estimation problem in the first step is to maximize equation (4.17), using (4.18) and (4.19) subject to $\theta_0 = 0$ (Assumption 2 for normalization). In practice, we set the

\[32\] The derivations are shown in Appendix, resulting in equations (C.12) and (C.14).
politician in our sample with DW-Nominate score closest to 0 as our normalizer. This is done to help comparisons with previous estimates in the literature. For starting values for this optimization, we use the results of estimating (4.17) for bills with both whip counts and roll calls.

To estimate it, we must also consider two additional details of the data. First, we must assign what consists of a “Yea” and of a “Nay”, given that the questions from whip counts and roll calls might not be clear cut. For this, we use the party leader votes to assign whether that bill was a “Yea” for the party or a “Nay”. In order of priority, if available, first we use the (majority/minority) party leader’s direction of voting, (majority/minority) party whip, and then the majority of the party, if needed. For the large majority of bills, it is sufficient to look at the party leader votes. Second, because bills in the theory can originate from status quo’s both to the left and to the right of a party median and we do not observe from which side, we must be able to assign whipping directions. Again, we use party leader decisions. Our theoretical framework proposes that party leaders are proposing the bills, and should theoretically say “Yea” if they proposed it. Hence, bills that have one party leader saying “Yea” at the Roll Call with the other saying “Nay” are assigned as proposed to the first. Bills that have both party leaders saying “Yea” are assigned to the minority, the Republicans. Finally, there is a small minority of bills that have both party leaders saying “Nay” with no guidance from theory, yet we do not omit them for the purpose of being conservative in avoiding selection.

Once the first step is estimated, we estimate $\sigma^2$ by applying the variance operator on equation (4.11). With it, we find the variance of our estimated $\gamma_1 - \gamma_2$ for bills that had both whip counts and roll calls. We then use the upper bound for $\sigma^2$ given by $\text{Var}(\gamma_1, \gamma_2)$ as its estimate. We also allow some of the parameters to be time-varying. Given our identification arguments and our interest in changes over time to whipping technology, we estimate different $y_{D,\max}, y_{R,\max}$ by Congress.

33 For example, often for the minority party, but not always, a whip count will be framed in the negative: “Will you vote against...” These questions can change directions from whip counts to roll calls.

34 In the theory, these bills can originate from either party. Nonetheless, we show in our empirical results that the results are consistent with proposals by the Republicans. This is because the estimated mass of $\gamma_2$ for these bills lies to the right of the Republican median.

35 We assign these bills to the Democrats, as the majority party, as these may be token votes conceded to certain members for idiosyncratic reasons outside the model. Our choice holds under an argument of negative agenda control (for example, [53]). Absent support by both the majority party and the minority, the bill would most likely never reach the floor to begin with. Therefore, not observing a party D leader voting “Yea” for these bills is just a limitation of the proxy that we use for this assignment.
4.7 Results

We now present Maximum Likelihood estimates based on our approach. Our first step lever whip counts as being revealing of true legislator ideology. That different - and we argued, revealing - information is contained in whip counts relative to roll calls can be intuitively displayed. For bills presented by the majority party that have both whip counts and roll calls, Figure 4.3 plots the distribution of individual vote choices aligned with the party leadership at the whip count phase and the roll call phase. The number of members voting with the leadership dramatically increases. The shift in the figure is from around 160 votes aligned with the leadership at the whip count phase at the mean of the distribution to around 218 votes aligned with the leadership on average. Notice that 218 is the simple majority threshold for the chamber - what is needed to pass a bill at roll call. Around 58 members are persuaded to toe the party line on average, exactly the phenomenon we aim to capture.

Table 4.3 presents our estimates from the Maximum Likelihood Estimation for each party of the ideology and whipping parameters of the model. In this step we recover from 315 whip counts and 5424 roll call votes the estimated legislator ideologies $\theta_i$ for 711 members (the Table reports the party medians for each Congress). We also recover the party discipline parameters $y^\text{max}_D$ and $y^\text{max}_R$ for each Congress and two standard deviation parameters for the aggregate shocks, $\sigma_1$ and $\sigma_2$. All parameters are precisely identified. The model overall identification and convergence was tested repeatedly in Monte Carlo simulations, recovering the assumed parameter vectors and experimenting on a wide range of initial values.

As it can be seen from the Table, party polarization in terms of distance $\theta_{m,R} - \theta_{m,D}$ clearly widens over time. In addition, $y^\text{max}_D$ and $y^\text{max}_R$ for each Congress are positive and statistically different from 0, the alternative hypothesis for a model absent party discipline (i.e. with no whipping).

Figure 4.4 plots kernel densities of the estimated legislator ideologies, $\theta_i$, by party and over time from our full model. It also offers, as a way of comparison, the corresponding ideological distribution which would obtain when estimating a misspecified model in which we impose by construction $y^\text{max}_D = 0$ and $y^\text{max}_R = 0$. Figure 4.4 is the most intuitive representation of the bias induced by misspecification of a model where estimation of legislator ideal points does not take into consideration party discipline. For reference, the reader may consider the DW-Nominate optimal classification scores, Heckman-Snyder linear probability model scores or Markov chain Monte Carlo approaches based on roll calls alone. The
models are comparable as they are under the same assumptions: that voting is sincere and legislators know the position of the bill they have to vote. As it can be seen from Figure 4.4, the distance between party medians is accentuated by the omission of $y_{D}^{\text{max}} + y_{R}^{\text{max}}$ in the misspecified model (represented by the dashed kernel densities) relative to the model correcting for party discipline (represented by the solid kernel densities). The ideology distributions are much closer together in reality than in the misspecified model. As a proof of concept, Figure 4.5 shows the estimated legislator ideologies $\theta_i$ from our full model and from the misspecified model absent party discipline compared to DW-Nominate scores. The misspecified model and DW-Nominate trace each other accurately. However, our full model reveals a gap in density over the ideological middle ground driven by the loading on legislator ideology of a significant component of party discipline omitted using DW-Nominate. This is substantial bias in DW-Nominate, amounting to around 0.30 in DW-Nominate units.

The contribution to changes in the ability of party leader in persuading their members through whipping is apparent in our estimates. Figure 4.6 reports estimated party discipline $y_{p}^{\text{max}}$ by party across all our Congresses and their 95% confidence intervals. The $y_{D}^{\text{max}}$ and $y_{R}^{\text{max}}$ are not only precisely estimated, but also statistically different when compared at the beginning of our sample in 1977 relative to the end of our sample in 1986. The trend in $y_{p}^{\text{max}}$ for both parties is clearly positive, tracing an increase in the reach of party leaders over the rank and file. This is a consequential finding.

The change in party discipline turns out to be a major factor in explaining party polarization. Table 4.4 presents a decomposition of political polarization based on differences in the distributions of legislator ideologies, represented by $\theta_{m,R} - \theta_{m,D}$, and party discipline, given by $y_{D}^{\text{max}} + y_{R}^{\text{max}}$. The share of polarization due to party discipline ranges from 67 percent to 71 percent and appears to be also increasing over time.

In terms of further probing of our model, Table 4.5 reports a useful check. The theoretical model predicts that the equilibrium marginal voter for a bill where a party goes for a whip count should be closer to the proposer party median relative to bills where the party goes for roll call directly. The intuition is that the former types of policies should

---

36 This rise in polarization and party discipline in the mid 1970s coincides with large reforms conducted in the House of Representatives. During this period, power was heavily concentrated in the party leadership’s hands. Amongst the changes, leaders were now responsible for committee assignments, including the Rules Committee (instead of it being by seniority), larger control of the agenda progress was given to the Speaker, new tactics emerged such as packaging legislation into “megabills” and the Democratic Steering and Policy Committee was formed. The latter met regularly to gather information and determine tactics and policies, with the leadership controlling half the votes. One strong motivation for these reforms was policy: to guarantee that more liberal policies could pass and not be held back by certain Committee chairmen. See [153] for a thorough description of the reforms and motivation.
be riskier in terms of passage but have sufficiently high value to warrant the use of whip counts as option to explore the likelihood of success of the bill. Averaging over the relevant $t = 1, \ldots, T$ and comparing the estimated $\gamma_{1,t}$ under whip count and the estimated $\gamma_{2,t}$ under roll call only bills, the model predicts that $\Sigma_t | \gamma_{1,t} - \theta_{m,p} | < \Sigma_t | \gamma_{2,t} - \theta_{m,p} |$ for both parties $p \in \{D, R\}$. This is a (subtle) implication of the theory that we do not impose anywhere in estimation. It appears strongly satisfied in Table 4.5.

Additional probing of the model is in terms of fit. Table 4.6 reports in-sample model fit based on individual vote choices correctly predicted by the model. The overall fit for roll calls correctly predicted (with and without whip counts) is 82.4% and for whip counts, 66%. Because whip counts are fewer and the MLE does not penalize in any sense incorrectly predicting a whip count versus a roll calls, the fit is higher in the more numerous roll call sample. Overall, the fit of the model is very good, especially considering that not a single roll call is dropped (either lopsided or close). This differs for extant approaches that condition on (occasionally hard to justify) selected subsamples of votes. For comparison, over our sample the DW-Nominate prediction rate is 85.9%, but the procedure drops 892 roll calls that we instead use.

We conclude this section with a specification selection test for constant party discipline parameter $y_{p}^{\text{max}}$. This alternative is broader than simply including a party discipline parameter equal to zero (i.e. $y_{p}^{\text{max}} = 0$, no whipping for $p \in \{D, R\}$). Our full model is estimated against a model with strategic agenda setting in terms of choice of the status quo $q_t$ to pursue and optimally selected alternative $x_t$ but with $y_p^{\text{max}}$ constant over time. A likelihood ratio test, presented in Table 4.7, rejects the alternative model with constant $y_p^{\text{max}}$ relative to our baseline full model at high confidence levels (p-value < 0.001). This particularly includes the rejection of the hypothesis $y_{D}^{\text{max}} = y_{R}^{\text{max}} = 0$.

### 4.8 Counterfactuals

We assess the importance of party discipline with counterfactual exercises.

We analyze the importance of party discipline for the approval of legislation. To do so, we keep the policy alternatives to be voted on as in the data, but assume that parties cannot whip their members for given $x_t$. The legislators vote solely according to their ideologies. This exercise illustrates the extent that “Yea” votes are driven exclusively by party discipline, complementing the analysis of polarization in Table 4.4. We focus on a series of key bills from our sample. The results suggest that whipping was decisive in the outcomes of key votes, but not always in the same direction.
4.8.1 The importance of party discipline for the approval of legislation

Our first exercise looks at important legislation within our sample. It considers counterfac-
tual roll call outcomes had parties not been able to whip. Within our model, this means
parties cannot discipline individual legislators (i.e. we set $y_{D}^{\text{max}} = y_{R}^{\text{max}} = 0$). Here, we still
assume that party leaders present the same actual legislation as in our data, subject to the
same shocks. This means that we maintain $\gamma_{2,t} = MV_{t} - \eta_{1,t} - \eta_{2,t}$ at their estimated values.

Among the bills that we consider are key legislation in international relations, security
and for the economy. These include the lifting of the arms embargo to Turkey, the Panama
Canal Treaty, the funding of the Contras of Nicaragua, as well as important economic poli-
cies. The latter include the National Energy Act of 1977, that brought considerable changes
to the industry, and the 1984 Reagan’s Tax Reform.

The first and second columns of Table 4.8 show that our model fits these votes well.
The third column of Table 4.8 presents the results of the counterfactual exercise. It shows
that party discipline is quantitatively important for the outcomes of these bills as, in some
cases, the approval of these bills would have been reverted. There are also a series of less
obvious considerations.

Interestingly, the number of “Yea” votes can increase or decrease depending on the bill.
For a bill such as the National Energy Act, absent party discipline, we would have had 7
less “Yea” votes in support. For the Tax Reform Act of 1984, the result would have been
a decrease closer to 100 votes. For other votes, such as the Budget Authorization banning
aid to the Contras (HR 5399), the counterfactual displays an increase in “Yea” votes absent
party discipline. The explanation for the different results depends on the policy being voted
on and how that leads to different numbers of Democrats and Republicans being whipped
by their parties.

To see why this is the case, consider H.R. 5399 banning aid to the Contras. For this bill,
the Democrats were whipping in favor of the policy, while the Republicans were whipping
for the status quo. The estimated value of $\gamma_2$ is 0.678. As this bill is realized relatively far
to the right, there are not many Democrats between the cutpoint of no whipping, 0.678, and
the cutpoint with whipping, $0.678 + y_{D}^{\text{max}}$. However, there are many Republicans in between
the whipped cutpoint, $0.678 - y_{R}^{\text{max}}$, and their no whipping cutpoint, 0.678. Therefore, if
neither party is able to whip, there is little change to the Democrats’ number of “Yea”
votes, but we observe a large increase in “Yea” votes for Republicans who no longer are
whipped to the status quo. An analogous argument, with the opposite signs, holds for the

---

37This number rationalizes the large number of both Democrats and Republicans voting “Yea”, even if the
Republican leadership voted against it.
National Energy Act and for the 1984 Tax Reform.

Even larger effects can occur when parties are whipping in the same direction. This is the case for the lifting of the Turkey arms embargo (H.R. 12514, 1978), which has a decrease of about 200 “Yea” votes without whipping. For this bill, both parties whipped in the same direction. The estimated $\gamma_2$ is $-0.84$, a value at which very few politicians would have voted “Yea” absent party discipline. Therefore, removing whipping from both parties removes most “Yea” votes. A similar case is the Panama Canal Treaty (H.R. 111 in 1979).

4.9 Conclusion

Polarization of political elites is a major empirical phenomenon, that has recently reached historical highs. It has consequential implications, ranging from heightened policy uncertainty (and its deleterious consequences on investment and trade) to gridlock and inability of political elites to respond to shocks and crises.

There are contrasting views of what has been driving polarization. Some researchers point squarely at ideological polarization of legislators as a consequence of more polarized electorates, possibly in a perverse cycle of segmentation of the voting population along economic and social cleavages driving the election of extremists. Other researchers caution on the role of ideology and emphasize changes in the rules of controlling the legislative agenda, increases in the leadership’s grip on policy platforms, and the capacity of parties to more precisely reward and punish members through appointments and campaign resources. This chapter provides an identification strategy useful in separating these different drivers (all of which are at play). It provides a structural economic assessment of their role over the initial phase of modern congressional polarization, at its inflection point between the 95th to 99th Congresses.

This exercise requires an effort in solving extant political economy problems speaking to internal organization of parties - for instance in terms of internal aggregation of information from the rank-and-file and persuasion of party members on the fence. Our theoretical setting attempts to rationalize these problems within a unified structure. It offers a tractable but realistic environment that we also estimate, leveraging whip count information. A series of counterfactual exercises indicates a quantitative relevant role for party discipline, almost twice as important as legislator ideology in explaining polarization.

Future research, including by the authors, should address the possibility of extending

38This can be seen by both leaders (Rhodes (GOP) and Wright (Dem)) voting in favor of the policy.
our estimation methodology to periods where whip count data as precise and comprehensive as the one we employ here may not be available. In a separate paper we are working on an approach able to project some of the methods developed in this chapter beyond the 99th Congress. One aim is providing tools for the assessment of party discipline and unbiased estimation of legislator ideology designed to be integrated within extant optimal classification or Markov chain Monte Carlo methods.
4.10 Tables and Figures

**Figure 4.3:** Votes with the Majority Party at Whip Counts and Roll Calls

Notes: The figure presents the kernel density of the number of votes from Democrats with their party leader at the whip count and at the roll call stages. This is done for bills that had both whip counts and roll calls. The vertical line is plotted at 218, the majority needed to pass a bill in the House of Representatives.

**Figure 4.4:** Estimated ideologies, \( \theta_i \), per Party over Time

Notes: We show the distributions of estimated ideologies, \( \theta_i \), for each party and Congress in the thick lines. In the dashed lines, we show the estimated \( \theta_i \) under a misspecified model that assumes no whipping and only uses Roll Call votes. The misspecified model overestimates polarization by a (theoretical) factor of \( \gamma_{D}^{\text{max}} + \gamma_{R}^{\text{max}} \), see main text. In practice, we reestimate the misspecified model to allow for numerical and robustness corrections.
Figure 4.5: Estimated Ideologies from the Model $\theta_i$ compared to DWNominate

Notes: We show the correlation between our estimates of ideology to those of DWNominate. In the left panel, we fit our model with only roll call data (a misspecified version, assuming there is no whipping). The correlation of this model to the DWNominate estimates is 0.951. In the right panel, we present the estimates from our full model. The correlation of ideology estimates from our full model to DWNominate is 0.829. The wedge generated by party discipline is shown by the difference across both graphs. Quantitatively, two politicians from different parties that our full models estimates as having the same ideology are estimated by DWNominate to be approximately 0.3 DWNominate units apart.

Figure 4.6: Estimated $y_{max}$ per Party over Time

Notes: We show the time series estimates of the whipping technology parameter, for each party. These parameters are in units of our ideology estimates.
### Table 4.1: Summary Statistics on Bill Selection

<table>
<thead>
<tr>
<th></th>
<th>Congress 95</th>
<th>96</th>
<th>97*</th>
<th>98*</th>
<th>99*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Total Number of Bills Whip Counted</td>
<td>131</td>
<td>58</td>
<td>28</td>
<td>50</td>
<td>48</td>
</tr>
<tr>
<td>B: Number of Bills Whip Counted, but not Roll Called</td>
<td>50</td>
<td>16</td>
<td>8</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>C: Total Number of Bills Roll Called</td>
<td>1540</td>
<td>1276</td>
<td>812</td>
<td>906</td>
<td>890</td>
</tr>
</tbody>
</table>

Notes: The table presents summary statistics on bill selection. It shows how many bills were whip counted, whip counted but not roll called, and bills that were roll called over Congresses 95-99. *We do not have data for Republican Whip Counts for Congresses 97-99, see the Data section.

### Table 4.2: Number of Whips per Party

<table>
<thead>
<tr>
<th>Whips</th>
<th>Congress 95</th>
<th>96</th>
<th>97</th>
<th>98</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrats (appointed)</td>
<td>14</td>
<td>14</td>
<td>20</td>
<td>26</td>
<td>41</td>
</tr>
<tr>
<td>Democrats (elected)</td>
<td>21</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>Republicans (appointed)</td>
<td>16</td>
<td>17</td>
<td>23</td>
<td>22</td>
<td>25</td>
</tr>
</tbody>
</table>

Notes: The table presents the number of whips per Party over the different Congresses. Data is from Meinke (2008). Both party leaderships appointed whips, however, between the 95th and 106th Congresses, the Democrats also elected assistant/zones whips independently of the party leaders (Meinke (2008)).
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Congress</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95</td>
</tr>
<tr>
<td>$\gamma^{max}$, Democrats</td>
<td>0.54 (0.027)</td>
</tr>
<tr>
<td>$\gamma^{max}$, Republicans</td>
<td>0.57 (0.007)</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.45 (0.006)</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.74 (0.178)</td>
</tr>
<tr>
<td>Party Median - Democrats, $\theta_m^D$</td>
<td>-0.55 (-0.101)</td>
</tr>
<tr>
<td>Party Median - Republicans, $\theta_m^R$</td>
<td>0.09 (0.030)</td>
</tr>
</tbody>
</table>

Notes: The table presents the main estimates for the first step parameters. Standard errors are in parentheses. For time-varying parameters, such as the party specific $\gamma^{max}$, we write each Congress specific parameter. $\sigma_1, \sigma_2$ are not time varying, so we present its estimate centered in the table.
Table 4.4: Decomposition of Polarization in Ideologies and Whipping

<table>
<thead>
<tr>
<th>Congress</th>
<th>95</th>
<th>96</th>
<th>97</th>
<th>98</th>
<th>99</th>
</tr>
</thead>
</table>

Implications of Table 4.3 for Polarization

A: Polarization due to ideology ($\theta_{m,R} - \theta_{m,D}$)  
0.547 0.643 0.650 0.694 0.733

B: Polarization due to whipping ($y_{\text{max}}^R + y_{\text{max}}^D$)  
1.110 1.180 1.326 1.689 1.812

C: Share of Polarization due to whipping ($B/(A+B)$)  
0.670 0.647 0.671 0.709 0.712

Notes: The table shows how (perceived) polarization changes over Congresses. The perceived polarization is the change in “perceived” ideology due to both whipping and drifting ideologies. The proportion of this aggregate distance due to whipping is approximately 2/3 throughout the sample period.

Table 4.5: Distance of $\gamma_1, \gamma_2$ to the Party Medians

<table>
<thead>
<tr>
<th></th>
<th>Average distance of $\gamma_1, i$, to the Party Median</th>
<th>Average distance of $\gamma_2, i$, only Roll Calls, to the Party Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrats</td>
<td>0.440</td>
<td>0.824</td>
</tr>
<tr>
<td>Republicans</td>
<td>0.707</td>
<td>1.677</td>
</tr>
</tbody>
</table>

Notes: The theoretical model gives a prediction that the marginal voter with whip counts should be closer to the party median than those with only roll calls (i.e. $\sum_i |\gamma_i - \theta_{m,p}| < \sum_i |\gamma_i - \theta_{m,p}|$ for both parties $p \in \{\text{Dem, Rep}\}$). This holds true in our estimates, even though they were not imposed in estimation.
Table 4.6: Model Fit

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>% Correctly Predicted Votes (“Yea/Nay”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Model</td>
<td>Roll Call Votes</td>
<td>0.824</td>
</tr>
<tr>
<td></td>
<td>Whip Count Votes</td>
<td>0.660</td>
</tr>
</tbody>
</table>

Notes: The table presents the model fit in terms of correctly predicted “Yea” and “Nay” votes at both the Whip Counts and Roll Call stages. The fit is better for Roll Calls as they are the vast majority of bills, and we do not penalize the likelihood to improve inference for Whip Counts.

Table 4.7: Likelihood Ratio Test for Constant $y^{max}$

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimated $y^{max}$</th>
<th>Log-Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Varying $y^{max}$</td>
<td>See Table 4.4</td>
<td>$-8.769 \times 10^5$</td>
</tr>
<tr>
<td>Constant $y^{max}$</td>
<td>Dem: 0.700, Rep: 0.689</td>
<td>$-8.795 \times 10^5$</td>
</tr>
</tbody>
</table>

p-value for LR test, with 8 degrees of freedom: 0.00

Notes: We test whether the whipping parameter, $y^{max}$, is constant across all Congresses in our sample. To do so, we fit a restricted version of our model where each party’s $y^{max}$ is the same throughout all periods. We compare it to our original model, and reject the hypothesis of a constant $y^{max}$ with a Likelihood Ratio test.
### Table 4.8: Counterfactual: Voting Outcomes on Salient Bills

<table>
<thead>
<tr>
<th>Bill</th>
<th>Yea Votes (Data)</th>
<th>Yea Votes (Model Predicted)</th>
<th>Yea Votes (Counterfactual, no whipping)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aid to Turkey/Lifting of Arms Embargo, 1978 (H.R. 12514)</td>
<td>212</td>
<td>242</td>
<td>29</td>
</tr>
<tr>
<td>Foreign Intelligence Surveillance Act of 1978 (H.R. 7308)</td>
<td>261</td>
<td>291</td>
<td>330</td>
</tr>
<tr>
<td>National Energy Act, 1978 (H.R. 8444)</td>
<td>247</td>
<td>284</td>
<td>277</td>
</tr>
<tr>
<td>Panama Canal Treaty, 1979 (H.R. 111)</td>
<td>224</td>
<td>259</td>
<td>59</td>
</tr>
<tr>
<td>Tax Reform Act of 1984 (H.R. 4170)</td>
<td>319</td>
<td>409</td>
<td>286</td>
</tr>
<tr>
<td>Contra Aid, 1984 (H.R. 5399)</td>
<td>294</td>
<td>285</td>
<td>392</td>
</tr>
</tbody>
</table>

Notes: The table shows how the outcomes of votes on certain key bills in our sample would have changed without whipping. To do so, we consider a subset of 6 bills and show (i) how many votes it received in the actual roll call, (ii) how many votes the model predicted it would receive, and (iii) how many votes the model predicts it would receive, absent whipping. For (iii), we set the whipping parameters $y_{D}^{\text{max}} = y_{R}^{\text{max}} = 0$, while keeping the realized marginal voter $\hat{\gamma}_{t} = MV_{t} - \eta_{1,t} - \eta_{2,t}$ at their estimated values. Hence, the table just shows the effect of whipping, but without the party reoptimizing the agenda under that assumption.
Chapter 5

Information Accumulation and the Timing of Voting Decisions

5.1 Introduction

Understanding how voters acquire information and learn is key to understanding voter decisions. A large theoretical and empirical literature proposes how information and learning can shape electoral outcomes, incentives for politicians and voters, and ultimately, policy outcomes.\(^1\)

Empirical evidence since the 1940’s, however, suggests there is only a limited role for learning by voters. The minimum effects hypothesis, (\([114]\), \([22]\) and, more recently, \([101]\)) finds only small changes to voter behaviour when they are exposed to more information.\(^2\) These results are hard to reconcile with observed behaviour, suggestive that voters must respond to information. The latter includes the expenditures in advertisement, focus groups, polling and debates by political campaigns, which aim at informing the electorate, together with previously cited research.

In this chapter, I provide an answer to this apparent contradiction. I use a structural model of information acquisition, which highlights how the timing in which information is acquired explains both sets of results. Some voters may not use new information that

---

\(^1\) These include the model of \([68]\), on information and turnout, models of learning about candidates (\([5]\), \([80]\)) and recent empirical evidence of learning in primaries (\([60]\), \([108]\)) and elections (\([103]\)).

\(^2\) As \([101]\) conclude from their meta-study of experimental evidence on the effects of campaigning information on voters’ decisions, “the best estimate for the persuasive effects of campaign contact and advertising...is zero”. The two exceptions that the authors find are that persuasive effects may exist if (i) information is introduced months before election day, although that effect decays, and (ii) an election day effect may exist when campaigns spend disproportionately targeting persuadable voters in the presence of extremist views.
becomes available, as they may have decided on their candidate long before election day. In this case, they may appear to have persistent preferences and not respond to information. Other voters might choose to use that new information to update their beliefs as they are still undecided. Both the voting decision and its timing are observed with the survey data used in this paper. This allows a model-based empirical specification that can capture both the timing of decisions (and how that presents evidence of learning or changing of beliefs) as well as the voting decisions themselves (and the extent to which that information is used to change voting behaviour).

The theoretical model has the key feature of endogenous information accumulation. Voters accumulate information until the cost is larger than the benefit. When that occurs, they decide on whom to vote for. Voters benefit from information as it yields more precise beliefs about the current state of their world, upon which they must choose their candidates. But acquiring information is costly, and as voters learn, the benefit of information decreases. It may decrease until this benefit no longer outweighs the cost. For some voters, this occurs early in the campaign, while for others, it might only be on election day. Such tradeoff implies an optimal stopping time behaviour by voters.

I estimate the model using panel survey data from Israel, which samples the same set of voters at the beginning of the campaign and after election day. Observed outcomes include when they decided on whom to vote for, and for whom they voted. I estimate the model using a simulation-based maximum likelihood approach. The results show that I am able to match several observed features of the timing of voting decisions. These include the dispersion in those decisions across the population (40% of voters know all along who they will vote for, while many others decide in the last week) and the characteristics of late deciders (they are often younger and more moderate). Furthermore, the structural model allows me to disentangle the sources of such timing decisions. For some voters, it is the cost of acquiring information that makes them decide earlier in the campaign. For others, it is having tighter priors, which makes information less valuable.

I find that younger voters have the lowest costs of information. Meanwhile, more educated and politically knowledgeable voters decide earlier because they have tighter priors to begin with. The distribution of information signals is also estimated. I find that information is noisy and gets noisier over time.

Finally, I consider the implications of the results for pre-election blackouts. Pre-election

---

3Information early in the campaign may reach a greater number of voters, as more voters are still undecided, information later may be more salient, and swing the remaining undecided voters towards or away from a candidate.

4See [34] for a survey of the evidence on the characteristics of late deciders.
blackouts are a widely used policy throughout the world ([(2)]) that bans electoral polls or advertisements for a period just before election day. I show how this policy can lead to welfare losses, even if it was designed for “fairness” (i.e. so that all voters would have the same amount of information before deciding). This welfare loss is driven by the ban affecting only the subset of voters who would have acquired additional information during the blackout. I find that the effect of increasing the blackout from 1 day, to 1 week in the empirical context is of 2% of the welfare of the affected voters, and around 0.7% of the welfare averaged across all voters.

This chapter speaks to recent works about the timing of endorsements, learning in elections and campaign effects. [108] also focuses on the effects of the timing of information upon voting decisions. They are interested in how votes in early contests in the US primaries may disproportionately influence outcomes, since they may change the decisions of voters in subsequent states. [60], also considers the impacts of learning via primaries. However, their focus is on how the timing of primaries affects voters’ incentives for coordination. My present work complements both of these papers, as I focus on a wider class of elections that do not have primaries or such publicly observed signals, and show how learning by different subgroups may lead to campaigning effects. This helps rationalize how extant evidence of learning may coexist with the previously mentioned results that voters do not respond to information in randomized experiments (e.g. [101]). Meanwhile, works that have studied the timing of endorsements, such as the theoretical approach in [159] and [47] show evidence that the timing of information does matter. This is evidence that voters respond to information by updating their beliefs.

This work is closely related to a literature in political science that studies the characteristics of late deciders in elections. For example, [73] point out that the timing of voting decisions can play a key role in explaining the lack of campaign effects: who is learning, and when they do so, matters for whether voters will change their decisions. Nonetheless, these papers usually focus on the characteristics of late deciders from comparisons of survey data ([44], [169], [29]) or in the form of logistic or probit regressions (for example, [86] and [73]). Lacking in this literature is a compelling theoretical framework for why this happens. As [34] summarize, even studies with large samples “do little to clarify the situation, as they come to differing conclusions about the character and behavior of these voters”. A notable exception is [124] which emphasizes the importance of rational voters who acquire information. Similar ideas about learning and the impacts on a decision maker who has limited time are present in models of marketing, such as in [141]. Meanwhile, the model by [58] also studied endogenous information acquisition, focusing instead on the
impacts of information acquisition on turnout.

The rest of the chapter is organized as follows. In Section 5.2, I present the structural model of costly information acquisition. In Section 5.3, I discuss how this model can be estimated. In Sections 5.4 and 5.5, I describe the empirical setting and the Data. Sections 5.6 through 5.8 present the Identification, Estimation and Results. Some extensions are discussed in Section 5.9, with the policy implications of the model discussed in Section 5.10. I conclude in Section 5.11. Proofs and additional Tables and Figures are shown in Appendix.

5.2 Model

The model is one of dynamic information acquisition, with an environment reminiscent of [143]. There is a set of voters (citizens) denoted $i = 1, \ldots, N$. Time is discrete, $t = 0, 1, \ldots, T$, with $t = 0$ denotes the beginning of the political campaign, and $T$ the election.

There is an underlying unobserved state of nature $x \in \mathbb{R}$. This can be interpreted as the state of the country, the state of the economy, or whether it is welfare enhancing or not to take up a policy. For simplicity, it is one dimensional, representing a possible projection of the issues voters care about onto a left-right spectrum. It is drawn at the beginning of time from a $\mathcal{N}(0, 1/\tau)$ distribution, where $\tau$ denotes the precision of this distribution. Starting from their priors, voters choose whether (and how much) they want to be informed about the issue at hand.

The realized state $x$ is unobserved to the voter. She can attempt to learn it over time, to make the best informed voting decision. Voters have priors given by $\mathcal{N}(\mu_i, 1/\tau_i)$. I assume $\bar{\tau} = \tau$, denoting that on average, the precision of voters is as nature's.$^5$

At $t = 0$, after $x$ has been drawn, voters can choose whether they are going to take part in the political process (i.e. observe and learn about politics from the campaign). This means that they choose whether to access the learning technology or not. This (access) is deemed costly, parametrized by $\kappa > 0$. We will describe this further.$^6$

If they choose to have access, each voter can then accumulate information from $t = 0, \ldots, T - 1$, by acquiring a new signal at each period. The cost of acquiring a signal (given

---

$^5$Priors can be from previous experiences that make them persist with their choices (or perception of the world), such as due to cognitive dissonance ([140]), due to past experiences from before the elections (in the generational model [20]), from party identification which influences their views of the world ([42]) or from media from other sources or political issues ([1]).

$^6$Allowing one to learn involves signing up for costly subscriptions, for reputation losses or for time commitments. Assuming this cost is not essential for the model, but useful for empirical purposes. It allows me to capture voters who do not care or pay attention at all to campaigning.
they have the technology to do so), is given by $c_i > 0$.

Note that this model assumes that voters start the information accumulation process together, at the beginning of the campaign. They end this process at different times. An alternative set-up of this same model would be one in which voters choose different starting points, but have the same stopping time (election day, $T$). The latter interpretation, however, is inconsistent with the data on the timing of voting decisions. In the data we observe, such as the one used in this paper (e.g. Table 5.1 for Israel), the American National Election Studies and others, the voters choose whether they decided on their candidate at different points of the campaign, relative to the election day. They do not reply whether they started making the decision at different points of the campaign.

5.2.1 Timing

I now review the timing of the model, for convenience to the reader. In the following sections, I then expand on each decision branch by deriving the associated equations that determine what the voter chooses to do.

At $t = 0$, the voter decides whether to have the option of learning. This will be seen by equation (5.6), in a following section. If she does not, she chooses the politician/party that maximizes her utility according to her priors.

If she does allow information, from $t = 0$ to $t = T - 1$ (the day before the election), she chooses whether to acquire information, by solving the (sequential) problem of acquiring a signal, given that the utility at a period is given by equation (5.5), derived in the next section.

The beliefs are updated following Bayes’ rule (see equation (5.3)).

Finally, given the beliefs, at $t = T$ the voter chooses the candidate that maximizes his utility given his beliefs.

5.2.2 Environment - Voters

On election day, $T$, each voter $i$ has quadratic preferences of choosing party $j$ at period $t$ under beliefs about $x$, given by:

$$
\mathbb{E}(u_{i,j,t} | \{e_{i,1},...,e_{i,T}\}) = -\mathbb{E} \left[ (a_j - b_i - x)^2 \big| \{e_{i,1},...,e_{i,T}\} \right] - \sum_{t=0}^{T-1} c_i y_{i,t},
$$

where $a_j$ is the policy proposed by party $j$, $b_i$ is her ideology, $e_{i,t}$ a signal received at
period \( t \) and \( x \) is the state of nature. \( c_i \) is the cost of acquiring a signal; and \( y_{i,t} \) refers to whether a signal was acquired by \( i \) (value of 1 if yes, and 0 if not) at period \( t \). The sum in the second component represents the total amount of signals bought, with \( e_{i,1} \) being a signal received at period 1.

If she decides to take part, a signal about \( x \) can be obtained at each period \( t = 0, \ldots, T - 1 \), by paying the cost \( c_i \) and follows:

\[
e_{i,t} = x + \varepsilon_{i,t},
\]

with \( \varepsilon_{i,t} \sim N(0, \sigma^2) \) and \( \varepsilon_{i,t} \) i.i.d across time and individuals. This signal can represent a new source of information, the processing of information available at that day ([130]), or a compilation of news feed from the day. The costly acquisition can mean simply that it is costly to process or take new information in, even if available freely. The information structure is taken as exogenous.

Beliefs are updated following Bayes’ rule, as in [5], [80]. Let the signal history at period \( t \) be denoted \( (\mathcal{H}_{i,t}) \) and \( i \)'s likelihood of the state being \( x \) represented by \( \mathcal{L}_i(x) \). Then, by Bayes’ rule:

\[
\mathcal{L}_i(x | \mathcal{H}_{i,t} \cup \{e_{i,t+1}\}) \propto \mathcal{L}_i(e_{i,t+1} | \mathcal{H}_{i,t}, x) \mathcal{L}_i(\mathcal{H}_{i,t} | x) \mathcal{L}_{i,0}(x) \quad (5.3)
\]

\[
\mathcal{L}_i(e_{i,t+1}, e_{i,t}, \ldots, e_{i,1} | x) \mathcal{L}_{i,0}(x) \quad (5.4)
\]

with the densities described before and \( \mathcal{L}_{i,0}(x) \) being the likelihood with the priors of \( i \).

Once the sequence of information has been gathered; at the moment of elections (\( T \)), the voter will make a decision on who to vote for.

We begin by reviewing the choice of the voter in the branch in which she accumulates information (or partakes in the technology to).

---

7 Note that here, the voter wants to learn about \( x \). This differentiates the model from other ones of Bayesian learning such as [5], [80], [19] where the voter is learning about the differences in campaign policy. I allow them to be learning, instead, of the issues upon which the policy is being made upon. This is consistent with the data, in Table D.3 that voters seem to attribute more significance in these elections to the economy/security, and not to valence issues.

8 I am empirically supported in this assumption by evidence of updating of information ([103]) and rational choice in voting decisions in the context of private information in elections, such as the evidence about the swing voter’s curse, [91] and [9]. In our model, there are no strategic interactions, so the environment is considerably simpler and would be expected to lead to outcomes close to rational ones [46].
Accumulating Information

Since $x$ is unknown; the voter will have her expected utility of choosing $j$, at moment $T$, given the received signal history $(\mathcal{H}_{i,T})$, given by:

\[
\begin{align*}
\mathbb{E}_i(u_{i,j,T}(y) \mid \mathcal{H}_{i,T}) &= \mathbb{E}_i \left( -(a_j - b_i - x)^2 \mid \mathcal{H}_{i,T} \right) + \sum_{t=0}^{T-1} c_i y_{i,t} \\
&= -(a_j - b_i - \mathbb{E}_i[x \mid \mathcal{H}_{i,T}] )^2 - \mathbb{E}_i[(x - \mathbb{E}_i(x \mid \mathcal{H}_{i,T}))^2 \mid \mathcal{H}_{i,T}] - \sum_{t=0}^{T-1} c_i y_{i,t} \\
&= -(a_j - b_i - \mathbb{E}_i[x \mid \mathcal{H}_{i,T}] )^2 - \text{Var}_i[x \mid \mathcal{H}_{i,T}] - \sum_{t=0}^{T-1} c_i y_{i,t},
\end{align*}
\tag{5.5}
\]

where the expectation is taken according to $i$’s beliefs about $x$, conditional on $\mathcal{H}_{i,T}$; and the second line follows from adding and subtracting the conditional expectation of $x$ inside the quadratic term.

The utility function (and hence, the first term) can be interpreted as a civic duty ([67], [58]) component in the voting: here, the voter gains utility in making the correct choice. This correct choice for her is dependent upon what she knows and her ideology.

Note that the utility, in the second line, is written as the sum of 3 components: (i) the utility from voting in $j$, given the knowledge at $T$; (ii) the (expected) quadratic loss function from trying to estimate $x$, given her information, and (iii) the cost of acquiring another signal.

Term (i) is taken as a state of what the voter would gain from voting for $j$ with her information set at $T$. As in Proposition 8 of [143], only the last two terms impact the decision of information accumulation in the sequential decisions. This is because the first term is not known in advance to the voter: at any period $t$, she expects $\mathbb{E}(x \mid \mathcal{H}_{i,t})$ to be the same as $\mathbb{E}(x \mid \mathcal{H}_{i,t+1})$. This is because the signals are i.i.d., and she cannot anticipate that the mean is different than what she already believes in.

The second and third terms are the ones that determine the information accumulation. Their values in any period $t + 1$ are known at $t$ (as we will show in the next section). These terms have an underlying clear statistical interpretation: the voter tries to estimate $x$ by acquiring signals. Her estimate, given the history of her signals is $\mathbb{E}(x \mid \mathcal{H}_{i,T})$. She is trying to minimize the loss from thinking that the true value is $\mathbb{E}(x \mid \mathcal{H}_{i,T})$ when it is actually $x$. This expected loss can be decreased by acquiring new signals. The gain is exactly the gain in the precision of the beliefs.
Out of the Political Process

If the citizen is not taking part in the political process, then she will choose \( j \) to maximize:

\[
\max_{j \in \{1,\ldots,J\}} -(a_j - b_i - \mu_i)^2,
\]

where I have replaced the prior of \( x \) being \( \mu_i \). Denote as \( \tilde{p} \) as the party that solves the above.

At the beginning of \( t = 0 \) (after \( x \) has been drawn, but before the choice to acquire the first signal). Each voter has to choose whether to be “In” or “Out” of the process. The voter chooses to be “In” if she deems that it is likely she will change her mind on who to vote. This is because the benefits of information are sufficient to pay the cost, \( \kappa > 0 \), of accessing the learning technology.

She will choose “In” if her vote choice at \( T, v_i^*, \) satisfies:

\[
P(v_i^* \neq \tilde{p} \mid m^*) - \kappa > 0, \tag{5.6}
\]

where \( m^* \) is the number of signals she anticipates acquiring. This captures that, although the agent does not know the signals that will come; she can anticipate how much information she will need. If it is unlikely that she will change her mind, even when confronted with all the information she would like to have, then there is no reason to actively take part in costly learning.

5.2.3 Political Parties and Voting

I model political parties by attributing them a policy value \( a_j \in \mathbb{R} \), for \( j = 1,\ldots,J \) where \( J \) is the number of parties. It can be interpreted as either their campaigned policy if the state was \( x \) (such as their economic policy if \( x \) represents the state of the economy). Another interpretation is the ideological position of that party.

This captures that as the citizen learns, she finds what party most suits her beliefs and ideology. I can consider an outside option for those who choose not to vote, but our sample consists mostly of individuals who vote; so I do not consider turnout\(^9\).

\(^9\)Israeli elections, the empirical application of this chapter, have a high turnout, around 65%-70% in general. Furthermore, our data shows that over half of the individuals who suggested they were not going to turnout, ended up voting.
5.2.4 Voting Decisions

At $t = T$, the agent solves:

$$
\max_{j \in \{1, \ldots, J\}} -(a_j - b_i - \mathbb{E}_i[x | \mathcal{H}_i(T)])^2,
$$

(5.7)

where this comes from equation (5.5), noting that at $T$ the last two terms are sunk. I denote the party voted by $i$ as:

$$
v^*_i = \arg\max_{j \in \{1, \ldots, J\}} -(a_j - b_i - \mathbb{E}_i[x | \mathcal{H}_i(T)])^2,
$$

(5.8)

and I denote the optimal timing in which the voter makes her decision (i.e. stops acquiring more information), as $t^*_i$, which is the last period in which a voter acquired information.

I now characterize the solution to the problem. The characterizations of the functional forms and of comparative statics provided next will guide the empirical approach.

5.2.5 Results and Solution

An Overview of the Theoretical Results

The model builds from the need of voters to understand what is truly happening (the state of nature). All voters, starting from their priors, understand how much they do not know and what information can give them. They do not know what the information will be (the values of the signals), but they understand that more information will lead to better decisions.

First, I present two standard results that will help find closed form solutions to the beliefs about $x$, and that will be used to construct the empirical specification. Proofs are presented in Appendix. Assume that at period $t$, agent $i$ has received $m_i$ signals.

**Lemma 5.2.1.** The variance of $i$’s beliefs about $x$ is given by $\text{Var}_i[x | \mathcal{H}_i(t)] = \sigma^2/m_i + \tau_i \sigma^2$ and is decreasing in the number of signals, in $\tau_i$ (the precision of the priors about $x$), and increasing in $\sigma^2$ (the variance of the signal). It also does not depend upon the (realized) values of the signals acquired.

This lemma, a well known statistical result from Normal distributions (see [59]) states that, as the voter has more precise prior beliefs about $x$, so will her posteriors be more precise. More precision of either the signals or the prior lead to more precise posteriors.

Note that this Lemma implies that the values of the signals do not matter for the computation of the variance (and, hence, the decision to acquire information). This follows from
the normal distribution assumption on the signals: the risk of a Normal distribution does not depend upon its values (see [59]). This is important as it implies that the decision of accumulating information does not depend upon the value of the signals themselves, and hence, a voter knows in advance the amount of information she wants to have, as well as the decision trajectory at every moment. If there are no signals, there are no gains in information and, hence, no changes to the optimal policy. The following Lemma formalized this.

**Lemma 5.2.2.** The voter’s optimal decision is characterized by a fixed amount of signal acquisitions such that a signal is acquired if (and only if):

$$\text{Var}[x | \mathcal{H}_i] - \text{Var}[x | \mathcal{H}_i \cup \{e_{i,t+1}\}] > c_i,$$

for each period $t = 0, \ldots, T - 1$, and the optimal number of signals to be acquired is characterized by:

$$m^*_i = \left\lfloor \left( \frac{\sigma^2}{c_i} \right)^{1/2} - \frac{\tau_i \sigma^2}{\sigma^2} \right\rfloor$$

where $\lfloor w \rfloor$ denotes the largest non-negative integer smaller than $w$.

By receiving an additional signal, the agent pays cost $c_i$ and updates her beliefs about $x$ by Bayes’ rule. An additional signal increases the precision of her beliefs by:

$$\text{Var}[x | \mathcal{H}_i] - \text{Var}[x | \mathcal{H}_i \cup \{e_{i,t+1}\}] = \frac{\sigma^2}{m_i + \tau_i \sigma^2} - \frac{\sigma^2}{m_i + 1 + \tau_i \sigma^2}$$

Then, I can substitute the values of the variance to get the following Corollary:

**Corollary 5.2.2.1.** The stopping time for voter $i$ is given by the $m$ such that:

$$m^*_i = \begin{cases} 
0 & \text{if } c_i > \frac{1}{\tau_i} - \frac{\sigma^2}{1+\tau_i \sigma^2} \\
\text{m such that } & \frac{\sigma^2}{m-1+\tau_i \sigma^2} - \frac{\sigma^2}{m+\tau_i \sigma^2} > c_i > \frac{\sigma^2}{m+\tau_i \sigma^2} - \frac{\sigma^2}{m+1+\tau_i \sigma^2} \\
T & \text{if } c_i < \frac{\sigma^2}{T-1+\tau_i \sigma^2} - \frac{\sigma^2}{T+\tau_i \sigma^2}
\end{cases}$$

This result shows that the optimal stopping time for accumulating information does not depend upon the values of the signals, and leads to a closed form solution for the stopping time, similar to [59]. The cost of an extra signal is compared to the precision gains.

As time goes on, voters stop as they acknowledge they know enough (or the benefit of knowing more is little compared to the cost), and are ready to make their decisions on who
to vote. This is done by looking at the best fit of a political party/politician, which balances one’s ideology and one’s belief about the state. Note that, as \( T < \infty \) and information is costly, beliefs need not converge. In fact, in general there will be dispersion of posterior beliefs about \( x \) across voters, even conditional on ideology and cost. After every citizen stops, they may have conflicting views because of the signals they received, leading to dispersion in choices for even seemingly identical individuals (as we see in the data). This can explain why works such as [86], who focus only on controlling for observed characteristics, might find that late deciders have “less predictable” choices.

With the results above, I have fully characterized the optimal stopping problem of information accumulation. Now, I provide a useful characterization of the posterior mean belief, useful to model the voting decision in the next section.

**Lemma 5.2.3.** \( E_{t}(x \mid \mathcal{H}_{t}) = \frac{\sum_{i=1}^{m_{t}} e_{i,t} + \sigma_{t}^{2} \mu_{t}}{m_{t} + \sigma_{t}^{2}} \). It has ambiguous sign in the number of signals, depending upon the value of the history of signals themselves.

This result shows that the expected value of \( x \) is a weighted mean of the signals; with weights that are functions of the precision of the signals and of the prior. As importantly, the term is linear in the values of the signals. The priors become less relevant with more signals. Although I do not observe the values of the signals in the data, the fact that it is a sum of them allows us to use the Normal distribution property of the signals. This means that the conditional belief of a citizen follows a normal distribution.

This Bayesian learning result also accommodates the exceptions for the null/zero effect in [101]. After reviewing all experimental evidence for U.S. general elections, they found no effect for campaigns except when (i) information is introduced months before election day, although that effect decays and is not present on election day, and (ii) an election day effect may exist when campaigns spend disproportionately targeting persuadable voters in the presence of extremist views.

Lemma 5.2.3 and Corollary 5.2.2.1 explains these results. Information introduced early on is received by more voters who are updating their beliefs. Hence, the campaign effects may persist (since they are part of voters’ beliefs through the result in Lemma 5.2.3). However, information introduced on election day might not show up on average effects, since the majority of voters is not accumulating that information and does not update their beliefs with it.

Corollary 5.2.2.1 and Lemma 5.2.3 will allow us to construct the likelihood I will use for the estimation of the model. As signals are i.i.d. across agents and time, Lemmas 5.2.1 and 5.2.3 are valid for every individual.
Finally, we look at the following comparative statics which are relevant to the literature.

**Lemma 5.2.4.** As the cost of acquiring information \( c_i \) increases, the voter does not choose to acquire more information. For a large enough ideology \( |b_i| \), if \( |b_i| \) increases (i.e. the voter is more ideologically extreme), the voter will acquire less information.

This is due to the more extreme voter staying out of the information acquisition. Conditional on acquiring some information, then there is no effect of ideology on the stopping time.

This concurs with the early observations in the literature on issues driving the decisions of late voters, but ideology of the early deciders as seen in [44], as cited by [169]: “(1) Partisan precommitment is sufficient to preclude campaign effects, (2) in the absence of precommitment, those exposed to the campaign will make their decisions primarily on the basis of campaign-specific information”. In the model, partisan precommitment is equivalent to a prior \( \mu_i \) close to that of a specific party, with a large enough precision \( \tau_i \). In the data, many “know all along” who they would vote for, many months before even the campaign was to start.

From [144], we have that ideology matters in the time to stop. This is also true in our data. So I reconcile these two pieces of evidence.

### 5.3 Moving to the Data

In the data, I observe individual characteristics, such as age, education, gender, which I denote by \( z_i \). I observe the ideology, on a one-dimensional spectrum, \( b_i \). I observe the political party positions on this same spectrum, and finally, I also observe the stopping times of voters (when they decided on who to vote), as well as who they voted for.

I make the following commonly used parametric assumption about cost:

\[
\begin{align*}
    c_i &= e^{z_i' \beta + \eta_i},
\end{align*}
\]

(5.10)

where \( z_i \) is a vector of observables (e.g. education, age, media access) which affects the cost of processing and acquiring information; and \( \eta_i \sim \mathcal{N}(0, \sigma_\eta^2) \) i.i.d. across individuals captures unobserved heterogeneity.

This specification is parsimonious, and makes sure that the cost must be positive. Under (5.10), we can calculate the probability of a voter stopping at any period, given the results from the previous section. Given the distribution of signals, one can also calculate the probability of voting for a given party. Denote the probability of voter \( i \) choosing party \( j \)
and stopping at period \( t \) as \( P(v_i = j, t_i = t \mid z_i, x; \theta) \). The explicit form for this is presented in Appendix, as Lemma D.2.1.

Having observed individual decisions on who to vote for and stopping times of decisions, \( \{v_i, t_i\} \), I can construct the likelihood of these choices directly from the lemmas above. We would like to estimate the parameters \( \theta = (\sigma, \sigma_\eta, \beta, \{\mu_i, \tau_i\}) \). \( \tau \) follows from knowing \( \tau_i \).

Since \( x \) is unknown, I integrate it out (as we know it follows a Normal distribution of mean 0, variance \( 1/\tau \), which we denote below as \( F(\cdot) \)). It follows that an individual likelihood, when we observe \( i \) making the choice for \( j \) and \( t \) when accumulating information is:

\[
L_i(\theta; v_i, t_i, z_i) = \left( \int P(v_i = j, t_i = t \mid z_i, x; \theta) dF(x) \right)^{I(v_i = j, t_i = t)} \left( \int P(v_i \neq j, t_i = t \mid z_i, x; \theta) - \kappa > 0 \right),
\]

where the first indicator is 1 if the voting choice and stopping time correspond to the observed one, and the second indicator denotes the choice of “In”. If the voter chooses “Out”, then their voting and stopping choices are trivial (equal to 1 for the best according to their prior and involving no parameters), as seen in Lemma D.2.1.

Since the signals are i.i.d. across individuals, I can write the likelihood as:

\[
L(\theta; v_i, t_i, z_i) = \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{t=0}^{T} \left( \int P(v_i = j, t_i = t \mid z_i, x; \theta) dF(x) \right)^{I(v_i = j, t_i = t)} \left( \int P(v_i \neq j, t_i = t \mid z_i, x; \theta) - \kappa > 0 \right),
\]

where the indicators capture that the vote was for \( j \) and decision was made at \( t \), with choices being made in the information accumulation stage.

Since we do not observe \( x \), but yet it is present in (D.7) and we know its prior distribution, the parameters of the model can be estimated via simulation techniques. I will use Maximum Simulated Likelihood. Before describing the estimation approach, I first discuss the data and context of the empirical application, which will guide the identification.
5.4 Data

The main source of data in this chapter comes from a series of surveys conducted by Israel National Election Studies, Tel Aviv University. These consist of a two-period panel dataset, conducted before and after the election for the same individuals. The stated aim of these surveys is to “investigate voting patterns, public opinion, and political participation in Israel”. Although the same individuals are surveyed before and after the election in each year, the sample is different across years. I will focus on data from 2006. The choice to focus on 2006 comes from the extended number of measures for parties in 2006, the extended number of questions in the 2006 survey compared to the others, as well as an appropriate setting that allows more variation for identification. This will be described more thoroughly in the Identification Section.

This dataset has been used by such work as [17], in the context of understanding voter’s different beliefs about government formation, [87] for the effects of terrorism on electoral outcomes and preferences; and are the basis for the series of books by Arian and Shamir (who were also behind the collection of this data) discussing the election and electoral context for each year (for example, [14]).

The Appendix provides further details on the Israeli political system and the context. It is particularly useful for the data to be from Israel, because of its proportional representation, multiparty and nationwide election. Proportional representation allows us to focus on non-strategic components of the decision, as the utility function of our voters will reflect this as they only wish to vote to the party which most closely represents their interests.

The multiple parties in the sample gives us useful variation in decisions and across the policy spectrum, allowing us to identify key parameters of the model. Finally, the nationwide election means I focus on the learning of issues, and not politicians. This is because the nationwide election presents a closed list of politicians, who run under the banner of the party’s national policy. Table D.3 shows that these were, in fact, the more relevant considerations in voting. These contrast significantly with the data used for research on late deciders, based mostly of US ([169], [34], [44]) or Canada ([73]).

In the pre-election survey, citizens are asked about their views on the state of the country, political preferences, individual characteristics (from education, age, to work status, place of birth, religiosity, gender) and further information on their knowledge and participation of the political process (for example, whether they know who is the Speaker of the Knesset; if they know the threshold to enter government, if they access media, if they are affiliated to a political party). The summary statistics are presented in Table D.1.
In the post-election survey, they are asked about who they voted for, and importantly for this chapter, *when they made their decision on whom to vote*. For 2006, the distribution of the answers to this question are in Figures D.1a - D.1d, and in Table 5.1.

**Table 5.1:** When Did You Finally Decide to Vote for the Party? 2006 survey

<table>
<thead>
<tr>
<th>Timing</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Election Day</td>
<td>278</td>
<td>21</td>
</tr>
<tr>
<td>2-3 days before the Elections</td>
<td>120</td>
<td>9</td>
</tr>
<tr>
<td>A Week before the Elections</td>
<td>117</td>
<td>9</td>
</tr>
<tr>
<td>A Month before the Elections</td>
<td>102</td>
<td>8</td>
</tr>
<tr>
<td>A Few Months before the Elections</td>
<td>127</td>
<td>9</td>
</tr>
<tr>
<td>From the Beginning I knew what I would do</td>
<td>594</td>
<td>44</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,338</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

We can see that there is variation in the timing of decisions, as many voters have decided “All along”, but almost 40% deciding in the last week. It seems that for 2006, 2009 and 2015, there are strong movements and outcomes in the last week. This would suggest that the very competitive elections in 2015 were decided by those deciding in the last week, as Likud wins with a narrow of majority. Meanwhile, the election in 2013 seems to have been decided from the beginning, as most voters already were sure of voting Likud at that time. For 2006, there seems to be large variations in the timing of decisions and across parties.

Regarding the reliability of this answer in survey data, this is deemed appropriate by the literature in similar contexts. [72] showed using data from the 1997 Canadian Election Study, that the answer to the question of when a voter decided on whom to vote for was highly reliable (around 80% accuracy) when compared to the one constructed from the multi-wave version (that asks voting intentions at many moments of the campaign). For Germany, [155] also presented high values of reliability. These contrast with results from the U.S., which has shown much lower results of concordance (between 40% - 60%, as seen in [147],[45]). According to [72], this is due to the different contexts studied. Parliamentary systems (such as Israel, Canada and Germany), have much shorter campaigns when compared to the U.S. In the Israeli case, the campaigns are 3 months long, while the U.S. ones are much longer (often close to 10 months, from the first primaries in January upto election day in November).
5.5 Identification

I wish to identify the following variables \( \{ \tau_i, \mu_i \}_{i \in N}, \beta, \sigma^2, \sigma_\eta, \kappa \). I have constructed the likelihood above, and deem to observe the votes, the timing decisions \( t^*_i \), individual characteristics \( (z_i) \), the party policy positions \( \{a_j\}_{j=1}^J \).

For that, I use a 2-step procedure. First, I begin by identifying all parameters conditional on knowing the priors \( \{ \tau_i, \mu_i \}_{i \in N} \). In the next subsection, I discuss the separate identification of the priors.

Let us focus first on equation (D.7). Note that there are no \( \sigma_\eta, \beta \) in this term (they only show up in equation (D.6)). Conditional on having the same stopping time, if we observe two agents who have the same ideology \( b_i \) in the data, then different voting decisions must be due to the different signals they have received. The dispersion in voters’ choices for those who have the same number of signals and the same stopping time identifies \( \sigma \).

Consider two individuals who are identical in their observables \( z_i \), but who have different stopping times. Then this must be due to individual heterogeneity (and its dispersion across the population). This identifies \( \sigma_\eta \). Intuitively, a different number of signals obtained, for apparently identical individuals must mean they are different in an unobserved way.

Given \( \sigma \), the priors and \( \sigma_\eta \), we can look at equation (D.6). \( \beta \) is the only unknown. It is identified from variations in observables (for example, education), mapping out to differential outcomes in stopping times of accumulating information. Any differences of stopping times (given we have controlled for unobservables), must be coming from the observed characteristics.

Finally, \( \kappa \) is not identified. When an individual chooses not acquire information, she could be doing so because: (i) \( \kappa \) is too high, (ii) the benefit of an additional signal is too low (i.e. \( \sigma^2 \) is too high), (iii) or her cost \( c_i \) is too high. I cannot separately identify whether the cost of choosing into the acquisition process is the one which leads to not acquiring information, or whether \( \kappa \) was low enough that she would want to, but the information was not valuable enough to do so.

However, I can still estimate the parameters \( \theta \) from the likelihood above. This is done conditionally on those who accumulate some information - i.e. decide at some point of the campaign. Note that, if we observe some information accumulation (i.e. the voter did not respond she knew “all along”), she must be in the branch of the game tree of information accumulation. Finally, since it wouldn’t be optimal to pay \( \kappa \) and not accumulate information (as one knows \( m^* \)) in advance, it must be the case that those who accumulate some
information are the set of agents that satisfy the restriction.

So it must be that \( P(v_i^* \neq \tilde{p} \mid t_i^* = t, b_i) - \kappa > 0 \) for the voters with an optimal stopping time \( t_i^* > 0 \). Hence, I can estimate all the other parameters, with the indicator function related to the choice of “In” being 1 if \( t_i^* > 0 \).

Identification of the Priors

The identification arguments in the previous section relied on having identified the priors, \( \{ \mu_i, \tau_i \}_{i=1}^N \). I now proceed with the identification of the priors themselves in a separate step.

The identification of \( \mu_i \) comes directly from the survey data. I use the answer to the question: “In your opinion, what is Israel’s general situation?” This question identifies exactly the content of \( \mu_i \) in the model, which is the median belief about the state of the country. I categorize the qualitative answers to this question (very good, good, so-so, bad, very bad), shown in Table 5.2, into a quantifiable measure on the real line.

<table>
<thead>
<tr>
<th>Answer</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Good</td>
<td>69</td>
<td>4</td>
</tr>
<tr>
<td>Good</td>
<td>434</td>
<td>23</td>
</tr>
<tr>
<td>So-So</td>
<td>740</td>
<td>39</td>
</tr>
<tr>
<td>Not Good</td>
<td>280</td>
<td>15</td>
</tr>
<tr>
<td>Bad</td>
<td>390</td>
<td>20</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,913</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Notes: The table shows the variation in answers (used as first-stage estimates) for the median prior of the state of the country/nature.

The identification of \( \tau_i \) (the precision) is not as immediate. This is because there is no information in the data about how “sure” voters are about the state of the country, the ideal measure for \( \tau_i \). To solve this identification problem, I use an additional source of variation prior to the beginning of the campaign. The intuition is the following. Assume that, before \( t = 0 \) in our model - that is, before the game begins - there were two identical voters with

\(^{10}\)To do so, I begin by setting the median (so-so) to 0, with the others at symmetric intervals of length 1 apart (i.e. 2,1,0,-1,-2).
the same median prior, $\mu_i$. One of those voters is exposed to additional signals before the campaign began, exogenously (i.e. a treatment). The treated voter then has median beliefs, observed in the data at $t = 0$, given by a different prior, $\mu'_i$. This $\mu'_i$ is different than $\mu_i$, the observed prior of the untreated voter, because of the additional information the treated voter received.

This difference, $\mu'_i - \mu_i$, will help us identify the precision of voter beliefs. This is because as voters update their beliefs following Bayes’ rule, the distance $\mu'_i - \mu_i$ is inversely proportional to $\tau_i$. To see this formally, denote $q_i$ as this exogenous amount of signals that this treated voter received. Using a first order linear approximation of $\mu'_i$ around $q_i = 0$ we find that

$$
\mu'_i(q_i) \approx \mu_i + \frac{\mu_i}{\tau_i \sigma^2} q_i.
$$

The lower the precision of prior beliefs, the faster the voter updates beliefs with new information.

The variation I use empirically are terrorist attacks across different regions of the country, before the beginning of the 2006 electoral campaign. I use fatalities of Israeli civilians in terrorist attacks in their cities of residence as an exogenous variation in (pre-election) signals. This is possible for two reasons: (i) the pre-2006 period was during the Second Intifada, with terrorist attacks and fatalities across different parts of Israel, and (ii) I can observe the city in which voters reside in the data.

As shown by Table 3 in [87], this variation also appears to be exogenous. Terror does not seem to be related to the location of residence. This suggests that demographic conditions or migration patterns are not related to terrorism at the local level. Another threat to identification could be that the timing of terrorist attacks could impact the timing of elections. Further empirical evidence suggests, though, that terrorist attacks do not influence the timing of the elections ([6]).

This variation is related to voters’ beliefs about the state of the country, as shown in Table D.3. The issues that voters care about in that election included security, which is encompassed by the random variable $x$ in the model. Terrorist attacks are informative about

---

11. The new prior $\mu'_i$ is derived by Bayes’ rule on the original prior $\mu_i$ (as in Lemma 5.2.3).
12. The attacks used are those prior to January 2006, the beginning of the campaign. I will focus on attacks in the 2 years before January 2006. These choices follow [87].
13. Data from B’Tselem (e.g. [87]) computes the number of fatalities of Israeli civilians during this period, shown in Table D.4.

136
the state of the country: it informs voters about the state of security and the provision of public goods. Such attacks are also relevant: they not only change informational content, but can shift voters’ actions, as terrorist attacks lead to increased vote share to the right bloc ([23], [87], [105]). In this chapter, the mechanism for these actions are through updates in beliefs.

The likelihood in (5.12) requires knowledge of the priors to be estimated. This suggests we must first recover estimates of the priors. To implement this first step, we must first parametrize prior beliefs as a function of individual characteristics, \( z_i \). This will allow me to recover latent beliefs \( \mu_i \) for those who are treated. I begin by assuming that:

\[
\mu_i = z_i' \delta + v_i \tag{5.14}
\]

\[
\frac{1}{\tau_i} = z_i' \gamma + \xi_i \tag{5.15}
\]

where \( \mathbb{E}(v_i \mid z_i, q_i) = \mathbb{E}(\xi_i \mid z_i, q_i) = 0 \). These conditions are satisfied if the variation is indeed exogenous. I normalize the variance of the signals before the campaign to 1 \( (\sigma_p^2 = 1) \). Substituting into equation (5.13) leads to:

\[
\mu_i'(q_i) = z_i' \delta + (z_i' \delta)(z_i' \gamma)q_i + \tilde{v}_i, \tag{5.16}
\]

where \( \tilde{v}_i = v_i + v_i \xi_i q_i + v_i z_i' \gamma q_i + z_i' \delta \xi_i q_i \), which is a mean zero component. \( q_i \) is the treatment of there being civilian fatalities in the location of \( i \)’s residence between January 2004-January 2006. Equation (5.16) encompasses the intuition described previously. It will also form the basis for the estimation of the priors, together with (5.14)-(5.15).

### 5.6 Estimation

Estimation of the parameters \( \{\mu_i, \tau_i\}, \beta, \sigma, \delta, \gamma \) follows a two-step procedure. In the first stage, I use the data to recover \( \{\mu_i\} \) coming from the individual (mode) answer to the question “In your opinion, what is Israel’s general situation?” which captures exactly the function of \( \mu \) in the model. The answers are \{Very good, Good, So – So, Not good, Bad\}. I take the median to be 0, and each answer to be spaced a measure 1 apart from each other. The distribution of this answer (and hence, of \( \mu_i \)) is seen in Table 5.2.

With this measured, I run an OLS regression of equation (5.16), which recovers estimates of \( \delta \) and \( \gamma \). Using equation (5.15), I recover estimates of \( \tau_i \) for all voters. Due to
the complexity of the cross sectional structure of \((z'_i \delta)(z'_i \gamma)\), I focus on a single dimensional variable generated from the first principal component of \(z_i\). This variable corresponds to over 95% of the explanatory power of \(z_i\). This is discussed further in the Appendix.

I then proceed to the second stage of estimating \((\beta, \sigma, \sigma_\eta)\). These remaining parameters are then estimated by Maximum Simulated Likelihood (MSL) on the analogous log-likelihood objective function to (5.12), given by:

\[
\ln L^s(\theta; v_i, t_i, z_i) = \sum_{i=1}^N \sum_{k=1}^J \sum_{t=0}^{T} I[v'_i = j, x'_i = t] I[t'_i > 0] \ln \left( \frac{1}{R} \sum_{r=1}^R P(v_i = j \mid t_i = t, z_i, x'_i; \theta) P(t_i = t \mid z_i; \theta) \right),
\]

(5.17)

where the terms are given by (D.7) and (D.6); \(R\) is the number of draws; and \(x\) is drawn from a Normal distribution with mean 0 and variance \(\tau\). Details about the Optimization routine to estimate the parameters are given in the Appendix.

The simulated part comes from not observing \(x\) in the data. Although I know the distribution of \(x\), since I do not observe its value, I have to integrate it out. One can simulate its realized values by drawing from the assumed Normal distribution.

The MSL estimator for the parameters is consistent, according to [165]; with \(n \to \infty\) and \(R \to \infty\). If \(R\) rises faster than \(\sqrt{n}\), then MSL and ML are equivalent. Note that the objective function is continuous and well defined, as it involves (smoothed) probabilities of voting one candidate (and not a discontinuous indicator function). Monte Carlo simulations have shown that this model and likelihood work well. The remaining parameters \(\delta\) and \(\gamma\) are consistent under the conditions of conditional mean zero errors \(\nu_i, \xi_i\) described.

The data does not have a daily outcome of when each individual decide on who to vote (which I interpret as the stopping of accumulating information). Instead, it provides intervals (last day, 2-3 days before the election, and so forth). The model can remain unaltered, but instead of having that the intervals to compare in (5.2.2) are from \(t\) to \(t + 1\), I adjust it to be the interval between the sets we see in the data. For example, those who stop “around one week before the election” have then stopped earlier than 3 days before the election (the next interval), but after around 1 month (30 days before). I set \(T = 90\) as it refers to the beginning of the campaign. This is the timing between the dissolution of the Knesset, on December 29, to the election on March 28.

For estimation, I have to construct values for the policy dimension of political parties\(^{14}\). This is a one-dimensional variable located in the left-right spectrum (with usual support of

\(^{14}\) Denoted there as \(a_j\), where \(j = 1, \ldots, J\) are the political parties.
0-10, with 0 being left, and 10, right). To construct this value, I create 2 different measures and show that they give very similar results. For the first, I use the average values of the answers to the question of where the voter locates the parties in the spectrum, for the 2006 survey data.\textsuperscript{15} They represent the average perceived location of each political party by the sample.\textsuperscript{16} My second measure is from the Duke Accountability and Linkages Project for the political parties in Israel (see [107]). These measures are computed from experts’ answers to surveys about political parties. I use the variable “dw”, which stands for overall left-right placement. Table 5.3 shows these measures and that they are highly correlated with each other, as well as with another measure constructed from the 2009 survey from the Israel National Election Studies.

\textsuperscript{15} This question is only presented in these two years.
\textsuperscript{16} The surveys for 2013 and 2015 do not ask this question. The data is also only available for the 8 “major” parties, although these represent the large majority of votes. In the estimation, I will focus on the parties that we do have data, as seen in Table D.2, which are a large majority of the total number of votes.
Table 5.3: Different Measurements for Policy Vectors

<table>
<thead>
<tr>
<th></th>
<th>2006 Computation</th>
<th>2009 Computation</th>
<th>DALP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kadima</td>
<td>4.987</td>
<td>5.314</td>
<td>5.135</td>
</tr>
<tr>
<td>Likud</td>
<td>6.824</td>
<td>7.188</td>
<td>7.800</td>
</tr>
<tr>
<td>Labor</td>
<td>3.615</td>
<td>4.053</td>
<td>3.400</td>
</tr>
<tr>
<td>Israel Beiteinu</td>
<td>7.003</td>
<td>7.792</td>
<td>8.985</td>
</tr>
<tr>
<td>Mafdal</td>
<td>7.030</td>
<td>7.003</td>
<td>8.942</td>
</tr>
<tr>
<td>UTJ</td>
<td>N/A</td>
<td>N/A</td>
<td>7.274</td>
</tr>
<tr>
<td>Shas</td>
<td>6.119</td>
<td>6.517</td>
<td>7.433</td>
</tr>
<tr>
<td>Meretz</td>
<td>2.378</td>
<td>3.500</td>
<td>1.158</td>
</tr>
</tbody>
</table>

Notes: The table presents different computed policy measures (\(\{a_j\}_{j \in J}\) in our model) that are used in the estimation. The first column uses the average across the surveyed population (in the 2006 Pre-Election Survey, in our main data) for the position of each political party on the Left (0) to Right (10) spectrum. The second column repeats this measure for the 2009 Pre-Election Survey. (These questions are unavailable for 2013 and 2015). The third column presents the measures on the left-right spectrum from the Duke Accountability Project (DALP), [107], computed from an aggregate of specialists’ opinions. This is re-scaled from the 1-10 spectrum to the 0-10 spectrum of the other measures. The correlation between the first and second column is 0.9835; between the first and the third is 0.994, and the second and third is 0.9792. This shows that the choice of measure does not seem to vary much. In Table 5.4 I present results using the DALP measure.

I show estimates that make these sets more flexible: using the exact days in the set, using values between those intervals, and so forth. The results do not significantly change.

5.7 Results

Table 5.4 shows the main set of results. Across specifications, the estimated variance of signals, \(\sigma^2\) and the variance of unobserved cost heterogeneity \(\sigma_\eta\) are estimated to be positive and relatively large. All results are robust across specifications. The standard deviation of signals is estimated to be around 2.5 (or 25% of the support of ideology, which is between \([0, 10]\)). Yet, since we observe the acquisition of information, it must be the case that it is still worth it for those with smaller costs to acquire it. This would mean that campaigns or available information do not inform very precisely the state of the country or issues, but are still useful for learning.

With regards to observable heterogeneity, these are also statistically significant. We can
see that age has a significant and positive coefficient, pointing out that older voters have a larger cost of acquiring/processing information about the political scenario. Education seems to have the inverse sign on the pointwise estimates, implying more educated ones have a smaller cost of acquiring information. However, this is not significant. This indicates that the effects of education are coming through the priors: more educated voters have different priors and initial beliefs than others. This explains the reduced form evidence from Table D.5 that more educated voters stop earlier. Meanwhile, the unobserved heterogeneity implies that the dispersion in stopping times comes not only from observable traits, as age and education, but within those groups there are significant differences.

Non-Hebrew speakers (interviews conducted in another language) and religious voters do not seem to have higher costs. However, in the reduced form results this was the case. This is because all of the correlation in the reduced form, under our structure, is shown to be due to different priors across these groups. All of these results seem robust to the addition of covariates (Column (2)), and changes in the definition of the stopping time dates (Columns (3)-(4)). They are also robust to changing the policy measure to the Duke Accountability Linkages Project one (DALP), as shown in Column (2) of Table 5.5.

Regarding the priors, the estimates of the precisions/variances are shown in Figure 5.1. They show that most voters have priors concentrated around \( \tau = 1 \), although a very few have estimated high precisions. The average precision is 2.35, implying that there is quite a small variance of beliefs on average.

### 5.8 Extensions

#### Allowing for Heteroskedastic Signals

So far, I have assumed that the information comes from a homoskedastic distribution, as \( \sigma^2 \) is the same for every period. The model allows for some variation of this.

We can allow a more general, yet parsimonious time-varying structure:

\[
\sigma^2(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2, \tag{5.18}
\]

for \( t = 1, \ldots, T \).

However, I must impose additional constraints to guarantee that the problem is well defined. The coefficients must be such that \( \sigma^2(t) \geq 0 \quad \forall t \) to guarantee non-negative variance. We must also have that our solution is well defined, which means that the term
Table 5.4: Results of the Structural Model

<table>
<thead>
<tr>
<th></th>
<th>Specification (1)</th>
<th>Specification (2)</th>
<th>Robust Date 1</th>
<th>Robust Date 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$</td>
<td>6.601***</td>
<td>6.713***</td>
<td>6.707***</td>
<td>5.074***</td>
</tr>
<tr>
<td>$\sigma_\eta$</td>
<td>2.517***</td>
<td>2.482***</td>
<td>2.435***</td>
<td>3.040***</td>
</tr>
<tr>
<td>$\beta$ - Constant</td>
<td>-6.98***</td>
<td>-6.781***</td>
<td>-7.175***</td>
<td>-7.421***</td>
</tr>
<tr>
<td>Age</td>
<td>0.038***</td>
<td>0.036***</td>
<td>0.035***</td>
<td>0.0435***</td>
</tr>
<tr>
<td>Education</td>
<td>0.026</td>
<td>0.020</td>
<td>0.017</td>
<td>0.0307</td>
</tr>
<tr>
<td>Gender (Female)</td>
<td>-0.533*</td>
<td>-0.432</td>
<td>-0.406</td>
<td>-0.523</td>
</tr>
<tr>
<td>Language (Arabic)</td>
<td>0.535</td>
<td>0.486</td>
<td>0.752</td>
<td></td>
</tr>
<tr>
<td>Language (Russian)</td>
<td>0.900</td>
<td>0.703</td>
<td>1.515</td>
<td></td>
</tr>
<tr>
<td>Religiosity (observes a little)</td>
<td>0.140</td>
<td>0.091</td>
<td>0.204</td>
<td></td>
</tr>
<tr>
<td>Religiosity (observes a lot)</td>
<td>-0.674</td>
<td>-0.679</td>
<td>-0.808</td>
<td></td>
</tr>
<tr>
<td>Religiosity (observes all of it)</td>
<td>0.230</td>
<td>0.174</td>
<td>0.135</td>
<td></td>
</tr>
<tr>
<td>Rooms Per</td>
<td>-0.180</td>
<td>-0.198</td>
<td>-0.212</td>
<td></td>
</tr>
<tr>
<td>Household Member</td>
<td>0.198</td>
<td>0.197</td>
<td>0.235</td>
<td></td>
</tr>
<tr>
<td>Knowledgeable</td>
<td>0.521</td>
<td>0.498</td>
<td>0.645</td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>819</td>
<td>819</td>
<td>819</td>
<td>819</td>
</tr>
</tbody>
</table>

Standard errors in parentheses, computed by Outer-Product Gradient Approximation.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table presents results across specifications of the main model, presented in Section 5.3. The results use 150 draws from the simulated distribution of the state of nature, $x$, assumed to be $\text{Normal}(0, 1/\tau)$, where $\tau$ is the average of the first stage estimates of $\tau_i$, which equals 2.35. The optimization routine used is shown in the Appendix. The policy measured used for $a$ was the average perceived location of each party according to the voters’ sample in 2006. Column (1) shows a simpler specification, where the cost function only depends on education, age and gender. Column (2) expands this specification, controlling for relevant variables in this setting. Column (3) and Column (4) change our definitions of stopping times: For Column (4), I take the later of the range: “a couple of months before” is 45 days before the election, about a month before is 15 days, about a week is 5 days, and a couple of days before the election is 2 days before. For Column (5), we use 75 days before, 45 days before, 10 days before and 3 days before, with 1 day before for “on the day”. The results are consistent across measures and specifications.
Figure 5.1: Distribution of Estimated Priors (Precision)
Notes: The graph presents the values for the estimated precision of priors ($\tau_i$) and of the associated variance of priors ($1/\tau_i$), estimated in the first stage as described in the Estimation section.

\[
\frac{\sigma^2}{t - 1 + \tau_i \sigma^2} - \frac{\sigma^2}{t + \tau_i \sigma^2}
\]

is non-decreasing in $t$. If that was not the case, then we could have non-monotonic stopping times, with more than one possible solution at each point, as well as an ill-defined
likelihood (as we take log of this term in Equation (5.12)).

The results with heteroskedastic signals under these restrictions are shown in Column 1 of Table 5.5. We can see that even though we allow $\sigma^2(t)$ to be decreasing over time (as long as the differences of the ratios above are decreasing), the results show $\sigma^2(t)' \geq 0$. We see that the coefficient on $\alpha_1$ is positive and significant, indicating that it seems that the signals get noisier closer to the election. The results of the other parameters, though, seem to hold.

This indicates that issue based knowledge is getting sparser as the election draws nearer. This is consistent with the model: voters with the least cost of processing are still willing to incorporate noisier information, as the gain in precision still compensates their low cost. Those deciding in the last day need whatever small informational contribution to help them make the decision, while those with larger costs and more extreme do not need that.

The magnitude is estimated to vary between reasonably precise information at the beginning of the campaign (with $t = 0$), with point estimate close to 0, until the end of the campaign where the standard deviation $\hat{\sigma} = 3$ (or variance of 9, since $90\alpha_1 \approx 9$). This is much noisier than the information received during most of the campaign and only affects the late deciders. Yet, these late deciders, such as swing voters, are fundamental in many of economics models ([68]) and the signals that ultimately swing them are the worse ones available. This inspires the consideration of improvements in information, as will be described in the counterfactual section.

5.8.1 Different Distributions for $\eta$

The model does not rely upon the assumption of the Normal distribution of $\eta$. Indeed, one could replace the Normal CDF in equation (D.6) by an arbitrary CDF $G(\cdot)$, yielding:

$$P(t_i = t | z_i, x; \theta) = \begin{cases} 1 - G((\ln(\frac{\sigma_0^2}{1 + \tau_0^2}) - z_i'\beta)), & \text{if } t = 0 \\ G((\ln(\frac{\sigma_0^2}{1 + \tau_0^2}) - z_i'\beta)) - G((\ln(\frac{\sigma_0^2}{1 + \tau_0^2} - z_i'\beta)), & \text{if } 0 < t < T \\ G((\ln(\frac{\sigma_0^2}{1 + \tau_0^2} - z_i'\beta)), & \text{if } t = T, \end{cases}$$

(5.19)

with almost no changes in the proof. For robustness, I consider a distribution $G$ that is Exponential, with parameter $\lambda$. This is done to show that the signs and significance of the estimates of the cost function are similar than those in the baseline model. Identification follows as before, with the exception that $\eta$ is no longer mean zero. $\lambda$ is now identified from the variance of $\eta$. Column 3 of Table 5.5 shows that the qualitative results still hold.
Table 5.5: Results of Extensions to the Structural Model

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heteroskedastic Signals</td>
<td>a DALP</td>
<td>η ~ Exp(λ)</td>
</tr>
<tr>
<td>σ²</td>
<td>5.692***</td>
<td>6.283***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.098)</td>
<td></td>
</tr>
<tr>
<td>α₀</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α₁</td>
<td>0.106***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α₂</td>
<td>8.565E-07</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.032E-05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ₁</td>
<td>5.146***</td>
<td>2.504***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.175)</td>
<td>(0.289)</td>
<td></td>
</tr>
<tr>
<td>λ</td>
<td>2.810***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.235)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β - Constant</td>
<td>-15.88</td>
<td>-7.069</td>
<td>-9.458***</td>
</tr>
<tr>
<td></td>
<td>(12.029)</td>
<td>(0.931)</td>
<td>(0.580)</td>
</tr>
<tr>
<td>Age</td>
<td>0.069***</td>
<td>0.037***</td>
<td>0.023***</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.010)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Education</td>
<td>-0.050</td>
<td>0.030</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>(0.092)</td>
<td>(0.049)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Gender (Female)</td>
<td>-1.045*</td>
<td>-0.471</td>
<td>-0.117</td>
</tr>
<tr>
<td></td>
<td>(0.585)</td>
<td>(0.294)</td>
<td>(0.243)</td>
</tr>
<tr>
<td>Language (Arabic)</td>
<td>1.535*</td>
<td>0.255</td>
<td>0.770***</td>
</tr>
<tr>
<td></td>
<td>(0.880)</td>
<td>(0.474)</td>
<td>(0.254)</td>
</tr>
<tr>
<td>Language (Russian)</td>
<td>2.621</td>
<td>1.289</td>
<td>1.540**</td>
</tr>
<tr>
<td></td>
<td>(3.820)</td>
<td>(1.568)</td>
<td>(0.766)</td>
</tr>
<tr>
<td>Religiosity (observes a little)</td>
<td>-0.117</td>
<td>0.129</td>
<td>-0.341</td>
</tr>
<tr>
<td></td>
<td>(0.665)</td>
<td>(0.345)</td>
<td>(0.215)</td>
</tr>
<tr>
<td>Religiosity (observes a lot)</td>
<td>-1.316**</td>
<td>-0.682</td>
<td>-0.217</td>
</tr>
<tr>
<td></td>
<td>(0.942)</td>
<td>(0.455)</td>
<td>(0.396)</td>
</tr>
<tr>
<td>Religiosity (observes all of it)</td>
<td>0.00</td>
<td>0.012</td>
<td>0.589</td>
</tr>
<tr>
<td></td>
<td>(1.634)</td>
<td>(0.761)</td>
<td>(0.642)</td>
</tr>
<tr>
<td>Rooms Per</td>
<td>-0.368</td>
<td>-0.209</td>
<td>0.120</td>
</tr>
<tr>
<td>Household Member</td>
<td>(0.407)</td>
<td>(0.206)</td>
<td>(0.091)</td>
</tr>
<tr>
<td>Knowledgeable</td>
<td>0.765</td>
<td>0.606</td>
<td>0.412*</td>
</tr>
<tr>
<td></td>
<td>(0.672)</td>
<td>(0.326)</td>
<td>(0.216)</td>
</tr>
<tr>
<td>N</td>
<td>819</td>
<td>819</td>
<td>819</td>
</tr>
</tbody>
</table>

Standard errors in parentheses, computed by Outer-Product Gradient Approximation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table presents results of the extensions to the baseline model, also using 150 draws for the state of nature, which is still drawn from a Normal distribution with mean 0, precision $\tau = 2.23$ estimated in the first stage. Column (1) presents the results, when heteroskedastic signals are allowed, as we define $\sigma^2(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2$. Column (2) presents the estimated parameters under the policy measure from the Duke Accountability Project. The third column allows $\eta$ to be Exponentially distributed, with parameter $\lambda$. These indicate that the main results of the baseline model are robust to changes in the structure.
5.9 Model Fit and Specification Tests

In this section, I perform exercises related to the model fit.

To see the value of the model fit, I plot the predicted timing of voting decisions with the estimated parameters against the data, in Figure 5.2. I also plot the predicted voting decisions themselves (for simplicity, under the baseline specification), grouped by Left, Centre, and Right Wing Parties (under the prior $x = 0$). We can see that the model fits well both decisions. It captures the key patterns of the timing of voting decisions (the large amount of voters deciding in the beginning, followed by the pattern of less voters deciding until the increase in the last day). Similarly, the predicted vote shares are close to the observed ones.

However, this model cannot capture two patterns, which although not essential, illustrate the limits of a Bayesian learning model. First, the model cannot explain the large number of voters deciding between 2-3 days before the election, compared to a similar mass deciding between a month and a week before the election. This is because the larger time frame for the latter should yield more voters deciding in a Bayesian environment than the amount of voters deciding in the 2-3 days time frame. I attribute this to cues and measurement errors in the answers to these questions. Furthermore, the model cannot capture why 2 parties in the right, that are close to each other, would have significantly different vote shares (given the reported ideologies by voters). In essence, only ideologies matter (and not other dimensions, such as ethnicity, history and salience of certain parties).

5.10 Policy Implications - Pre-election Silence (Blackouts)

The results so far point out that information is quite noisy and voters are very heterogeneous in terms of costs and priors. Altogether, this implies large dispersions in the timing of voting decisions, as well as dispersion in the votes themselves for seemingly identical individuals.

One could then think of whether changes in policy could actually improve decision-making by voters. In the model, this would have to come from changing either the signal structure, or the amount of information available to voters. Blackouts (or pre-election silence) constitute a wide spread policy, as shown in [2]. This policy is a (partial) ban on campaigning or electoral information within a time frame of election day (for example, a day). During this period, it is often the case that candidates may not campaign, polls may not be released and sometimes their names cannot be used in the media.

One motivation for such a policy is a concern for fairness: all voters should have access

---

17 Although reliable, the answers in this time frame are more subject to misreporting, e.g. [73]).
Figure 5.2: Model Fit

Notes: The first graph shows the predicted stopping times under the estimated parameters. The second graph presents the model fit of the timing of voting decisions, as well as the decisions themselves. It shows the average expected vote share after simulating the model 100 times under the estimated parameters of the baseline specification. It aggregates the votes by parties (Left is Meretz and Labor, Centre is Kadima, Right are the remaining parties. Simulation is needed to reproduce the signals that voters receive.

to the same information. This is a prevalent concern, for example, in countries with different time zones, where some information might be available to voters that are not to others at their time of decision. Another, is that noisy information just before the election could lead to undesired changes in voting decisions.

Taking this type of policy to our model, consider a scenario where information accumulation in the last period (i.e. no citizen can buy $e_{i,T}$). For voters whose optimal stopping
time is before $T$, there is no change to their utility or decisions. However, for those who
would have stopped at $T$, they now have to stop at $T - 1$. The difference in their (ex-ante)
utility, is given by Lemma 5.2.2: how much they would have gained by adding another
signal. That is,

$$\mathbb{E}_i^{\text{blackout}}(u_{i,j,T-1}(y) \mid \mathcal{H}_i,T-1) - \mathbb{E}_i(u_{i,j,T-1}(y) \mid \mathcal{H}_i,T-1) = \left( \frac{\sigma^2}{T - 1 + \tau_i \sigma^2} - \frac{\sigma^2}{T + \tau_i \sigma^2} \right) + c_i < 0.$$  

The ex-ante expected welfare loss from this policy is given by:

$$\int_{t_i'=T} \left( \frac{\sigma^2}{T - 1 + \tau_i \sigma^2} - \frac{\sigma^2}{T + \tau_i \sigma^2} + c_i \right) di. \quad (5.20)$$

The blackout policy does not allow a subset of voters, who wanted to learn more before
making a decision, the opportunity to do so. Noisy information, if it is known to be so, will
only be accumulated by those for whom it will give a gain in precision. Hence, even noisy
information can be valuable to some. Banning it can lead to a worse precision of beliefs,
and more often voting decisions that are not optimal (with the additional signal).

Given this, we can calculate what the effect of increasing the 1 day blackout policy to 1
week. To do so, we can just plug-in our estimates in equation (5.20), and compare it. I find
that when looking at the welfare (i.e. dividing equation (5.20) by the utility of the whole
population), we find that this change induces a 0.7% loss in welfare. This number is small
as the effects are only on the subgroup that would have decided between 7 days and 1 day
before the election, and the welfare impact of banning is marginal: those voters already had
all the signals up to 7 days before the election. Nonetheless, if we look at the welfare loss
on that subgroup of late deciders (rather than all voters), the loss is of 2.2%.

5.11 Conclusion

In this chapter, I presented a model of endogenous information acquisition and estimated
it with Israeli data. With the estimates, I showed the importance of the timing in decision
making. The model is able to capture several features of the data, including how different
subgroups decide when and whom to vote for. Furthermore, it is able to separate whether
that is due to the costs of information or to flat priors. We find that subgroups, such as
younger voters, have the smallest costs of information, while other subgroups, such as
religious voters, do not learn due to tight priors to begin with. I then used the model to study
a widely used policy: pre-election blackouts. The theoretical and empirical results illustrate that a policy that might be deemed to promote fairness and equal amounts of information, might actually lead to welfare losses and unequal beliefs. By banning information from voters who still wished to learn more, this generates around 0.7-2% of welfare losses in this empirical setting.

Addressing the issue raised in the Introduction, this model illustrates how it is possible that randomized controlled trials that find no effects of information on voters (e.g. [101]) can coexist with learning by voters. This is due to the timing dimension of voters’ decisions. If information in these experiments are introduced late in the campaign, such as a week before election day, most voters would have already decided. Hence, we should find a 0 average effect of information on voting decisions. Nonetheless, the subgroup of voters still deciding may still be acquiring and using new information. The timing of decision, therefore, matters both for campaigns and for our interpretation of its importance. A Bayesian model of learning is able to match their two cases of positive impacts of information: (i) when information is introduced months before election day - leading to a sizeable set of undecided voters using that information, with a decaying effect due to additional signals later on, and (ii) an effect when campaigns targeting persuadable voters later in the campaign.

In future work, I would like to address some unexplored issues in the model. First, it is likely that voters receive information and learn from neighbours, social networks and friends (such as in [85] and [95]), leading to non-standard convergence of beliefs. Without data on these connections, it is hard to incorporate this into an empirical model, especially when trying to identify correlations of signals between groups of individuals. Second, the model takes a reduced form approach to the supply side of information. We do not model politicians’ incentives to supply information. The empirical results suggest that parties should have different messaging incentives at different points of the campaign: this will depend on the subset of voters still deciding. This is found empirically, as $\sigma^2$ seems to vary across groups of voters and across time. Although one can interpret the information structure available to voters, as the result of some game theoretic outcome of political parties that send messages out to voters, it is a promising avenue for future research.
Bibliography


[16] P. Ban, D. J. Moskowitz, and J. James M. Snyder. The changing relative power of party leaders in congress. mimeo, 2016. → page81


153


[65] C. L. Evans. Congressional whip count database. In College of William and Mary, mimeo (Online), 2012. → pages 83, 97


[69] M. P. Fiorina. The political parties have sorted. *Hoover Institution Essay on Contemporary Politics, Series No. 3*, 2017. → pages 4, 18, 32


161


Appendix A

Appendix to Chapter 2

A.1 Proofs

Lemma A.1.1.

\[ \sum_{j \neq i} s_i s_j m_{ij}(s) = s_i. \]  \hspace{1cm} (A.1)

Proof.

\[ \sum_{j \neq i} s_i s_j m_{ij}(s) = \sum_{j \neq i, j \notin P(i)} s_i s_j m_{ij}(s) + \sum_{j \neq i, j \notin P(i)} s_i s_j m_{ij}(s) \]

\[ = \sum_{j \neq i, j \notin P(i)} s_i s_j \left( p(i) \frac{p(j)}{\sum_{k \in P(i), k \neq i} p(k) s_k} + (1 - p(i)) \frac{(1 - p(j))}{\sum_{k \neq i}(1 - p(k)) s_k} \right) + \]

\[ + \sum_{j \neq i, j \notin P(i)} s_i s_j (1 - p(i)) \frac{1 - p(j)}{\sum_{k \neq i}(1 - p(k)) s_k} \]

\[ = s_i \left( p(i) \left( \frac{\sum_{j \neq i, j \notin P(i)} p(j) s_j}{\sum_{k \neq i, k \in P(i)} p(k) s_k} \right) + (1 - p(i)) \frac{\sum_{j \neq i, j \notin P(i)} (1 - p(j)) s_j}{\sum_{k \neq i}(1 - p(k)) s_k} \right) + \]

\[ + s_i (1 - p(i)) \left( \frac{\sum_{j \neq i, j \notin P(i)} (1 - p(j)) s_j}{\sum_{k \neq i}(1 - p(k)) s_k} \right) \]

\[ = s_i \left( p(i) + (1 - p(i)) \left( \frac{\sum_{j \neq i, j \notin P(i)} \sum_{j \neq i, j \notin P(i)} (1 - p(j)) s_j}{\sum_{k \neq i}(1 - p(k)) s_k} \right) \right) \]

\[ = s_i \left( p(i) + (1 - p(i)) \left( \frac{\sum_{j \neq i} (1 - p(j)) s_j}{\sum_{k \neq i}(1 - p(k)) s_k} \right) \right) \]

\[ = s_i. \]

\[ \square \]
**Proposition 2.2.1**: The limit equilibrium is defined by equations (2.22)-(2.24).

**Proof of Proposition 2.2.1**. Recall that we have from equations (2.17) and (2.16), from the First Order Conditions, that:

\[
c = \frac{\alpha_i}{x_i^i} + \frac{s_i^2}{x_i^i},
\]

(A.2)

and

\[
\frac{s_i^*}{x_i^*} = \phi \sum_{j \neq i} s_j^* m_{ij}(s^*) x_j^*.
\]

(A.3)

We also use that:

\[
x_i^* = \alpha_i X_{P(i)}
\]

(A.4)

\[
s_i^* = \alpha_i S_{P(i)},
\]

(A.5)

for some \(X_{P(i)}, S_{P(i)}\), which comes from the fact that \(\frac{s_i^*}{x_i^*}\) and \(\frac{x_i^*}{\alpha_i}\) are the same for all agents within a party. Let \(P(i) \in \{1, 2\}\) be arbitrary.

Using (A.4) in (2.17) implies:

\[
c = \frac{\alpha_i}{x_i^i} + \frac{s_i^2}{x_i^i} = \frac{\alpha_i}{\alpha_i X_{P(i)}} + \frac{\alpha_i^2 S_{P(i)}^2}{\alpha_i^2 X_{P(i)}^2} = \frac{1}{X_{P(i)}} + \frac{S_{P(i)}^2}{X_{P(i)}^2}.
\]

Multiplying both sides by \(X_{P(i)}^2\) yields:

\[
c X_{P(i)}^2 = X_{P(i)} + S_{P(i)}^2,
\]

(A.6)

which is (2.24).

Let us now substitute (A.4) in (2.16):
\[
\frac{\alpha_s X_p(i)}{X_p(i)} = \phi \sum_{j \neq i} \alpha_j S_p(j) m_{ij}(s^*) \alpha_j X_p(j)
\]

\[
\frac{S_p(i)}{X_p(i)} = \phi \sum_{j \neq i} \alpha_j^2 X_p(j) S_p(j) m_{ij}(s^*)
\]

\[
= \phi \sum_{j \neq i} \alpha_j^2 X_p(j) S_p(j) \left( p(i) \frac{p(j)}{\sum_{k \in P(i), k \neq i} p(k) s_k^*} + (1 - p(i)) \frac{(1 - p(j))}{\sum_{k \neq i} (1 - p(k)) s_k^*} \right) I_{\{j \in P(i)\}}
\]

\[
+ \phi \sum_{j \neq i} \alpha_j^2 X_p(j) S_p(j) \left( (1 - p(i)) \frac{1 - p(j)}{\sum_{k \neq i} (1 - p(k)) s_k^*} \right) I_{\{j \notin P(i)\}}.
\]

Note that for the first two terms, \( p(i) = p(j) \) because they are only summed when \( j \in P(i) \). For the last, \( p(i) \neq p(j) \) as it is summed when \( j \notin P(i) \).

Rewriting the above with this implies:

\[
\frac{S_p(i)}{X_p(i)} = \phi \sum_{j \neq i} \alpha_j^2 X_p(j) S_p(j) \left( p(i) \frac{p(i)}{\sum_{k \in P(i), k \neq i} p(i) s_k^*} + (1 - p(i)) \frac{(1 - p(i))}{\sum_{k \neq i} (1 - p(k)) s_k^*} \right) I_{\{j \in P(i)\}}
\]

\[
+ \phi \sum_{j \neq i} \alpha_j^2 X_p(j) S_p(j) \left( (1 - p(i)) \frac{1 - p(j)}{\sum_{k \neq i} (1 - p(k)) s_k^*} \right) I_{\{j \notin P(i)\}}.
\]

Using that \( s_k^* = \alpha_k S_p(k) \) leads to:

\[
\frac{S_p(i)}{X_p(i)} = \phi \sum_{j \neq i} \alpha_j^2 X_p(j) S_p(j) \left( \frac{p(i)^2}{p(i) \sum_{k \in P(i), k \neq i} \alpha_k S_p(k)} + \frac{(1 - p(i))^2}{\sum_{k \neq i} (1 - p(k)) \alpha_k S_p(k)} \right) I_{\{j \in P(i)\}}
\]

\[
+ \phi \sum_{j \neq i} \alpha_j^2 X_p(j) S_p(j) \left( \frac{(1 - p(i))(1 - p(j))}{\sum_{k \neq i} (1 - p(k)) \alpha_k S_p(k)} \right) I_{\{j \notin P(i)\}}.
\]

Let us focus on the case of \( P(i) = 1 \), as the other case is symmetric.

\[
\frac{S_1}{X_1} = \phi \sum_{j \neq i} \alpha_j^2 X_1 S_1 \left( \frac{p_1}{\sum_{k \in P(i), k \neq i} \alpha_k S_1} + \frac{(1 - p_1)^2}{\sum_{k \neq i} (1 - p(k)) \alpha_k S_p(k)} \right) I_{\{j \in P(i)\}}
\]

\[
+ \phi \sum_{j \neq i} \alpha_j^2 X_1 S_2 \left( \frac{(1 - p_1)(1 - p_2)}{\sum_{k \neq i} (1 - p(k)) \alpha_k S_p(k)} \right) I_{\{j \notin P(i)\}}.
\]

Finally, we use that:
Proof of Proposition 2.2.2.\textbf{ }Recall that an interior equilibrium is a solution to (2.22) to (2.24).
So, rewriting these:

\[
S_1 = X_1 \phi \left( \frac{p_1 B_1 X_1}{A_1} + \frac{(1 - p_1)^2 B_1 S_1 X_1 + (1 - p_1)(1 - p_2) B_2 S_2 X_2}{(1 - p_1) A_1 S_1 + (1 - p_2) A_2 S_2} \right). \tag{A.7}
\]

\[
S_2 = X_2 \phi \left( \frac{p_2 B_2 X_2}{A_2} + \frac{(1 - p_2)^2 B_2 S_2 X_2 + (1 - p_1)(1 - p_2) B_1 S_1 X_1}{(1 - p_1) A_1 S_1 + (1 - p_2) A_2 S_2} \right). \tag{A.8}
\]

\[
cX_1^2 = X_1 + S_1^2, \quad cX_2^2 = X_2 + S_2^2. \tag{A.9}
\]

Substituting (A.7) into (A.9) leads to

\[
cX_1^2 = X_1 + X_1^2 \phi^2 \left( \frac{p_1 B_1 X_1}{A_1} + \frac{(1 - p_1)^2 B_1 S_1 X_1 + (1 - p_1)(1 - p_2) B_2 S_2 X_2}{(1 - p_1) A_1 S_1 + (1 - p_2) A_2 S_2} \right)^2. \tag{A.10}
\]

or

\[
cX_1 = 1 + X_1 \phi^2 \left( \frac{p_1 B_1 X_1}{A_1} + \frac{(1 - p_1)^2 B_1 S_1 X_1 + (1 - p_1)(1 - p_2) B_2 S_2 X_2}{(1 - p_1) A_1 S_1 + (1 - p_2) A_2 S_2} \right)^2. \tag{A.11}
\]

There is a similar expression for $S_2, X_2$. Note that the right hand side of (A.11) lies above the left hand side as we approach $X_1 = 0$ (same for $X_2$). To have an interior solution, we need the right hand side to sometimes fall at or below the left hand side for positive $X_1$.

Suppose that the equilibrium (when it exists) is such that $X_1 \geq X_2$, and the other case is analogous just reversing subscripts everywhere. Then the right hand side is less than what we get by replacing $X_2$ by $X_1$, and so we want

\[
cX_1 \geq 1 + X_1^3 \phi^2 \left( \frac{p_1 B_1}{A_1} + \frac{(1 - p_1)^2 B_1 S_1 + (1 - p_1)(1 - p_2) B_2 S_2}{(1 - p_1) A_1 S_1 + (1 - p_2) A_2 S_2} \right)^2. \tag{A.12}
\]

for some interior $X_1$. Rewriting

\[
cX_1 \geq 1 + X_1^3 \phi^2 \left( \frac{p_1 B_1}{A_1} + \frac{(1 - p_1)^2 B_1 + (1 - p_1)(1 - p_2) B_2 S_2}{(1 - p_1) A_1 + (1 - p_2) A_2 S_2} \right)^2. \tag{A.13}
\]

The right hand side is maximized either at $\frac{S_2}{S_1} = 0$ or $\frac{S_2}{S_1} = \infty$, and so it is sufficient to have

\[
cX_1 \geq 1 + X_1^3 \phi^2 \left( \frac{B_1}{A_1} + (1 - p_1) \max \left[ \frac{B_1}{A_1}, \frac{B_2}{A_2} \right] \right)^2. \tag{A.14}
\]
Let
\[ D_1 = p_1 \frac{B_1}{A_1} + (1 - p_1) \max \left[ \frac{B_1}{A_1}, \frac{B_2}{A_2} \right] \]  
(A.15)

Then (A.14) can be rewritten as
\[ cX_1 \geq 1 + X_1^3 \phi^2 D_1^2. \]  
(A.16)
for some positive \( X_1 \). Note that
\[ D_1 \leq D = \max \left[ \frac{B_1}{A_1}, \frac{B_2}{A_2} \right] \]  
(A.17)

So, it is sufficient to have
\[ cX_1 \geq 1 + X_1^3 \phi^2 D^2. \]  
(A.18)
for some positive \( X_1 \).

It is necessary and sufficient to check that the left hand side and right hand side are tangent at the point at which the slope of the right hand side is \( c \). This happens at \( X_1 = \sqrt{\frac{c}{3\phi^2 D^2}} \) and then the corresponding sufficient condition becomes:
\[ c \left( \frac{c}{3\phi^2 D^2} \right)^{1/2} \geq 1 + \left( \frac{c}{3\phi^2 D^2} \right)^{3/2} \phi^2 D^2, \]  
(A.19)
or
\[ \frac{2c^{3/2}}{3\sqrt{3}} \geq \phi D. \]  
(A.20)

This is the claimed expression.

Proof of Proposition 2.2.3. Assume by way of contradiction that there is \( \{s_i^{FB}, x_i^{FB}\} \) first best that is finite. Consider the allocation \( \{s_i', x_i'\} = \{\lambda s_i^{FB}, \lambda x_i^{FB}\} \), with \( \lambda > 1 \). The later increases all politician’s utility by a cubic rate (from equation (2.13)), while the costs increase quadratically. Hence, \( \{s_i', x_i'\} \) is feasible and yields a higher utility to all agents, which is a contradiction.

168
A.1.1 Best Response Dynamics

Best response dynamics are described as follows. Consider starting at some vectors \( s^0, x^0 \). Then the best response dynamics are described by:

\[
s_t^i = x_t^i - \sum_{j \neq i} m_{ij}(s_{t-1}^j)x_{t-1}^j, \quad (A.21)
\]

and

\[
x_t^i = \frac{\alpha_i}{c} + \frac{s_{t-1}^i}{c} \sum_{j \neq i} m_{ij}(s_{t-1}^j)x_{t-1}^j. \quad (A.22)
\]

It follows that if \( s^0 = 0 \), then \( m_{ij}(s_{t-1}^j) = 0 \) for all \( ij \) (recall Footnote 7) and we get immediate convergence to \( s_t^i = 0, x_t^i = \frac{\alpha_i}{c} \) for all \( t \). Otherwise, \( s_t^i, x_t^i \) will be positive for all \( t \).

To see how these best response dynamics work for a special case, let us consider the situation in which there is some \( S^0, X^0 \) such that \( s_t^0 = \alpha_i S^0 \) and \( x_t^0 = \alpha_i X^0 \) (which has to eventually hold at any limit point).

In that case, working with the limiting or continuum case, in which the matching function is symmetric within a party, and presuming that \( S_{k-1}^t > 0 \) for each party (which happens after the first period if some \( s_0^j > 0 \) and otherwise the solution is already described above), we end up with the following dynamics. For party \( k \) (letting \( k' \) denote the other party):

\[
S_k^t = X_k^t - \frac{1}{c} \phi \left( m_{kk}(S_{k-1}^t)B_k S_{k-1}^t X_{k-1}^t + m_{kk'}(S_{k-1}^t)B_{k'} S_{k-1}^t X_{k-1}^t \right), \quad (A.23)
\]

and

\[
X_k^t = \frac{1}{c} + \frac{s_{k-1}^t}{c} \phi \left( m_{kk}(S_{k-1}^t)B_k S_{k-1}^t X_{k-1}^t + m_{kk'}(S_{k-1}^t)B_{k'} S_{k-1}^t X_{k-1}^t \right). \quad (A.24)
\]

where

\[
m_{kk}(S_{k-1}^t) = \frac{p_k}{S_k A_k} + \frac{(1-p_k)^2}{(1-p_1)S_1 A_1 + (1-p_2)S_2 A_2}, \quad (A.25)
\]

and

\[
m_{kk'}(S_{k-1}^t) = \frac{(1-p_1)(1-p_2)}{(1-p_1)S_1 A_1 + (1-p_2)S_2 A_2}. \quad (A.26)
\]

This is also useful in determining the instability of equilibria.
A.2 Parametric Identification, with set identification of $\phi$

Recall from equation (2.28) that $\alpha_i = e^{\tilde{z}_i^\prime \beta_{P(i)}}$, and that we observe proxies for $(s^*_i, x^*_i)$.

We also have that $m_{ij} = \frac{(1-p_1)(1-p_2)}{A_{P(i)}S_{P(i)}} + \frac{(1-p(i))^2}{(1-p_1)A_1S_1 + (1-p_2)A_2S_2}$ if $i, j$ are from opposing parties and $m_{ij} = \frac{p(i)}{A_{P(i)}S_{P(i)}} + \frac{(1-p(i))^2}{(1-p_1)A_1S_1 + (1-p_2)A_2S_2}$ if they are from the same. Let us denote $m_{ij} = m_{12}$, if $i, j$ belong to different parties, $m_{ij} = m_{11}$ if they both belong to party 1, and $m_{ij} = m_{22}$ if they both belong to party 2.

We now proceed with identification of the parameters of the model. Applying (2.28) and (2.26) in (2.21), for an arbitrary politician $i$ from party $P(i)$ we obtain:

$$\tilde{s}_i e^{\lambda_i} = e^{\tilde{z}_i^\prime \beta_{P(i)}}S_{P(i)},$$

and

$$\log(\tilde{s}_i) = \log(S_{P(i)}) + \tilde{z}_i^\prime \beta_{P(i)} - \lambda_i. \quad (A.28)$$

Since $E\lambda_i = E\tilde{z}_i\lambda_i = 0$, we now have elementary moment conditions (like an OLS) to estimate $\alpha_i^2$. The moment conditions (A.27) for each party identify the respective party specific parameters. To see this more clearly, one can use the moment equations just described to get:

$$E(\log(\tilde{s}_i) - \log(S_{P(i)}) - \tilde{z}_i^\prime \beta_{P(i)}) = 0 \quad (A.29)$$

and

$$E\tilde{z}_i(\log(\tilde{s}_i) - \log(S_{P(i)}) - \tilde{z}_i^\prime \beta_{P(i)}) = 0. \quad (A.30)$$

As long as $E[1, \tilde{z}_i][1, \tilde{z}_i]'$ is invertible, then $\beta_{P(i)}$ and $\log(S_{P(i)})$ are identified. $\beta_{P(i)}$ is identified off the different $\tilde{z}_i$ for members of the same party. $S_{P(i)}$ is identified from the average proxy for effort within a party (the constant in the regression within a party).\footnote{Identification with nonparametric $\alpha$ is proved in a following Appendix.}

Similarly for $\tilde{x}_i$:

$$\tilde{x}_i e^{v_i} = e^{\tilde{z}_i^\prime \beta_{P(i)}}X_{P(i)}$$

and

$$\log(\tilde{x}_i) = \log(X_{P(i)}) + \tilde{z}_i^\prime \beta_{P(i)} - v_i. \quad (A.31)$$

which can be written as another moment condition in terms of the i.i.d. mean zero

\footnote{Implicitly, we require that $\tilde{z}_i$ do not include the constant, as that cannot be separately identified from $\log(S_{P(i)})$ without further assumptions. The average (log) socializing is party-specific.}
random variable $v_i$. $X_{P(i)}$ can be similarly identified for each party.

Since we know $\beta_{P(i)}, S_{P(i)}, X_{P(i)}$ for both parties, $c$ is identified from equation (2.24), where it is the only unknown.

However, $p_1, p_2$ cannot be uniquely identified from the system above. It is clear that $p_1, p_2, \phi$ show up only in the same 2 equations: (2.22) and (2.23). To identify $\phi$, we pursue a set identification approach. To do so, let us first notice that equations (2.22) - (2.23) can be rewritten as:

$$S_1 = \phi X_1 (B_1 S_1 X_1 m_{11} + B_2 S_2 X_2 m_{12})$$  \hspace{1cm} (A.32)

$$S_2 = \phi X_2 (B_2 S_2 X_2 m_{22} + B_1 S_1 X_1 m_{12})$$  \hspace{1cm} (A.33)

To proceed with set identification of $\phi$, we calculate all triples $m_{11}, m_{12}, m_{21}$ that are consistent with some $p_1, p_2 \in [0, 1] \times [0, 1]$. Hence, any $\phi$ that satisfies the above equations for some $(p_1, p_2) \in [0, 1] \times [0, 1]$ is identified.$^5$

A few comments about $m_{ij}(s^*)$ are in order. Even if $m_{ij}(s^*)$ was recovered uniquely, that would not guarantee a unique identification of $p_1, p_2$. There might be multiple $p_1, p_2$ that can yield the same meeting probabilities (usually a continuum of them, governed by a hyperbolic function). Second, $m_{ij}(s^*)$ is not a parameter of interest for us, since it is a normalization on the $g_{ij}$ function that governs the probability of linking.

Having identified all other parameters of the model, we can now prove the identification of the parameters $\gamma, \zeta$. To do so, we use the data on bill passage.

---

$^2$This follows from an alternative rewriting of the first order conditions presented in the model. Namely, we use that:

$$\frac{S_{P(i)}}{X_{P(i)}} = \phi \sum_{j \notin i} \alpha_j^2 S_{P(j)} X_{P(j)} m_{ij}(s^*)$$

$$= \phi \left( \sum_{j \notin i, j \in P(i)} \alpha_j^2 S_{P(j)} X_{P(j)} m_{ij}(s^*) + \sum_{j \notin P(i)} \alpha_j^2 S_{P(j)} X_{P(j)} m_{ij}(s^*) \right)$$

$$= \phi \left( S_{P(i)} X_{P(i)} m_{ij}(s^*) \sum_{j \notin i, j \in P(i)} \alpha_j^2 + S_{P(-i)} X_{P(-i)} m_{12}(s^*) \sum_{j \notin P(i)} \alpha_j^2 \right).$$

$^5$In Appendix, we also provide a solution to point identify $p_1, p_2$, but that relies on additional moment conditions, coming from the second moments of the error terms of the proxy variables $(\tilde{s}_i, \tilde{x}_i)$. However, such a solution requires additional structure outside of the theoretical model.
Identifying the components of $\phi$: $\gamma$ and $\zeta$.

We recall that $\phi = \frac{\zeta \gamma}{m}$, where $\gamma$ was the scale parameter of the distribution of $\varepsilon_i$ (the shock on bill approval, such that the same politician could have different bills passing or not), $m$ was the institutional threshold for approval of a bill (i.e. the minimum amount of support needed for approval), and $\zeta$ was the return to the politician from the voters of having the bill approved. We further assumed that $m = 1$.

If we want to identify the components of $\gamma$ and $\zeta$, we can use the probability of bill approval equation (2.9):

$$P(y_i = 1) = \frac{\gamma}{m} \sum_{j \neq i} g_{ij}(s^*) x^*_i x^*_j.$$  \hspace{1cm} (A.34)

This, in turn, can be rewritten using the First Order Condition for $s_i$ (equation (A.3)) as:

$$P(y_i = 1) = \frac{\gamma}{m} \sum_{j \neq i} g_{ij}(s^*) x^*_i x^*_j = \frac{\gamma}{m} s^*_i x^*_s \sum_{j \neq i} s^*_j x^*_j m_{ij}(s^*) = \frac{\gamma}{m} s^*_i x^*_s \left( \frac{s^*_i}{\phi x^*_i} \right) = \frac{\gamma}{m} \phi x^*_i = \frac{1}{\zeta} s^*_i. \hspace{1cm} (A.35)$$

where $P(y_i = 1)$ is the probability that a bill from politician $i$ is approved.

Since $s^*_i = \tilde{s}_i e^{\lambda_i}$, we can rewrite (A.35) as:

$$P(y_i = 1) = \frac{1}{\zeta} \tilde{s}_i^2 e^{2\lambda_i}. \hspace{1cm} (A.36)$$

Taking logs implies that:

$$logP(y_i = 1) = log \left( \frac{1}{\zeta} \right) + log(\tilde{s}_i^2) + 2\lambda_i. \hspace{1cm} (A.37)$$

Since $\mathbb{E}\lambda_i = 0$, taking expectations on both sides means the only unknown is $\zeta$, which
is identified.

\( \zeta \) must also satisfy the following restriction, due to \( \gamma \) being the scale parameter of the Pareto distributed \( \epsilon_i^6 \):

\[
\frac{m\phi s_i^2}{s_i^2} \geq \gamma, \quad \forall i
\]

\[
\zeta \geq \min_i \alpha_i^2 S_{P(i)}^2
\]

Since \( \phi \) had been previously (set) identified, and \( m = 1 \), then \( \gamma = \phi \zeta \) is (set) identified. This completes the identification of the model. Finally, we note the importance of the measurement errors for identification of this model.

### A.2.1 The need for exponential measurement errors even when \( \alpha \) is parametric.

The exponential form for the measurement errors is very useful for two main reasons. First, we bypass the truncation issue (having to guarantee that \( \tilde{s}_i \geq 0 \) for any \( \lambda \)). Second, it is very tractable (as seen in equation (A.27)).

One could think that measurement errors would not have to show up on \( s_i^* \) and/or \( x_i^* \). However, measurement errors on \( s_i^* \), \( x_i^* \) are needed even with \( \alpha \) being parametric. To see this, consider dividing equation (2.21) by (2.20):

\[
\frac{s_i^*}{x_i^*} = \frac{S_{P(i)}}{X_{P(i)}}.
\]

The \( \alpha \)'s cancel out, and the model is rejected as (A.38) does not hold in the data (without randomness). Hence, there must be an error term in \( s_i^* \) and \( x_i^* \).

---

\footnote{This, in turn, implies that the support for \( \epsilon_i \) is \([\gamma, \infty)\)}
A.3  Rewriting the Model in terms of Moment conditions over $i$

In this section, we provide the derivation for transforming the model from the equations in Proposition 2.2.1 to the moment equations described in section 2.4. We begin by considering the equations derived in the Identification section.

We first note that one could stack-up equations (A.27) and (A.31) across both parties, to get:

\[
\log(\tilde{s}_i) = \log(S_1) + (\log(S_2) - \log(S_1))I_{i \in P_2} + z_i^s \beta_1 + z_i^s I_{i \in P_2}(\beta_2 - \beta_1) - \lambda_i,
\]

\[
\log(\tilde{x}_i) = \log(X_1) + (\log(X_2) - \log(X_1))I_{i \in P_2} + z_i^x \beta_1 + z_i^x I_{i \in P_2}(\beta_2 - \beta_1) - v_i,
\]

where $I_{i \in P_2}$ is an indicator of whether $i$ is in party 2. We have simply introduced the dummy variable for party 2 to stack up the equations. Just like in Ordinary Least Squares, the parameter in front of the indicator recovers the difference of that variable across parties.

Note that the above equations can be rewritten as:

\[
\log(\tilde{s}_i) = \tilde{z}_i^s \beta_s - \lambda_i,
\]

\[
\log(\tilde{x}_i) = \tilde{z}_i^x \beta_x - v_i,
\]

where $\tilde{z}_i = [1, I_{i \in P_2}, z_i^s, z_i^x I_{i \in P_2}], \beta_s = [\log(S_1), \log(S_2) - \log(S_1), \beta_1, \beta_2 - \beta_1], \beta_x = [\log(X_1), \log(X_2) - \log(X_1), \beta_1, \beta_2 - \beta_1]$.

Recall that $E\lambda_i = E\tilde{z}_i \lambda_i = 0$, which is now rewritten together as $E\tilde{z}_i \lambda_i = 0$. Similarly, $Ev_i = E\tilde{z}_i v_i = 0$ is rewritten as $E\tilde{z}_i v_i = 0$.

We are now ready to rewrite the model in terms of moment conditions. Equations (A.39) and (A.40) are rewritten in terms of moments as:

\[
E\tilde{z}_i (\log(\tilde{s}_i) - \tilde{z}_i^s \beta_s) = 0,
\]

\[
E\tilde{z}_i (\log(\tilde{x}_i) - \tilde{z}_i^x \beta_x) = 0.
\]
\begin{align*}
0 &= c - \frac{1}{X_1} \frac{S_1^2}{X_1^3} \\
&= c - \frac{1}{X_1} \frac{s_i^2}{x_i^2} \\
&= c - \frac{1}{X_1} \frac{s_i^2 e^{2(\lambda_i - v_i)}}{x_i^2}.
\end{align*}

Simplifying and taking logs:

\[
\log(c - \frac{1}{X_1}) = 2 \log \left( e^{\lambda_i - v_i} \frac{\tilde{s}_i}{\tilde{x}_i} \right) = 2(\log(\tilde{s}_i) - \log(\tilde{x}_i) + \lambda_i - v_i).
\]

where we have used equations (2.20) and (2.21) and rewritten it in terms of our proxies $(\tilde{s}_i, \tilde{x}_i)$.\footnote{We can replace $X_1$ as well (as a separate equation), and get analogous equations. They will be linearly dependent, as we are already using (2.20) and (2.21) in the estimation, and the previous equation has been introduced.}

Since an analogous version holds for Party 2, one can rewrite the equation above across parties as:

\[
\mathbb{E} \left( 2(\log(\tilde{s}_i) - \log(\tilde{x}_i)) - \log \left( c - \frac{1}{X_1} \right) I_{i \in P_1} - \log \left( c - \frac{1}{X_2} \right) I_{i \in P_2} \right) = 0, \quad (A.43)
\]

where we have taken expectations on both sides over $\lambda_i, v_i$, across all agents. This can be rewritten as equation (2.32) by replacing $I_{i \in P_1} = 1 - I_{i \in P_2}$.

Finally, to rewrite equation (2.33), we simply take expectations over both sides of equation (A.37).

### A.4 Details on Estimation

We now provide further details on how the estimation procedure was implemented. This includes how we generate starting values for the numerical solution to the GMM optimizer, as well as how the Standard Errors were calculated.
A.4.1 OLS and plug-in Approach as Starting Values for Optimization

For the starting values for GMM optimization, we use a simple closed form estimate for the parameters of interest. This new estimator is the OLS estimator for a subset of the parameters, and a plug-in of those for the remaining parameters. Such an estimator is consistent, but inefficient. Yet, it is a good starting value for the optimization procedure.

The inefficiency comes from an OLS approach to equations (2.36) and (2.37) neglecting that $\beta_{Dem}, \beta_{Rep}$ are the same across both equations. To note that we can find this OLS estimator, (2.36) and (2.37) are the moment conditions associated with an OLS problem. The OLS estimator is then given by:

$$\hat{\beta}_s = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\log(\tilde{s}),$$  \hspace{1cm} (A.44)

$$\hat{\beta}_x = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\log(\tilde{x}),$$  \hspace{1cm} (A.45)

where $\tilde{Z}$ is an $N \times k$ matrix of $z_i'$, $\log(\tilde{s})$ and $\log(\tilde{x})$ are the vectors of $\log(\tilde{s}_i)$ and $\log(\tilde{x}_i)$ respectively.

By our definition of $\beta_s, \beta_x$, such an estimator is consistent for $\log(S_1), \log(X_1), \log(S_2), \log(X_2)$ (from the first two parameters in each of $\beta_s, \beta_x$). Similarly, we recover consistent estimates for $\beta_1, \beta_2$ in each equation. One replaces them into either equation (2.24) to recover $c$ (that was not present in any of the other equations).

A.4.2 Computation of the Optimal Weighting Matrix and of Standard Errors

The Optimal Weighting Matrix and Computation of Estimates for the Standard Errors

To compute the standard errors for our GMM estimates, we use a two step procedure that is common in the literature. First, we estimate the model using an inefficient weighting matrix $W = I$, with $I$ the identity matrix, with starting values from the OLS and plug-in approach (described above). Given these inefficient estimates, we then compute the optimal weighting matrix for GMM.

The optimal weighting matrix for GMM is defined as $W = \Omega^{-1}$, where $\Omega = \mathbb{E}(g(\tilde{s}_i, \tilde{x}_i, y_i, z_i, \theta)g(\tilde{s}_i, \tilde{x}_i, y_i, z_i, \theta)'$.

As it is well known, the asymptotic variance matrix (of $\sqrt{n}$ times) our parameters of interest is then given by $(\Gamma\Omega^{-1}\Gamma)^{-1}$, where $\Gamma = \mathbb{E}\frac{\partial g(\tilde{s}_i, \tilde{x}_i, \theta)}{\partial \theta}$.

We compute $\Gamma$ analytically, by taking derivatives of each moment equation in relation
to each parameter. We then replace the expectation by its empirical counterpart (the mean across all politicians).

**Finite Sample Corrections for the Standard Errors**

In finite samples, both $\Omega$ and $(\Gamma'\Omega^{-1}\Gamma')$ can be close to singular. Hence, we provide corrections that improve the finite sample performance.

For $\Omega$, we implement the correction used in [41]. This involves increasing the standard errors in $\Omega$ by adding a small perturbation to its eigenvalues. This perturbation is sufficient to remove singularity.

Such a procedure uses the spectral decomposition of $\Omega = \Lambda D \Lambda'$, where $\Lambda$ is a diagonal matrix of eigenvalues. We then add a small $\delta_\Omega > 0$ to the diagonal of $\hat{\Lambda}$, therefore increasing the eigenvalues of $\hat{\Omega}$. Since this procedure increases standard errors, the new standard errors are still valid for our parameters.

In practice, we pick $\delta_\Omega = 0.0001$, and use it on the eigenvalues that are smaller than $10^{-7}$. This is typically 1 or 2 of the eigenvalues of our estimated $\Omega$. This correction is used for both the calculation of the optimal weighting matrix, as well as for the standard errors of the parameters.

However, such a correction might still be insufficient to guarantee that our estimated variance matrix $\frac{1}{n}(\hat{\Gamma}'\hat{\Omega}^{-1}\hat{\Gamma})^{-1}$ has good finite sample performance, as $\Gamma$ is quite sparse. To solve this, we use a perturbed estimator, that is still consistent.

The estimator we use adds small perturbations to $\Gamma$, in a similar fashion to the correction for $\Omega$. That is, we add a sequence $\delta_{n,\Gamma}$ to the eigenvalues of (the inverse of) our estimated variance matrix, where $\delta_{n,\Gamma} \to 0$ as $n \to \infty$ is a sequence of small perturbations. This estimator is still consistent for our variance matrix. In practice, we pick $\delta_{n,\Gamma} = 10^{-7}$ and replace the eigenvalues that are smaller than $10^{-7}$ by it.

Our results are robust to the choices of $\delta_{n,\Gamma}, \delta_\Omega$.

**A.5 Computation of Comparative Statics**

Our estimates of $S_{Dem}, S_{Rep}, X_{Dem}, X_{Rep}$, are useful for our comparative statics and fitting our model. However, those estimates are not necessarily the values that solve the equilibrium system in Proposition 2.2.1. They are estimates that solve the system in expected value (the moment conditions), as we only observe proxies for social and legislative effort.

---

8The parameters $c$ and $\zeta$, for example, only appear in one moment condition each.
To calculate the values that are consistent with our model, we solve the system in Proposition 2.2.1 using those values as starting points, under the estimated values of \((c, \phi, \beta_{Dem}, \beta_{Rep})\). The solution are values \(S^*_Dem, S^*_Rep, X^*_Dem, X^*_Rep\) that, under the estimated parameters, solve the original game.

For example, when we are interested in the changes in the probability of bill approval, we use \(S_{p(i)}, \alpha_i\) and equation (2.33) to get our results in the following way:

\[
P(y_i = 1) = \frac{1}{\xi} (\alpha_i S_{p(i)})^2. \quad (A.46)
\]
A.6 Identification and Estimation using second moments of the proxies of $(s_t^x, x_t^x)$

In this section, we impose additional structure on the measurement errors of the proxies, denoted $\lambda_i, v_i$, such that we can recover point identification of $p_1, p_2, \phi$. In the following subsection, we then estimate this version of the model.

As before, let us introduce the measurement errors $\lambda_i, v_i$, such that we observe proxies of $(s_t^x, x_t^x)$ of the following form:

\begin{align}
\tilde{s}_i &= s_i^x e^{-\lambda_i} \quad (A.47) \\
\tilde{x}_i &= x_i^x e^{-v_i}. \quad (A.48)
\end{align}

However, we now assume that:

\begin{align}
\lambda_i &= \omega_i p(i) + u_i \quad (A.49) \\
v_i &= \omega_i p(i) + \eta_i \quad (A.50)
\end{align}

with $u_i, \omega_i, \eta_i$ being i.i.d. across $i$ and we have the (standard) assumptions, for all $i$:

\begin{align*}
\mathbb{E}(u_i | z_i, \omega_i, p(i)) &= 0, \\
\mathbb{E}u_i^2 &= \sigma_u^2 \\
\mathbb{E}(\eta_i | z_i, \omega_i, p(i)) &= 0, \\
\mathbb{E}\eta_i^2 &= \sigma_\eta^2 \\
\mathbb{E}(\omega_i | z_i, p(i)) &= 0, \\
\mathbb{E}\omega_i^2 &= \sigma_\omega^2,
\end{align*}

These conditions imply the restrictions $\mathbb{E}(\lambda_i | z_i) = \mathbb{E}(v_i | z_i) = 0$ used in the main text of the chapter.

The structure allows the measurement errors to vary by individual and by party. An intuition for this is that, when partisanship increases, it is harder to observe the true social/legislative effort of a politician. This implies we get noisier proxies for $s_t^x, x_t^x$ with more partisanship. When $p_i = 0$, we only have classical measurement error in the proxies.
With no partisanship, there are no differences in how well we observe socializing across politicians and parties.

The identification arguments presented in the previous approach still hold, up to the identification of $\phi$. Let us continue from there.

Under the assumptions stated above, note that:

\[
\mathbb{E} \lambda_i^2 = \mathbb{E} (\omega_i p(i) + u_i)^2 = \mathbb{E} (\omega_i^2 p(i)^2 + u_i^2 + 2u_i \omega_i p(i)) = p(i)^2 \sigma_\omega^2 + \sigma_u^2. \tag{A.51}
\]

Similarly,

\[
\mathbb{E} v_i^2 = p(i)^2 \sigma_\omega^2 + \sigma_\eta^2 \tag{A.52}
\]

\[
\mathbb{E} (\lambda_i v_i) = p(i)^2 \sigma_\omega^2. \tag{A.53}
\]

We make the following normalization, such that all variances are now written in terms of the variance of $\omega$: $\sigma_\omega = 1^9$. In that case, from (A.53) we get that:

\[
p(i)^2 = \mathbb{E} (\lambda_i v_i). \tag{A.54}
\]

Since we can take expectations over $i \in P_1$ or $i \in P_2$, we get identification of both $p_1$ and $p_2$. Plugging in $p(i)$ into (A.51) identifies $\sigma_u^2$. Plugging $p(i)$ into (A.52) identifies $\sigma_\eta^2$.

Note that the moment conditions given by equations (A.51), (A.52), (A.53) (for each party) are of very simple form to use.

### A.6.1 Estimation with Second Moment Conditions on the Proxies of $s_i^\ast, x_i^\ast$

To estimate this version of the model, we can retain all moment conditions presented in the main part of the text, while adding the moment equations (A.51), (A.52), (A.53), together with a moment equation derived for $\phi$. To derive the latter, equations (2.22) - (2.23) are rewritten together as:

\[
\mathbb{E} (\log(\tilde{s}_i) - \log(\tilde{x}_i) - \log(\phi) - \Psi_0 - \Psi_1 I_i \in P_2) = 0. \tag{A.55}
\]

\footnote{$p_1, p_2$ are always multiplied by $\sigma_\omega$.}
where
\[ \Psi_0 = \log \left( \frac{p_1 B_1 X_1}{A_1} + \frac{(1 - p_1)^2 B_1 S_1 X_1 + (1 - p_1)(1 - p_2) B_2 S_2 X_2}{(1 - p_1)A_1 S_1 + (1 - p_2)A_2 S_2} \right), \]
\[ \Psi_1 = \log \left( \frac{p_2 B_2 X_2}{A_2} + \frac{(1 - p_2)^2 B_2 S_2 X_2 + (1 - p_1)(1 - p_2) B_1 S_1 X_1}{(1 - p_1)A_1 S_1 + (1 - p_2)A_2 S_2} \right) - \Psi_0. \]

Estimation then follows the routines described in the main part of the text, except we now estimate \( \phi, p_1, p_2, \sigma_u^2, \sigma_\eta^2 \).

**A.6.2 Results under Restrictions on the Second Moments of the Proxies for \((s^*_i, x^*_i)\)**

These are presented in Table A.1 below.
Table A.1: Results, Specification 1, second moments of the \((s_i^*, x_i^*)\) proxy

<table>
<thead>
<tr>
<th></th>
<th>105</th>
<th>106</th>
<th>107</th>
<th>108</th>
<th>109</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c)</td>
<td>0.328</td>
<td>0.324</td>
<td>0.360</td>
<td>0.366</td>
<td>0.352</td>
<td>0.334</td>
</tr>
<tr>
<td></td>
<td>(0.0089)</td>
<td>(0.0083)</td>
<td>(0.0102)</td>
<td>(0.0114)</td>
<td>(0.0095)</td>
<td>(0.0085)</td>
</tr>
<tr>
<td>(\phi)</td>
<td>0.0532</td>
<td>0.0572</td>
<td>0.0641</td>
<td>0.0616</td>
<td>0.0611</td>
<td>0.0627</td>
</tr>
<tr>
<td></td>
<td>(0.00207)</td>
<td>(0.00239)</td>
<td>(0.00272)</td>
<td>(0.00382)</td>
<td>(0.00314)</td>
<td>(0.00260)</td>
</tr>
<tr>
<td></td>
<td>(0.9207)</td>
<td>(1.900)</td>
<td>(1.370)</td>
<td>(1.045)</td>
<td>(1.325)</td>
<td>(1.060)</td>
</tr>
<tr>
<td>(p_1)</td>
<td>0.0708</td>
<td>0.0585</td>
<td>0.0684</td>
<td>0.106</td>
<td>0.0555</td>
<td>0.0920</td>
</tr>
<tr>
<td></td>
<td>(0.0231)</td>
<td>(0.0247)</td>
<td>(0.0225)</td>
<td>(0.0222)</td>
<td>(0.0296)</td>
<td>(0.0179)</td>
</tr>
<tr>
<td>(p_2)</td>
<td>0.0804</td>
<td>0.0803</td>
<td>0.116</td>
<td>0.0984</td>
<td>0.113</td>
<td>0.0929</td>
</tr>
<tr>
<td></td>
<td>(0.0203)</td>
<td>(0.0197)</td>
<td>(0.0171)</td>
<td>(0.0203)</td>
<td>(0.0206)</td>
<td>(0.0246)</td>
</tr>
<tr>
<td>(\sigma_\delta^2)</td>
<td>0.228</td>
<td>0.213</td>
<td>0.207</td>
<td>0.204</td>
<td>0.215</td>
<td>0.219</td>
</tr>
<tr>
<td></td>
<td>(0.00537)</td>
<td>(0.00766)</td>
<td>(0.00736)</td>
<td>(0.00617)</td>
<td>(0.00691)</td>
<td>(0.0140)</td>
</tr>
<tr>
<td>(\sigma_\eta^2)</td>
<td>0.1814</td>
<td>0.162</td>
<td>0.192</td>
<td>0.251</td>
<td>0.208</td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td>(0.00374)</td>
<td>(0.00564)</td>
<td>(0.00443)</td>
<td>(0.0263)</td>
<td>(0.0131)</td>
<td>(0.00732)</td>
</tr>
<tr>
<td>(S_{Dom})</td>
<td>0.876</td>
<td>1.054</td>
<td>1.005</td>
<td>0.969</td>
<td>1.010</td>
<td>1.167</td>
</tr>
<tr>
<td></td>
<td>(0.0329)</td>
<td>(0.0362)</td>
<td>(0.0355)</td>
<td>(0.0342)</td>
<td>(0.0313)</td>
<td>(0.0300)</td>
</tr>
<tr>
<td>(S_{Rep})</td>
<td>0.851</td>
<td>0.983</td>
<td>0.936</td>
<td>0.847</td>
<td>0.885</td>
<td>1.005</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.0407)</td>
<td>(0.0503)</td>
<td>(0.0422)</td>
<td>(0.0475)</td>
<td>(0.0595)</td>
</tr>
<tr>
<td>(X_{Dom})</td>
<td>3.420</td>
<td>3.5858</td>
<td>3.152</td>
<td>3.127</td>
<td>3.295</td>
<td>3.757</td>
</tr>
<tr>
<td></td>
<td>(0.130)</td>
<td>(0.126)</td>
<td>(0.117)</td>
<td>(0.131)</td>
<td>(0.113)</td>
<td>(0.0951)</td>
</tr>
<tr>
<td></td>
<td>(0.180)</td>
<td>(0.187)</td>
<td>(0.220)</td>
<td>(0.193)</td>
<td>(0.211)</td>
<td>(0.265)</td>
</tr>
<tr>
<td>Ideology</td>
<td>-0.575</td>
<td>-0.492</td>
<td>-0.714</td>
<td>-0.771</td>
<td>-0.675</td>
<td>-0.465</td>
</tr>
<tr>
<td></td>
<td>(0.0891)</td>
<td>(0.0825)</td>
<td>(0.0841)</td>
<td>(0.0928)</td>
<td>(0.0819)</td>
<td>(0.0699)</td>
</tr>
<tr>
<td>Tenure</td>
<td>0.00374</td>
<td>0.00598</td>
<td>0.00454</td>
<td>0.005245</td>
<td>0.00347</td>
<td>8.415e-05</td>
</tr>
<tr>
<td></td>
<td>(0.00285)</td>
<td>(0.00242)</td>
<td>(0.00276)</td>
<td>(0.00311)</td>
<td>(0.00260)</td>
<td>(0.00220)</td>
</tr>
<tr>
<td>Grosewart</td>
<td>-0.0149</td>
<td>-0.0182</td>
<td>-0.0302</td>
<td>-0.0154</td>
<td>-0.0113</td>
<td>-0.00582</td>
</tr>
<tr>
<td></td>
<td>(0.0095)</td>
<td>(0.0097)</td>
<td>(0.0094)</td>
<td>(0.0114)</td>
<td>(0.0096)</td>
<td>(0.0085)</td>
</tr>
<tr>
<td>Ideology × Rep</td>
<td>0.0734</td>
<td>-0.0255</td>
<td>-0.0713</td>
<td>0.0925</td>
<td>0.121</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>(0.0763)</td>
<td>(0.0699)</td>
<td>(0.0876)</td>
<td>(0.0749)</td>
<td>(0.0786)</td>
<td>(0.0821)</td>
</tr>
<tr>
<td>Tenure × Rep</td>
<td>0.0110</td>
<td>0.00708</td>
<td>0.00487</td>
<td>0.00775</td>
<td>0.00176</td>
<td>-0.00019</td>
</tr>
<tr>
<td></td>
<td>(0.00329)</td>
<td>(0.00271)</td>
<td>(0.00347)</td>
<td>(0.0030)</td>
<td>(0.0031)</td>
<td>(0.0037)</td>
</tr>
<tr>
<td>Grosewart × Rep</td>
<td>-0.0180</td>
<td>-0.0064</td>
<td>-0.0205</td>
<td>-0.0239</td>
<td>-0.0325</td>
<td>-0.0329</td>
</tr>
<tr>
<td></td>
<td>(0.0103)</td>
<td>(0.0100)</td>
<td>(0.0115)</td>
<td>(0.0112)</td>
<td>(0.0104)</td>
<td>(0.0123)</td>
</tr>
<tr>
<td>(N)</td>
<td>437</td>
<td>434</td>
<td>436</td>
<td>438</td>
<td>434</td>
<td>437</td>
</tr>
</tbody>
</table>

Notes: Standard Errors in parentheses. The table presents the results from the GMM estimation under second moment conditions on the proxies of \((s_i^*, x_i^*)\). The optimal weighting matrix is used, and standard errors are estimated as discussed in Appendix. All other notes follow those in Table 2.2.
A.7 Identification with Nonparametric $\alpha$

In this section, we prove that our model is identified even with a nonparametric $\alpha$. This emphasizes that the use of the parametrization of $\alpha$ is not strictly needed theoretically, but is useful for data purposes. This is because a nonparametric $\alpha$ requires us to estimate $\alpha_i$ for each politician in each Congress, when we have only one set of observations per politician per term.

Let us maintain the same error structure as in the Parametric approach in the main text, given in equation (2.26). Let $i_1$ be a normalizer ($\alpha_{i_1} = 1$). Without loss, let him/her be from party 1.

Then, we can rewrite (2.21) for $i_1$ as:

$$s^*_{i_1} = \alpha_{i_1} S_1$$

$$\log(\tilde{s}_{i_1}) + \hat{\lambda}_{i_1} = \log(S_1).$$

Taking expectations on both sides yields:

$$\mathbb{E}\log(\tilde{s}_{i_1}) = \log(S_1),$$

which is now identified. Note that $\mathbb{E}\tilde{s}_{i_1}$ is known from the data, as it is a moment of the observed proxy.

Now using (2.21) for any $i$ in party 1, leads to:

$$\mathbb{E}\log(\tilde{s}_i) = \log(\alpha_i) + \mathbb{E}\log(\tilde{s}_{i_1}),$$

and $\alpha_i$ is identified for any politician in party 1. The intuition is that, in equilibrium, all politicians in party 1 choose social efforts that are proportional to $S_1$. This means that once we normalize someone, we know everyone else’s scale. It follows that $A_1 = \sum_{j \in P_1} \alpha_j$, $B_1 = \sum_{j \in P_1} \alpha_j^2$ are identified as well. This is because we know the party affiliations of each Congress member.

For any politician in party 1, we must also have from (2.20) that:
\[ \log(x_i^*) = \log(\tilde{x}_i) + v_i = \log(\alpha_i) + \log(X_1) \]
\[ \mathbb{E}(\log(\tilde{x}_i)) = \log(\alpha_i) + \log(X_1), \]

where we use that \( v_i \) is mean zero. It follows that \( X_1 \) is identified, and \( S_1/X_1 \) is now identified.

Applying this to equation (2.24) of Party 1 leads to:
\[ c = \frac{1}{X_1} + \left( \frac{S_1}{X_1} \right)^2, \quad \text{(A.58)} \]
implying that \( c \) is identified (since \( S_1, X_1 \) have been identified).

Let us now use equations (2.21) and (2.20) for a party member \( k \) from party 2. (2.21) divided by (2.20) yields:
\[ s_k^* = \frac{X_k}{X_2} S_2 \]
\[ \log(\tilde{s}_k) + \lambda_k = (\log(\tilde{s}_k) + v_k) \log \left( \frac{S_2}{X_2} \right) \]
\[ \mathbb{E}(\log(\tilde{s}_k)) = \mathbb{E}(\log(\tilde{s}_k)) \frac{S_2}{X_2} \]
\[ \quad \text{(A.59)} \]

Hence, \( S_2/X_2 \) is now identified. Using that in equation (2.24) of Party 2:
\[ c - \frac{S_2}{X_2} = \frac{1}{X_2}. \quad \text{(A.60)} \]

Since \( c \) is already known, and so is \( S_2/X_2 \), we have that \( X_2 \) is identified. It follows that \( S_2 \) is identified.

Applying this on equation (2.21) implies that \( \alpha_j \) is identified for all politicians in party 2. It follows that we now know \( A_2, B_2 \), which are functions of \( \alpha_j \) for party 2.

(Set) Identification of \( \phi \) follows the same way as the parametric approach.
Appendix B

Appendix to Chapter 3

B.1 Construction of the Estimator for the Asymptotic Covariance Matrix

We now explain our proposal to estimate the asymptotic covariance matrix, given in equation (3.17) for the model with agents of simple type.

We first explain our proposal to estimate $\Lambda$ consistently for the case of $\beta_0 \neq 0$. Then, we later show how the estimator works even for the case of $\beta_0 = 0$. We first write

$$v_i = R_i(\epsilon) + \eta_i, \quad (B.1)$$

where

$$R_i(\epsilon) = \lambda_{ii} \epsilon_i + \frac{\beta_0 \lambda_{ii}}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij} \epsilon_j.$$ 

Define for $i, j \in N$,

$$e_{ij} = \frac{E[R_i(\epsilon)R_j(\epsilon)|\mathcal{F}]}{\sigma^2_\epsilon},$$

where $\sigma^2_\epsilon$ denotes the variance of $\epsilon_i$. It is not hard to see that for all $i \in N$,

$$e_{ii} = \lambda_{ii}^2 + \frac{\beta_0^2 \lambda_{ii}^2}{n^2_P(i)} \sum_{j \in N_P(i)} \lambda_{ij}^2.$$
and for \( i \neq j \) such that \( N_P(i) \cap N_P(j) \neq \emptyset \), \( e_{ij} = \beta_0 q_{e,ij} \), where

\[
q_{e,ij} = \frac{\lambda_{ij} \lambda_{ii} \lambda_{jj}}{n_p(j)} + \frac{\lambda_{ij} \lambda_{ii} \lambda_{jj}}{n_p(i)} + \frac{\beta_0 \lambda_{ij} \lambda_{jj}}{n_p(i)n_p(j)} \sum_{k \in N_P(i) \cap N_P(j)} \lambda_{ik} \lambda_{jk}.
\]

Thus, we write

\[
\frac{1}{n^*} \sum_{i \in N^*} \mathbb{E}[v_i^2 | \mathcal{F}] = a_{e} \sigma_{e}^2 + \sigma_{\eta}^2, \quad \text{and} \quad (B.2)
\]

\[
\frac{1}{n^*} \sum_{i \in N^*, j \in N_P(i) \cap N^*} \mathbb{E}[v_i v_j | \mathcal{F}] = b_{e} \sigma_{e}^2,
\]

where \( \sigma_{\eta}^2 \) denotes the variance of \( \eta_i \),

\[
a_{e} = \frac{1}{n^*} \sum_{i \in N^*} e_{ii}, \quad \text{and} \quad b_{e} = \frac{1}{n^*} \sum_{i \in N^*, j \in N_P(i) \cap N^*} q_{e,ij},
\]

(Note that since not all agents in \( N_P(i) \) are in \( N^* \) for all \( i \in N^* \), the set \( N_P(i) \cap N^* \) does not necessarily coincide with \( N_P(i) \).) When \( \beta_0 \neq 0 \), the solution takes the following form:

\[
\sigma_{e}^2 = \frac{1}{n^* b_{e}} \sum_{i \in N^*, j \in N_P(i) \cap N^*} \mathbb{E}[v_i v_j | \mathcal{F}] \quad \text{and} \quad (B.3)
\]

\[
\sigma_{\eta}^2 = \frac{1}{n^*} \sum_{i \in N^*} \mathbb{E}[v_i^2 | \mathcal{F}] - a_{e} \frac{n^* b_{e}}{n^* b_{e}} \sum_{i \in N^*, j \in N_P(i) \cap N^*} \mathbb{E}[v_i v_j | \mathcal{F}].
\]

In other words, when \( \beta_0 \neq 0 \), i.e., when there is strategic interaction among the players, we can “identify” \( \sigma_{e}^2 \) and \( \sigma_{\eta}^2 \) by using the variances and covariances of residuals \( v_i \)'s. The intuition is as follows. Since the source of cross-sectional dependence of \( v_i \)'s is due to the presence of \( \epsilon_i \)'s, we can identify first \( \sigma_{e}^2 \) using covariance between \( v_i \) and \( v_j \) for linked pairs \( i, j \), and then identify \( \sigma_{\eta}^2 \) by subtracting from the variance of \( v_i \) the contribution from \( \epsilon_i \).

In order to obtain a consistent estimator of \( \Lambda \) which does not require that \( \beta_0 \neq 0 \), we derive its alternative expression. Let us first write

\[
\Lambda = \Lambda_1 + \Lambda_2, \quad (B.4)
\]

186
where
\[
\Lambda_1 = \frac{1}{n^*} \sum_{i \in N^*} \mathbb{E}[v_i^2 | \mathcal{F}] \tilde{\phi}_i \tilde{\phi}_i', \quad \text{and}
\]
\[
\Lambda_2 = \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N^* \setminus \{i\}} \mathbb{E}[v_i v_j | \mathcal{F}] \tilde{\phi}_i \tilde{\phi}_j',
\]
where \( N^*_{-i} = N^* \setminus \{i\} \). Using (B.1) and (B.3), we can rewrite
\[
\Lambda_2 = \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N^* \setminus \{i\}} e_{ij} \sigma^2_E \tilde{\phi}_i \tilde{\phi}_j
\]
\[
= \frac{\tilde{\beta}_0}{n^*} \sum_{i \in N^*} \sum_{j \in N^* \setminus \{i\}} q_{E,ij} \sigma^2_E \tilde{\phi}_i \tilde{\phi}_j
\]
\[
= \frac{s_E}{n^*} \sum_{i \in N^*} \sum_{j \in N^* \setminus \{i\}} q_{E,ij} \tilde{\phi}_i \tilde{\phi}_j',
\]
where
\[
s_E = \frac{\sum_{i \in N^*} \sum_{j \in N^* \setminus \{i\}} \mathbb{E}[v_i v_j | \mathcal{F}]}{\sum_{i \in N^*} \sum_{j \in N^* \setminus \{i\}} q_{E,ij}}.
\]

Now, it is clear that with this expression for \( \Lambda_2 \), the definition of \( \Lambda \) is well defined regardless of whether \( \tilde{\beta}_0 = 0 \) or \( \tilde{\beta}_0 \neq 0 \). We can then find the estimator of \( \Lambda \), \( \hat{\Lambda} \), by using the empirical analogues to the above, as shown in the main text.
B.2 Proofs of Theorems 3.2.1 - 3.3.1

In this section we provide the proofs of Theorems 3.2.1 - 3.3.1. Throughout the proofs, we use the notation $C_1$ and $C_2$ to represent a constant which does not depend on $n$ or $n^*$. Without loss of generality, we also assume that $N^*$ is $\mathcal{F}$-measurable. This loses no generality because due to Condition A of the sampling process in the chapter, the same proof goes through if we redefine $\mathcal{F}$ to be the $\sigma$-field generated by both $\mathcal{F}$ and $N^*$.

First, we make explicit the best response operator $M$. Given $w_i = (w_k^i)_{k \in N}$, let us define

$$M_iw_i = \frac{1}{np(i)} \sum_{k \in NP(i)} r_{ik}w_k^i \mathbb{1}\{j \in N_i^f(k)\}.$$ 

Recall that player $i$’s payoff is affected by his $G_P$-neighbors’ actions. Hence player $i$ perceives player $j$ as important to him even if player $j$’s action does not directly influence the payoff of player $i$, if player $j$’s type is observed by and influences many of player $i$’s $G_P$-neighbors. The expression $M_iw_i$ captures this perceived importance of player $j$ to player $i$ that comes through player $j$’s influence (as perceived by player $i$) on his $G_P$-neighbors.

Suppose that each agent $i$ has information set $I_i$ for some $k \geq 0$. Then the best response operator $M$ is given by the following relations:

$$w_{i,1} = \gamma_0 + \beta_0 M_iw_{i,1}, \quad (B.5)$$
$$w_{i,e} = 1 + \beta_0 M_iw_{i,e}, \quad w_{i,2} = \beta_0 M_iw_{i,2},$$

and for all $j \in N_i$,

$$w_{ij,1} = \beta_0 M_{ij}w_{ij,1}, \quad (B.6)$$
$$w_{ij,e} = \beta_0 M_{ij}w_{ij,e}, \quad w_{ij,2} = \delta_0 r_{ij}/np(i) + \beta_0 M_{ij}w_{ij,2}, \quad \text{if } j \in NP(i),$$
$$\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \text{if } j \in N_i \setminus NP(i),$$

where we write $w_{ij} = (w_{ij,1}, w_{ij,2}, \epsilon_{ij})'$, $w_{ij,1}, w_{ij,2}$ and $w_{ij,e}$ being weights given by player $i$ to player $j$’s type components, $X_{i,1}, X_{i,2}$ and $\epsilon_{ij}$.

Proof of Theorem 3.2.1: From the optimization of agent $i$, we have

$$S_i^{BR}(\mathcal{I}) = X_i^{1} \gamma_0 + X_i^{2} \delta_0 + \beta_0 \mathbb{E}\left[Y_i^{\beta} | \mathcal{I}_i\right] + \epsilon_i + \eta_i, \quad (B.7)$$

188
where $\hat{Y}_i^B$ denotes the weighted average (over $N_P(i)$) of $s_k(S_k)$ where the weights are given by beliefs of player $i$ and $r_{ik}$. Thus we write

$$E[\hat{Y}_i^B | S_i] = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} \sum_{j \in N_i(i)} w_{kj}^i T_j 1\{j \in N_j^i(k)\}.$$ 

Plugging in this in (B.7), we have

$$s_i^{BR}(S_i) = \left( \gamma_0 + \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} w_{kl}^i X_{i,1} 1\{i \in N_i^k(k)\} \right) X_{i,1}$$

$$+ \left( 1 + \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} w_{kl}^{i,1} 1\{i \in N_i^k(k)\} \right) \varepsilon_i + A_n + B_n,$$

where

$$A_n = \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} w_{kl}^{i,2} X_{i,2} 1\{i \in N_i^k(k)\},$$

$$B_n = \beta_0 \sum_{j \in N_i(i)} \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} (w_{kj,1}^i X_{j,1} + w_{kj}^{i,1} \varepsilon_j) 1\{j \in N_j^i(k)\}$$

$$+ \beta_0 \sum_{j \in N_i(i)} \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} w_{kj,2}^i X_{j,2} 1\{j \in N_j^i(k)\}$$

$$+ \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} \delta_k^i X_{k,2}.$$ 

By setting the coefficients of $X_{j,1}, \varepsilon_j$ and $X_{j,2}$ to be $w_{j,1}^i, w_{j,\varepsilon}^i$ and $w_{j,2}^i$, we obtain that

$$w_{ii,1} = \gamma_0 + \beta_0 \cdot \mathcal{M} w_{i,1}^i, \quad \text{(B.8)}$$

$$w_{ii,\varepsilon} = 1 + \beta_0 \cdot \mathcal{M} w_{i,\varepsilon}^i,$$

$$w_{ii,2} = \beta_0 \cdot \mathcal{M} w_{i,2}^i,$$

and for all $j \in N_i(i)$,

$$w_{jj,1} = \beta_0 \cdot \mathcal{M} w_{j,1}^i, \quad \text{(B.9)}$$

$$w_{jj,\varepsilon} = \beta_0 \cdot \mathcal{M} w_{j,\varepsilon}^i,$$

$$w_{jj,2} = \begin{cases} \delta_j T_j / n_P(i) + \beta_0 \cdot \mathcal{M} w_{j,2}^i, & \text{if } j \in N_P(i), \\ \beta_0 \cdot \mathcal{M} w_{j,2}^i, & \text{if } j \in N_i(i) \setminus N_P(i), \end{cases} \quad \text{(B.10)}$$

$$w_{ij,2} = \begin{cases} \delta_{ij} T_j / n_P(i) + \beta_0 \cdot \mathcal{M} w_{j,2}^i, & \text{if } j \in N_P(i), \\ \beta_0 \cdot \mathcal{M} w_{j,2}^i, & \text{if } j \in N_i(i) \setminus N_P(i), \end{cases} \quad \text{(B.11)}$$
where

\[ \mathcal{M}_i w^j_i = \frac{1}{n_p(i)} \sum_{k \in N_p(i)} r_{ik} w^j_i \{ j \in N'_j(k) \} \]

Now, we apply the behavioral assumptions to this operator to obtain the following:

\[ \mathcal{M}_i w^j_i = \frac{1}{n_p(i)} \sum_{k \in N_p(i)} r_{ik} t^j_i \{ i \in N_p(k) \} \]
\[ = \frac{1}{n_p(i)} \sum_{k \in N_p(i)} r_{ik} t^j_i \]
\[ = \frac{1}{n_p(i)} \sum_{k \in N_p(i)} \frac{1}{n_p(k)} w^j_i = c_{ii} w^j_i. \]

By plugging this into (B.8), we have

\[ w^j_{ii,1} = \gamma_0 + \beta_0 w^j_{ii,1} c_{ii}, \quad (B.12) \]
\[ w^j_{ii,1} = 1 + \beta_0 w^j_{ii,1} c_{ii}, \]

and

\[ w^j_{ii,2} = w^j_{ii,2} \cdot \frac{\beta_0}{n_p(i)} \sum_{k \in N_p(i)} \frac{1\{ i \in N'_j(k) \}}{n_p(k)} \]
\[ = w^j_{ii,2} \cdot \frac{\beta_0}{n_p(i)} \sum_{k \in N_p(i)} \frac{1\{ i \in N_p(k) \}}{n_p(k)} = \beta_0 w^j_{ii,2} c_{ii}. \]

The last equation gives \( w^j_{ii,2} = 0 \), because \( |\beta_0 c_{ii}| < 1 \), and the first two equations give

\[ w^j_{ii,1} = \frac{\gamma_0}{1 - \beta_0 c_{ii}} \quad (B.13) \]

and

\[ w^j_{ii,1} = \frac{1}{1 - \beta_0 c_{ii}}. \quad (B.14) \]

Also, we turn to \( \mathcal{M}_i w^j_i \):

\[ \mathcal{M}_i w^j_i = \frac{1}{n_p(i)} \sum_{k \in N_p(i)} w_{ij}^j \{ j \in N_p(k) \} + \frac{r_{ij} w^j_j}{n_p(i)}, \quad (B.15) \]
where the last term corresponds to the case \( j = k \in N_P(i) \). Using the definition \( c_{ij} \), we rewrite

\[
M_{ij}^w = c_{ij}w_{ij,1} + \frac{w_{ii,1}r_{ij}1\{j \in N_P(i)\}}{n_p(i)} = c_{ij}w_{ij,1} + \frac{\gamma_0}{1 - \beta_0c_{ii}} \frac{1}{n_p(i)} r_{ij}1\{j \in N_P(i)\}
\]

and

\[
M_{ij}^{w,\varepsilon} = c_{ij}w_{ij,\varepsilon} + \frac{w_{ii,\varepsilon}r_{ij}1\{j \in N_P(i)\}}{n_p(i)} = c_{ij}w_{ij,\varepsilon} + \frac{1}{1 - \beta_0c_{ii}} \frac{1}{n_p(i)} r_{ij}1\{j \in N_P(i)\}.
\]

We plug this into (B.9) to obtain

\[
w_{ij,1} = \frac{\beta_0\gamma_0r_{ij}1\{j \in N_P(i)\}}{n_p(i)(1 - \beta_0c_{ij})(1 - \beta_0c_{ii})},
\]

and

\[
w_{ij,\varepsilon} = \frac{\beta_0r_{ij}1\{j \in N_P(i)\}}{n_p(i)(1 - \beta_0c_{ij})(1 - \beta_0c_{ii})}.
\]

Finally, let us consider \( w_{ij,2} \). Note that from (B.15),

\[
M_{ij}^{w,2} = c_{ij}w_{ij,2},
\]

because \( w_{ii,2} = 0 \). By plugging this into (B.9), we obtain that

\[
w_{ij,2} = \begin{cases} \delta_0r_{ij}/n_p(i) + \beta_0c_{ij}w_{ij,2}, & \text{if } j \in N_P(i), \\ \beta_0c_{ij}w_{ij,2}, & \text{if } j \in N_P^2(i) \setminus N_P(i), \end{cases}
\]

where the last zero follows from the equality \( w_{ij,2} = \beta_0c_{ij}w_{ij,2} \) with \( |\beta_0c_{ij}| < 1 \). Therefore, we have

\[
w_{ij,2} = \delta_0r_{ij}1\{j \in N_P(i)\} \frac{1}{n_p(i)(1 - \beta_0c_{ij})}.
\]

From the form of a linear strategy for \( s_i^{BR}(\mathcal{I}) \) with the weights as solved thus far, we obtain the desired result. \( \blacksquare \)

**Proof of Theorem 3.2.2:** Suppose each agent is first-order sophisticated (FS) type; i.e.,
each \(i \in \mathcal{N}\) believes that each \(k \neq i\) is simple type and chooses strategies according to:

\[
s_k^i(\mathcal{I}_{k,0}) = \sum_{j \in \bar{\mathcal{N}}_i(k)} T^i_{kj} w^i_{kj} + \eta_k.
\]

Then \(i\)'s best response takes the form

\[
s^i_{BR,FS}(\mathcal{I}_{i,1}) = \sum_{j \in \bar{\mathcal{N}}_i(i)} T^i_{ij} w^i_{ij} + \eta_i,
\]

where \(\mathbb{E}[\hat{Y}^B_i | \mathcal{I}_{i,1}]\) is written as in the proof of Theorem 3.2.1 noting only that FS has a larger information set and the weights \(w^i_{kj}\) now denote FS-type beliefs. Hence, we can express the best response of FS types as

\[
s^i_{BR,FS}(\mathcal{I}_{i,1}) = \sum_{j \in \bar{\mathcal{N}}_i(i)} T^i_{ij} w^i_{ij} + \eta_i,
\]

where, for each \(i \in \mathcal{N}\), and each \(j \in \bar{\mathcal{N}}_i(i)\), the \(w^i_{ij}\)'s are given according to (B.8) - (B.9) from Theorem 3.2.1. Hence, the best responses of FS types are linear when utility is quadratic and FS types believe simple types play according to linear strategies.

Since an agent \(i\) with FS-type believes that all other agents are of simple type, we have that \(w^i_{kj} = w_{kj}\), where \(w_{kj}\)'s are the weights given to \(j\) when the agent \(k\) is of simple type. The latter weights are already found in Theorem 3.2.1. Therefore, using \(\lambda_{ij} = r_{ij}/(1 - \beta_0 c_{ij})\), together with the Theorem 3.2.1 weights, in equations (B.8) - (B.9), we obtain the FS weights as follows:

\[
w^i_{1,1} = \gamma_0 + \frac{\beta_0}{n_{P}({i})} \sum_{k \in \mathcal{N}_{P}({i})} \frac{\beta_0 \gamma_k r_{ik} \lambda_{k1} 1\{i \in \mathcal{N}_{P}({k})\}}{n_{P}({k})(1 - \beta_0 c_{kk})} 1\{i \in \bar{\mathcal{N}}_i(k)\},
\]

\[
w^i_{1,2} = \frac{\beta_0}{n_{P}({i})} \sum_{k \in \mathcal{N}_{P}({i})} \frac{\beta_0 \gamma_k r_{ik} \lambda_{k1} 1\{i \in \mathcal{N}_{P}({k})\}}{n_{P}({k})(1 - \beta_0 c_{kk})} 1\{i \in \bar{\mathcal{N}}_i(k)\},
\]

\[
w^i_{1,2} = \frac{\beta_0}{n_{P}({i})} \sum_{k \in \mathcal{N}_{P}({i})} \frac{\delta_0 r_{ik} \lambda_{k1} 1\{i \in \mathcal{N}_{P}({k})\}}{n_{P}({k})} 1\{i \in \bar{\mathcal{N}}_i(k)\},
\]
and for each $j \in N_P(i)$,

$$w_{ij,1} = \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} w_{kj,1} \mathbb{1}\{j \in N_P(k)\}$$

$$= \frac{\beta_0}{n_P(i)} \left( \sum_{k \in N_P(i)} w_{kj,1} \mathbb{1}\{j \in N_P(k)\} + w_{jj,1} \mathbb{1}\{j = k\} \right)$$

$$= \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} \beta_0 \gamma_0 r_{ik} \tilde{\lambda}_{kj} \mathbb{1}\{j \in N_P(k)\} + \beta_0 \gamma_0 \mathbb{1}\{j \in N_P(i)\}$$

Analogously,

$$w_{ij,e} = \frac{\beta_0}{n_P(i)} \mathbb{1}\{j \in N_P(i)\} \frac{n_p(i)(1 - \beta_0 c_{jj})}{n_P(i)(1 - \beta_0 c_{jj})},$$

and as for $w_{ij,2}$, if $j \in N_P(i)$

$$w_{ij,2} = \delta_0 r_{ij}/n_P(i) + \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} \frac{r_{ik} \delta_0 \tilde{\lambda}_{kj} \mathbb{1}\{j \in N_P(k)\}}{n_P(k)} \mathbb{1}\{j \in N_I(k)\},$$

and if $j \in N_I(i)\setminus N_P(i)$,

$$w_{ij,2} = \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} \frac{r_{ik} \delta_0 \tilde{\lambda}_{kj} \mathbb{1}\{j \in N_P(k)\}}{n_P(k)} \mathbb{1}\{j \in N_I(k)\}. $$

Next, defining

$$\tilde{\lambda}_{ij} = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} \frac{r_{ik} \tilde{\lambda}_{kj} \mathbb{1}\{j \in N_P(k)\}}{n_P(k)},$$

and

$$\tilde{\lambda}_{ij} = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} \frac{r_{ik} \tilde{\lambda}_{kj} \mathbb{1}\{j \in N_P(k)\}}{n_P(k)(1 - \beta_0 c_{kk})},$$

we may write the weights as

$$w_{ii,1} = \gamma_0 + \beta_0^2 \gamma_0 \tilde{\lambda}_{ii},$$

$$w_{ii,e} = 1 + \beta_0^2 \tilde{\lambda}_{ii},$$

and $w_{ii,2} = \beta_0 \delta_0 \tilde{\lambda}_{ii}$.  

193
Lastly, for each \( j \in N_{P_2}(i) \), we have

\[
 w_{ij,1} = \frac{\beta_0 \gamma_0 \lambda_{ij} 1 \{ j \in N_P(i) \}}{n_P(i)} + \beta_0^2 \gamma_0 \tilde{\lambda}_{ij},
\]

\[
 w_{ij,\varepsilon} = \frac{\beta_0 \lambda_{ij} 1 \{ j \in N_P(i) \}}{n_P(i)} + \beta_0^2 \tilde{\lambda}_{ij},
\]

and

\[
 w_{ij,2} = \begin{cases} 
 \beta_0 \delta_0 \tilde{\lambda}_{ij}, & j \in N_{P_2}(i) \setminus N_P(i) \\
 \delta_0 \gamma_{ij}/n_P(i) + \beta_0 \delta_0 \tilde{\lambda}_{ij}, & j \in N_P(i).
\end{cases}
\]

Substituting these weights back into the best response function for FS types, we obtain the desired result.

We introduce auxiliary lemmas which are used for proving Theorem 3.3.1.

**Lemma B.2.1.** For any array of numbers \( \{a_{ij}\}_{i,j \in N} \) and a sequence \( \{b_i\}_{i \in N} \) of numbers, we have for any subsets \( A, B \subset N \) and for any undirected graph \( G = (N,E) \),

\[
 \sum_{i \in B} \sum_{j \in N(i) \cap A} a_{ij} b_j = \sum_{i \in A} \left( \sum_{j \in N(i) \cap B} a_{ji} \right) b_i,
\]

where \( N(i) = \{ i \in N : i j \in E \} \).

**Proof:** Since the graph \( G \) is undirected, i.e, \( 1 \{ j \in N(i) \} = 1 \{ i \in N(j) \} \), we write the left hand side sum as

\[
 \sum_{i \in B} \sum_{j \in N(i) \cap A} 1 \{ j \in N(i) \} a_{ij} b_j = \sum_{j \in A} \sum_{i \in B} 1 \{ i \in N(j) \} a_{ij} b_j.
\]

Interchanging the index notation \( i \) and \( j \) gives the desired result. ■

**Lemma B.2.2.** Suppose that the conditions of Theorem 3.3.1 hold. Then,

\[
 \Lambda^{-1/2} \frac{1}{\sqrt{n^*}} \sum_{i \in N^*} \phi_{iV} \rightarrow_{d} N(0, I_M).
\]

**Proof:** Choose any vector \( b \in \mathbb{R}^M \) such that \( \|b\| = 1 \) and let \( \tilde{\phi}_i b = b' \tilde{\phi}_i \). Define

\[
 a_i = \lambda_i \phi_{iV} 1 \{ i \in N^* \} + \beta_0 \sum_{j \in N_P(i) \cap N^*} \phi_{jV} \lambda_{ji} \lambda_{jj} / n_P(j).
\]
Using Lemma B.2.1, we can write
\[
\frac{1}{\sqrt{n^*}} \sum_{i \in N^*} \tilde{\varphi}_{i,b} v_i = \sum_{i \in N^*} \xi_i, \tag{B.16}
\]
where we recall \( N^o = \bigcup_{i \in N^*} \overline{N}_p(i), \) and
\[
\xi_i = (a_i \varepsilon_i + \tilde{\varphi}_{i,b} \eta_i 1\{i \in N^*\}) / \sqrt{n^*}.
\]
By the Berry-Esseen Lemma for independent random variables (see, e.g., [157], p.259),
\[
\sup_{t \in \mathbb{R}} \left| P \left\{ \sum_{i \in N^o} \xi_i \sigma_{\xi,i} \leq t \right\} - \Phi(t) \right| \leq \frac{9 \mathbb{E} \left[ \sum_{i \in N^o} |\xi_i|^3 \right]}{(\sum_{i \in N^o} \sigma_{\xi,i}^2)^{3/2}}, \tag{B.17}
\]
where \( \sigma_{\xi,i}^2 = \text{Var}(\xi_i | \mathcal{F}). \) It suffices to show that the last bound vanishes in probability as \( n^* \to \infty. \) First, observe that
\[
\sum_{i \in N^o} \sigma_{\xi,i}^2 = \frac{1}{n^*} \sum_{i \in N^o} (a_i^2 \sigma_{\varepsilon,i}^2 + \tilde{\varphi}_{i,b}^2 \sigma_{\eta,i}^2 1\{i \in N^*\}) \geq \frac{\sigma_{\eta}^2}{n^*} \sum_{i \in N^o} \sigma_{\xi,i}^2 = \sigma_{\eta}^2 > 0,
\]
because \( \frac{1}{n^*} \sum_{i \in N^o} \phi_{i,b}^2 = 1. \) Observe that
\[
\mathbb{E} \left[ \sum_{i \in N^o} |\xi_i|^3 \right] \leq \frac{4 \max_{i \in N^o} \mathbb{E}[|\varepsilon_i|^3 | \mathcal{F}]}{(n^*)^{3/2}} \sum_{i \in N^o} |\tilde{\varphi}_{i,b}|^3 |a_i|^3 \tag{B.18}
\]
\[
+ \frac{4 \max_{i \in N^o} \mathbb{E}[|\eta_i|^3 | \mathcal{F}]}{(n^*)^{3/2}} \sum_{i \in N^o} |\tilde{\varphi}_{i,b}|^3
\]
\[
\leq C_1 \max_{i \in N^o} \mathbb{E}[|\varepsilon_i|^3 | \mathcal{F}] \left( \sum_{i \in N^o} |a_i|^3 \right) + C_1 n^o \max_{i \in N^o} \mathbb{E}[|\eta_i|^3 | \mathcal{F}],
\]
for some constant \( C_1 > 0, \) by Assumption 3.3.3. Now, using the fact that for \( i, j \in N^o \) such that for \( i \neq j, \) \( 0 < \lambda_{ij} \leq C/(1 - \beta_0) \) (which is due to \( 0 \leq c_{ij} \leq 1 \) for all \( i,j \in N^o \) and \( \beta_0 \in (-1,1) \) and by Assumption 3.3.4), and for \( i = j, \) \( 0 < \lambda_i \leq 1/(1 - \beta_0), \) and that
\[
|\tilde{\varphi}_{i,b}| \leq C,
\]
195
for some constant $C > 0$, we bound the leading term as (for some constants $C_2, C_3 > 0$)

\[
\frac{C_2}{n^*} \sum_{i \in N^*} |a_i|^3 \leq \frac{C}{(1 - \beta_0)^3} \frac{1}{n^*} \sum_{i \in N^*} \left( \Phi_{i,b} 1\{i \in N^*\} + \beta_0 \sum_{j \in NP(i) \cap N^*} \frac{\tilde{\Phi}_{j,b}\tilde{\lambda}_{j,i}\tilde{\lambda}_{j,j}}{nP(j)} \right)^3 \leq \frac{C_3}{(1 - \beta_0)^3} + \frac{C_3}{(1 - \beta_0)^6} R_n^3,
\]

where

\[
R_n = \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in NP(i) \cap N^*} \frac{r_{ij}}{nP(j)}.
\]

Using Lemma B.2.1, we rewrite

\[
R_n = \frac{1}{n^*} \sum_{i \in N^*} \frac{1}{nP(i)} \sum_{j \in NP(i)} r_{ij} < C,
\]

for some constant $C > 0$. Hence we find that for some constant $C_1 > 0$,

\[
\frac{1}{n^*} \sum_{i \in N^*} |a_i|^3 \leq \frac{C_1}{(1 - \beta_0)^6}.
\]

Therefore, for some constant $C_2 > 0$,

\[
E \left[ \sum_{i \in N^*} |\xi_i|^3 | F \right] \leq \frac{C_2}{\sqrt{n^*} (1 - \beta_0)^6} \max_{i \in N^*} E[|\xi_i|^3 | F] + \frac{C_2 n^*}{(n^*)^{3/2}} \max_{i \in N^*} E[|\eta_i|^3 | F].
\]

Thus we conclude that the bound in (B.17) is $O_P((n^*)^{-1/2} + n^5 (n^*)^{-3/2})$. However, for some constant $C > 0$,

\[
n^5 \leq \sum_{i \in N^*} |\tilde{N}_P(i)| \leq C n^*,
\]

by Assumption 3.3.4. Hence we obtain the desired result. ■

**Lemma B.2.3.** Suppose that the conditions of Theorem 3.3.1 hold. Then for some constant $C_1 > 0$,

\[
E \left[ ||S_{\Phi^v}||^2 | F \right] = O((n^*)^{-1}), \text{ and } E \left[ ||S_{Z^v}||^2 | F \right] = O((n^*)^{-1}),
\]

where $Z_i^* = \sum_{j \in NP(i) \cap N^*} Z_j$ and $S_{Z^v} = \frac{1}{n^*} \sum_{i \in N^*} Z_i^* v_i$.
Suppose that the conditions of Theorem 3.3.1 hold. Then the following holds.

\[ \mathbb{E} [ ||S_{\tilde{\Phi}_i}||^2 | \mathcal{F} ] \leq \frac{\sigma_\xi^2}{(n^*)^2} \sum_{i \in \mathbb{N}^*} \sum_{j \in \mathbb{N}^* : N_P(i) \cap N_P(j) \neq \emptyset} |\epsilon_{ij}|||\tilde{\Phi}_i|||\tilde{\Phi}_j|| + \frac{1}{(n^*)^2} \sum_{i \in \mathbb{N}^*} (|\epsilon_{ii}| \sigma_\xi^2 + \sigma_\eta^2)|||\tilde{\Phi}_i||^2. \]

However, since \( ||\tilde{\Phi}_i|| \leq C \) by Assumption [3.3.3], the leading term on the right hand side is bounded by for some constants \( C_1, C_2 > 0 \),

\[ \frac{C_1 \sigma_\xi^2 \beta_0}{(n^*)^2 (1 - \beta_0)^d} \sum_{i \in \mathbb{N}^*} \sum_{j \in \mathbb{N}^* : N_P(i) \cap N_P(j) \neq \emptyset} |N_P(i) \cap N_P(j)| \frac{1}{n_P(i) n_P(j)} \leq \frac{C_2}{n^*}, \]

and the second term by \( C \sigma_\eta^2 / n^* \) for some constant \( C > 0 \). Hence the first bound follows.

Let us turn to the second bound. Observe that by Assumption [3.3.3], we have some \( C > 0 \) such that for all \( i \in \mathbb{N}^* \), \( ||Z_i|| \leq C \). Following the same proof as before, we find that \( \mathbb{E} [ ||S_{Z_i^\prime}||^2 | \mathcal{F} ] \) is bounded by

\[ \frac{C_1 \sigma_\xi^2 \beta_0}{(n^*)^2 (1 - \beta_0)^d} \sum_{i \in \mathbb{N}^*} \sum_{j \in \mathbb{N}^* : N_P(i) \cap N_P(j) \neq \emptyset} |N_P(i) \cap N_P(j)||N_P(i) \cap N^*||N_P(j) \cap N^*| \frac{1}{n_P(i) n_P(j)} \]

\[ + \frac{C_1 \sigma_\xi^2}{n^* (1 - \beta_0)^3}. \]

The leading term is bounded by (for some constants \( C_1, C_2, C_3 \))

\[ \frac{C_1 \sigma_\xi^2 \beta_0}{(n^*)^2 (1 - \beta_0)^d} \sum_{i \in \mathbb{N}^*} \sum_{j \in \mathbb{N}^* : N_P(i) \cap N_P(j) \neq \emptyset} |N_P(i) \cap N_P(j)| \leq \frac{C_2 \sigma_\xi^2 \beta_0}{(n^*)^2 (1 - \beta_0)^d} \sum_{i \in \mathbb{N}^*} |N_P(i) \cap N^*| \leq \frac{C_2 \sigma_\xi^2 \beta_0}{n^* (1 - \beta_0)^d}. \]

Thus we obtain the desired result. \( \blacksquare \)

**Lemma B.2.4.** Suppose that the conditions of Theorem [3.3.1] hold. Then the following holds.

(i) \( \frac{1}{n^*} \sum_{i \in \mathbb{N}^*} (v_i^2 - v_i^2) \tilde{\Phi}_i \tilde{\Phi}_i^* = O_P(1/\sqrt{n^*}). \)

(ii) \( \frac{1}{n^*} \sum_{i \in \mathbb{N}^*} \sum_{j \in N_P(i) \cap N^*} (\tilde{\nu}_i \tilde{v}_j - v_i v_j) \tilde{\Phi}_i \tilde{\Phi}_j^* = O_P(1/n^*). \)

(iii) \( \frac{1}{n^*} \sum_{i \in \mathbb{N}^*} (v_i^2 - \mathbb{E} [v_i^2 | \mathcal{F} ] ) \tilde{\Phi}_i \tilde{\Phi}_i^* = O_P(1/\sqrt{n^*}). \)

(iv) \( \frac{1}{n^*} \sum_{i \in \mathbb{N}^*} \sum_{j \in N_P(i) \cap N^*} (v_i v_j - \mathbb{E} [v_i v_j | \mathcal{F} ] ) \tilde{\Phi}_i \tilde{\Phi}_j^* = O_P(1/\sqrt{n^*}). \)

197
Proof: (i) First, write \( \tilde{v} - v = -Z(\tilde{\rho} - \rho_0) \), where \( \tilde{\rho} - \rho_0 = \left[ S_{Z\Phi}S_{Z\Phi}' \right]^{-1} S_{Z\Phi}S_{\phi} \). Hence

\[
\left\| \frac{1}{n^*} \sum_{i \in N^*} (\tilde{v}_i - v_i)^2 \Phi_i \tilde{\Phi}_i' \right\| \leq \frac{C_1}{n^*} \sum_{i \in N^*} (\tilde{v}_i - v_i)^2,
\]

for some constant \( C_1 > 0 \). As for the last term, note that

\[
\frac{1}{n^*} \sum_{i \in N^*} \mathbb{E} \left[ (\tilde{v}_i - v_i)^2 \mid \mathcal{F} \right] \quad \text{(B.19)}
\]

\[
= \frac{1}{n^*} \text{tr} \left( S_{Z\Phi}' [S_{Z\Phi}S_{Z\Phi}']^{-1} S_{ZZ} [S_{Z\Phi}S_{Z\Phi}']^{-1} S_{Z\Phi}A \right) = O_P \left( \frac{1}{n^*} \right),
\]

by Lemma B.2.3. However, we need to deal with

\[
\left| \frac{1}{n^*} \sum_{i \in N^*} (\tilde{v}_i^2 - v_i^2) \right| \leq \sqrt{\frac{1}{n^*} \sum_{i \in N^*} (\tilde{v}_i - v_i)^2} \sqrt{\frac{1}{n^*} \sum_{i \in N^*} (\tilde{v}_i + v_i)^2}. \quad \text{(B.20)}
\]

Note that

\[
\frac{1}{n^*} \sum_{i \in N^*} (\tilde{v}_i + v_i)^2 \leq \frac{2}{n^*} \sum_{i \in N^*} (\tilde{v}_i - v_i)^2 + \frac{8}{n^*} \sum_{i \in N^*} v_i^2
\]

\[
= O_P \left( \frac{1}{n^*} \right) + \frac{8}{n^*} \sum_{i \in N^*} v_i^2,
\]

by (B.19). As for the last term,

\[
\frac{1}{n^*} \sum_{i \in N^*} \mathbb{E}[\tilde{v}_i^2 \mid \mathcal{F}] \leq \frac{2}{n^*} \sum_{i \in N^*} \mathbb{E}[R_i(\varepsilon)^2 \mid \mathcal{F}] + \frac{2}{n^*} \sum_{i \in N^*} \mathbb{E}[\eta_i^2 \mid \mathcal{F}].
\]

The last term is bounded by \( 2\sigma_{\eta_i}^2 \), and the first term on the right hand side is bounded by

\[
\frac{2\sigma_{\varepsilon}^2}{(1 - \beta_0)^2} + \frac{2}{n^*} \sum_{i \in N^*} \mathbb{E} \left[ \left( \frac{\beta_0}{np(i)} \lambda_{ij} \lambda_{ij} \varepsilon_j \right)^2 \mid \mathcal{F} \right] \leq C.
\]

Combining this with (B.19) and (B.20), we obtain the desired result.
(ii) Let us first write

\[
\frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N_P(i) \cap N^*} (\tilde{v}_j - v_i v_j) = \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N_P(i) \cap N^*} (\tilde{v}_i - v_i)(\tilde{v}_j - v_j) + \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N_P(i) \cap N^*} (\tilde{v}_i - v_i) v_j + \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N_P(i) \cap N^*} v_i (\tilde{v}_j - v_j) = A_{n,1} + A_{n,2} + A_{n,3}, \text{ say.}
\]

As for the leading term, by Cauchy-Schwarz inequality,

\[
|A_{n,1}| = \sqrt{\frac{1}{n^*} \sum_{i \in N^*} (\tilde{v}_i - v_i)^2 \left( \frac{1}{n^*} \sum_{i \in N^*} \left( \sum_{j \in N_P(i) \cap N^*} (\tilde{v}_j - v_j) \right)^2 \right)}.
\]

Note that

\[
\frac{1}{n^*} \sum_{i \in N^*} \mathbb{E} \left[ \left( \sum_{j \in N_P(i) \cap N^*} (\tilde{v}_j - v_j) \right)^2 \vert \mathcal{F} \right] \leq \frac{1}{n^*} \sum_{i \in N^*} |N_P(i) \cap N^*| \sum_{j \in N_P(i) \cap N^*} \mathbb{E} \left[ (\tilde{v}_j - v_j)^2 \vert \mathcal{F} \right] = \frac{1}{n^*} \sum_{i \in N^*} \left( \sum_{j \in N_P(i) \cap N^*} |N_P(j) \cap N^*| \right) \mathbb{E} \left[ (\tilde{v}_i - v_i)^2 \vert \mathcal{F} \right],
\]

where the inequality above uses Jensen’s inequality and the equality above uses Lemma B.2.1. Hence the last term is bounded by

\[
\max_{i \in N^*} \frac{|N_P(i) \cap N^*|}{n^*} \sum_{i \in N^*} \mathbb{E} \left[ (\tilde{v}_i - v_i)^2 \vert \mathcal{F} \right] \leq O_P \left( \frac{1}{n^*} \right) \text{ by (B.19).}
\]

Thus we conclude that

\[
|A_{n,1}| = O_P \left( \frac{1}{n^*} \right).
\]
Now, let us turn to $A_{n,2}$. Observe that

$$A_{n,2} = -\frac{1}{n^*} \sum_{i \in N^*} Z_i' \sum_{j \in N(i) \cap N^*} v_j (\tilde{\rho} - \rho_0)$$

$$= - \left( \frac{1}{n^*} \sum_{i \in N^*} Z_i' v_i \right) (\tilde{\rho} - \rho_0) = -S_{Z,v}(\tilde{\rho} - \rho_0)$$

using Lemma B.2.1. From the proof of (i), we obtain that

$$\tilde{\rho} - \rho_0 = O_P \left( \frac{1}{\sqrt{n^*}} \right).$$

Hence combined with Lemma B.2.3, we have

$$|A_{n,2}| = O_P \left( \frac{1}{n^*} \right).$$

Since by Lemma B.2.1, $A_{n,2} = A_{n,3}$, the proof of (ii) is complete.

(iii) Note that

$$\text{Var} \left( \frac{1}{n^*} \sum_{i \in N^*} R_i^2(\varepsilon)|\mathcal{F} \right) \leq \frac{2}{(n^*)^2} \sum_{i \in N^*} \text{Var} \left( \lambda_{ii}^2 \varepsilon_i^2 |\mathcal{F} \right)$$

$$+ \frac{2}{(n^*)^2} \sum_{i \in N^*} \text{Var} \left( \left( \frac{\beta_0 \lambda_{ii}}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij} \varepsilon_j \right)^2 |\mathcal{F} \right).$$

The leading term is $O_P((n^*)^{-1})$. The last term is bounded by

$$\frac{2}{(n^*)^2} \sum_{i \in N^*} \beta_0^2 \lambda_{ii}^2 \frac{1}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij} \varepsilon_j^2 |\mathcal{F} \right) = O_P((n^*)^{-1}).$$

Since $v_i = R_i(\varepsilon) + \eta_i$ and $\varepsilon_i$'s and $\eta_i$'s are independent, we obtain the desired rate.

(iv) For simplicity of notation, define

$$V_{ij} = (v_i v_j - E[v_i v_j |\mathcal{F}]) \tilde{\phi}_i \tilde{\phi}_j.$$
Then we write

\[ E \left[ \left( \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N^p(i) \cap N^*} V_{ij} \right)^2 \right] \bigg| \mathcal{F} \right]

\[ = \frac{1}{(n^*)^2} \sum_{i_1 \in N^*} \sum_{i \in N^p(i) \cap N^*} \sum_{j_1 \in N^*} \sum_{j \in N^p(i) \cap N^*} E [V_{i_1,j_1} V_{i,j} \big| \mathcal{F}] . \]

The last exception is zero, whenever \((i_2, j_2)\) is away from \((i_1, j_1)\) by more than two edges. Hence we can bound the last term by (using Assumption 3.3.4)

\[ \frac{C_1}{n^*} \max_{i \in N} E [v_i^2 \big| \mathcal{F}] \leq \frac{C_2}{n^*} \]

for some constants \(C_1, C_2\) which do not depend on \(n\).

**Lemma B.2.5.** Suppose that the conditions of Theorem 3.3.1 hold. Then,

\[ \hat{\Lambda} - \Lambda = O_P \left( \frac{1}{\sqrt{n^*}} \right) . \]

**Proof:** We write

\[ \hat{\Lambda}_1 - \Lambda_1 = \frac{1}{n^*} \sum_{i \in N^*} (\hat{v}_i^2 - E[v_i^2 \big| \mathcal{F}]) \hat{\phi}_i \hat{\phi}_i' \text{ and} \]

\[ \hat{\Lambda}_2 - \Lambda_2 = \frac{\hat{s}_e - s_e}{n^*} \sum_{i \in N^*} \sum_{j \in N^p(i) \cap N^*} q_{ij} \hat{\phi}_i \hat{\phi}_j.' \]

By Assumption 3.3.2 and Lemma B.2.4(ii)(iv), we have

\[ \hat{s}_e - s_e = O_P(1/\sqrt{n^*}). \]

The desired result follows by using this and applying Lemma B.2.4(i)(iii) to \(\hat{\Lambda}_1 - \Lambda_1\).

**Lemma B.2.6.** Suppose that the conditions of Theorem 3.3.1 hold. Then the following holds.

(i) \( \frac{1}{n^*} \sum_{i \in N^*} (\hat{v}_i^2 - v_i^2) \hat{\phi}_i \hat{\phi}_i' = O_P(1/\sqrt{n^*}). \)

(ii) \( \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N^p(i) \cap N^*} (\hat{v}_j - v_j) \hat{\phi}_j \hat{\phi}_j' = O_P(1/n^*). \)

**Proof:** First, write \( \hat{\phi} - \phi = -Z(\hat{\rho} - \rho_0) \), where

\[ \hat{\rho} - \rho_0 = [S_{Z\phi} \hat{\Lambda}^{-1} S_{Z\phi}']^{-1} S_{Z\phi} \hat{\Lambda}^{-1} S_{Z\phi} . \]
Following the same arguments as in the proof of Lemma B.2.4(i)(ii) and Lemma B.2.5, we obtain the desired result. ■

Proof of Theorem 3.3.1: Let us consider the first statement. We write
\[
\frac{1}{\sqrt{n^*}} \hat{\phi}' \hat{v} = \frac{1}{\sqrt{n^*}} \hat{\phi}' (\hat{v} - v) + \frac{1}{\sqrt{n^*}} \hat{\phi}' v
\]
\[
= -\frac{1}{\sqrt{n^*}} \hat{\phi}' Z (\hat{\rho} - \rho_0) + \frac{1}{\sqrt{n^*}} \hat{\phi}' v = \sqrt{n^*} (I - P) \hat{\Lambda}^{-1/2} S_{\phi v},
\]
using (B.21), where
\[
P = \hat{\Lambda}^{-1/2} S'_{Z\phi} \left[ S_{Z\phi} \hat{\Lambda}^{-1} S'_{Z\phi} \right]^{-1} S_{Z\phi} \hat{\Lambda}^{-1/2}.
\]
Note that \(P\) is a projection matrix from \(\mathbb{R}^M\) onto the range space of \(\hat{\Lambda}^{-1/2} S'_{Z\phi}\). Hence combining Lemmas B.2.2 and B.2.5, we obtain the desired result. The second result follows from Lemma B.2.2 and (B.21).

Let us turn to the third statement. First, note that
\[
\frac{1}{\sqrt{n^*}} \sum_{i \in N^*} \sum_{j \in \tilde{N}(i)} (\hat{v}_i \hat{v}_j - v_i v_j) = O_P(1/\sqrt{n^*}),
\]
by following precisely the same proof as that of Lemma B.2.4(ii). (Recall that \(\tilde{N}(i)\) is defined in Condition D in the main text.) Now, we let
\[
\sigma^2 = \text{Var} \left( \frac{1}{\sqrt{n^*}} \sum_{i \in N^*} \sum_{j \in \tilde{N}(i)} \eta_i \eta_j \right)
\]
and write
\[
\frac{1}{\sigma \sqrt{n^*}} \sum_{i \in N^*} \sum_{j \in \tilde{N}(i)} v_i v_j = \frac{1}{\sqrt{n^*}} \sum_{i \in N^*} r_i,
\]
where
\[
r_i = \frac{1}{\sigma} \sum_{j \in \tilde{N}(i)} \eta_i \eta_j,
\]
because \(v_i = \eta_i\) under the null hypothesis. Note that \(E[r_i|\mathcal{F}] = 0\). Let \(G_p\) be a graph on \(N^*\) such that \(i\) and \(j\) are adjacent if and only if \(j \in \tilde{N}(i)\) or \(i \in \tilde{N}(j)\). Then \(\{r_i\}_{i \in N^*}\) has \(G_p\) as a
dependency graph conditional on $\mathcal{F}$. Now we show the following:

$$(n^*)^{-1/4} \sqrt{\mu_3^2 + (n^*)^{-1/2} \mu_4^2} \to_p 0,$$

(B.22)

where for $p \geq 1$,

$$\mu_p = \max_{i \in N^*} (E[|r_i|^p | \mathcal{F}])^{1/p}.$$

Then by Theorem 2.3 of [145], we obtain that

$$\frac{1}{\sigma \sqrt{n^*}} \sum_{i \in N^*} \sum_{j \in N(i)} v_i v_j \to d N(0, 1),$$

as $n^* \to \infty$. First, note that

$$\sigma^2 = E \left( \left( \frac{1}{\sqrt{n^*}} \sum_{i \in N^*} \sum_{j \in N(i)} \eta_i \eta_j \right)^2 \right)$$

$$= \frac{1}{\sqrt{n^*}} \sum_{i \in N^*} \sum_{j \in N(i)} \sum_{i \in N} \sum_{j \in N(i) \cup N(j)} E[\eta_i \eta_j \eta_i \eta_j | \mathcal{F}].$$

Note that in the quadruple sum, $i_1 \neq j_1$ and $i_2 \neq j_2$. There are only two ways the last conditional expectation is not zero: either $i_1 = i_2$ and $j_1 = j_2$ or $j_1 = i_2$ and $i_1 = j_2$, because $\eta_i$’s are independent across $i$’s and its conditional expectation given $\mathcal{F}$ is zero. Hence the last term is equal to

$$\frac{2 \sigma_\eta^4}{n^*} \sum_{i \in N^*} |\bar{N}(i)| = 2 \sigma_\eta^4 \bar{d}_{av}$$

(B.23)

Hence for any $p \geq 2$,

$$\mu_p^p = \frac{1}{\sigma^p} \max_{i \in N^*} E \left[ \left( \sum_{j \in \bar{N}(i)} |\eta_i \eta_j| \right)^p \right] \leq \frac{\max_{i,j \in N^*} E[|\eta_i \eta_j|^p | \mathcal{F}]}{\sigma^p} \leq \frac{\max_{i,j \in N^*} E[|\eta_i \eta_j|^p | \mathcal{F}]}{2^p \sigma_\eta^2 \bar{d}_{av}^p}.$$

Note that $\bar{d}_{av} \geq 1$ because $\bar{N}(i) \neq \emptyset$ for all $i \in N^*$. Thus (B.22) follows. Now, by Lemma

203
and in the light of the expression (B.23), it is not hard to see that

$$2\hat{s}^4(\beta_0) = \sigma^2 + o_p(1).$$

The desired result follows from this and the Bonferroni procedure. ■
Appendix C

Appendix to Chapter 4

C.1 Proofs

Proof of Lemma 4.3.1:
Consider first \( k_t > k'_t \). Given the increasing cost of exerting influence, a whip exerts the minimum amount of influence necessary to ensure a vote for \( k_t \), provided this amount is less than or equal to \( y^\text{max}_p \). The minimum amount of influence is such that the member is indifferent, \( u(k_t, \omega_{t,i} + y_{i,t}) = u(k'_t, \omega_{t,i} + y_{i,t}) \) or \( |\omega_{t,i} + y_{i,t} - k_t| = |\omega_{t,i} + y_{i,t} - k'_t| \). This equality is satisfied if and only if \( \omega_{t,i} + y_{i,t} = MV_t = \frac{k_t + k'_t}{2} \). If \( \omega_{t,i} \geq MV_t \), the required influence is weakly negative (absent influence, the member votes for \( k_t \)) and so no influence is exerted. If \( \omega_{t,i} < MV_t \), a positive amount of influence, \( y_{i,t} = MV_t - \omega_{t,i} > 0 \) is required which increases linearly in \( MV_t - \omega_{t,i} \). Therefore, a member is whipped if and only if their ideology is such that \( MV_t - y^\text{max}_p \leq \omega_{t,i} < MV_t \). For \( k_t < k'_t \), the argument is reversed: only members for which \( MV_t < \omega_{t,i} \leq MV_t + y^\text{max}_p \) are whipped. □

Proof of Lemma 4.3.2:
Consider the mass, \( f(\theta) \), of members at some \( \theta \), each of whom has an independent signal of \( \hat{\eta}_t^1 \) due to their independent ideological shocks. The average number of ‘yes’ reports from \( N \) at \( \theta \) members is given by \( \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} I \left( u(x_t, \theta + \delta_t^1 + \hat{\eta}_t^1) \geq u(q_t, \theta + \delta_t^1 + \hat{\eta}_t^1) \right) \) where \( I() \) represents the indicator function. By the law of large numbers, as \( N \to \infty \), this average converges to:
\[ f(\theta)E[I(u(x_t, \theta + \delta_1^1 + \hat{\eta}_t) \geq u(q_t, \theta + \delta_1^1 + \hat{\eta}_t))] = f(\theta) Pr(u(x_t, \theta + \delta_1^1 + \hat{\eta}_t) \geq u(q_t, \theta + \delta_1^1 + \hat{\eta}_t)) \]
\[ = f(\theta) Pr(\theta + \delta_1^1 + \hat{\eta}_t \geq MV_t) \]
\[ = f(\theta) (1 - G(MV_t - \theta - \hat{\eta}_t^1)). \]

Therefore, after observing the number of ‘yes’ reports for a given \( \theta \), \( \hat{\eta}_t^1 \) is known with probability one. \( \Box \)

**Proof of Lemma 4.3.3:**

Consider \( x_t > q_t \). Let \( G_{1+2}(\cdot) \) denote the cdf of \( \delta_1^1 + \delta_2^2 \) (with corresponding pdf, \( g_{1+2}(\cdot) \)). For a given \( MV_t \), the number of votes for \( x_t \) from a given party’s members is known with probability one due to independent idiosyncratic shocks and a continuum of members. To see this fact, consider the continuum of party \( p \)’s members located at each \( \theta \), each with independent shocks, \( \delta_1^1 \) and \( \delta_2^2 \). With \( N \) voters at \( \theta \), the average number of votes from these members is given by \( \lim_{N \to \infty} \frac{f(\theta)}{N} \sum_{i=1}^{N} I(\theta + \eta_1^1 + \eta_2^2 + \delta_1^1 + \delta_2^2 \geq MV_t \pm \gamma_{\max}^p) \), where the sign with which \( \gamma_{\max}^p \) enters depends upon the direction that party \( p \) whips. By the law of large numbers, as \( N \to \infty \), this average converges to:

\[ f(\theta)E[I(\theta + \eta_1^1 + \eta_2^2 + \delta_1^1 + \delta_2^2 \geq MV_t \pm \gamma_{\max}^p)] = f(\theta) Pr(\theta + \eta_1^1 + \eta_2^2 + \delta_1^1 + \delta_2^2 \geq MV_t \pm \gamma_{\max}^p) \]
\[ = f(\theta) (1 - G_{1+2}(MV_t - \eta_1^1 - \eta_2^2 \pm \gamma_{\max}^p)). \]

Denote the realized marginal voter after the aggregate shocks as \( \tilde{MV}_t = MV_t - \eta_1^1 - \eta_2^2 \). Then, the number of votes for \( x_t \) from party \( D \)’s members is given by \( Y_D(\tilde{MV}_t) = N_D \int_{-\infty}^{\infty} \left( 1 - G_{1+2}(\tilde{MV}_t - \theta \pm \gamma_{\max}^D) \right) f_D(\theta) d\theta \). The corresponding expression for party \( R \) is \( Y_R(\tilde{MV}_t) = N_R \int_{-\infty}^{\infty} \left( 1 - G_{1+2}(\tilde{MV}_t - \theta \pm \gamma_{\max}^R) \right) f_R(\theta) d\theta \). The total number of votes for \( x_t \) is then given by \( Y(\tilde{MV}_t) \equiv Y_D(\tilde{MV}_t) + Y_R(\tilde{MV}_t) \).

\( Y(\tilde{MV}_t) \) is strictly decreasing in \( x_t \). To see this, consider the votes from party \( D \)’s members, \( Y_D(x_t) \):

\[ \frac{\partial Y_D(MV_t)}{\partial x_t} = \frac{1}{2} \frac{\partial}{\partial MV_t} N_D \left[ \int_{-\infty}^{\infty} \left( 1 - G_{1+2}(\tilde{MV}_t - \theta \pm \gamma_{\max}^D) \right) f_D(\theta) d\theta \right] \]
\[ = -\frac{N_D}{2} \int_{-\infty}^{\infty} g_{1+2}(\tilde{MV}_t - \theta \pm \gamma_{\max}^D)f_D(\theta) d\theta \quad (C.1) \]
(C.1) is strictly less than zero given that that ideological shocks are unbounded, independent of the (finite) amount or direction of whipping. The same is true of the derivative of \( Y_R(M\hat{V}_t) \), ensuring \( Y(M\hat{V}_t) \) strictly decreases in \( x_t \) for \( x_t > q_t \). For \( x_t < q_t \), we have \( Y_D(M\hat{V}_t) = N_D \left[ \int_{-\infty}^{\infty} G_{1+2}(MV_t - \theta \pm y_D^{\max}) f_D(\theta) d\theta \right] \) and \( Y_R(M\hat{V}_t) = N_R \left[ \int_{-\infty}^{\infty} G_{1+2}(MV_t - \theta \pm y_R^{\max}) f_R(\theta) d\theta \right] \) so that \( Y(M\hat{V}_t) \) increases in \( x_t \). Since for \( q_t < \theta_{m,p} \) we must have \( x_t > q_t \) and for \( q_t > \theta_{m,p} \) we must have \( x_t < q_t \), we see that the number of votes for \( x_t \) strictly decreases the closer it gets to the proposing party’s ideal point. □

Proof of Proposition 4.3.4:

For \( q_t = \theta_{m,D} \), clearly \( x_t^{\text{no count}} = x_t^{\text{no count}} = \theta_{m,D} \) are the unique optimal alternative policies because party \( D \) can do no better than its ideal point.

In the case of no whip count, and \( q_t < \theta_{m,D} \) so that \( x_t > q_t \), we can rewrite party \( D \)'s expected utility as

\[
EU_{D}^{\text{no count}}(q_t, x_t) = \left( 1 - \Phi \left( \frac{MV_t - \hat{MV}_{R,R}}{\sigma} \right) \right) (u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D})) + u(q_t, \theta_{m,D}) - C_b
\]

The derivative with respect to \( x_t \) is given by

\[
\left( 1 - \Phi \left( \frac{MV_t - \hat{MV}_{R,R}}{\sigma} \right) \right) u_x(x_t, \theta_{m,D}) - \frac{1}{2\sigma} \phi \left( \frac{MV_t - \hat{MV}_{R,R}}{\sigma} \right) (u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D}))
\]

where \( \phi() \) denotes the pdf of the standard normal distribution. At \( x_t = q_t \), the derivative is strictly positive given \( q_t < \theta_{m,D} \) and the fact that \( \hat{MV}_{R,R} \) is finite. At \( x_t = \theta_{m,D} \), it is strictly negative given \( u(q_t, \theta_{m,D}) < 0 \). Together these facts ensure an interior solution, which we now show is unique. Any interior solution must satisfy the first-order condition,

\[
\left( 1 - \Phi \left( \frac{MV_t^{\text{no count}} - \hat{MV}_{R,R}}{\sigma} \right) \right) u_x(x_t^{\text{no count}}, \theta_{m,D}) - \frac{1}{2\sigma} \phi \left( \frac{MV_t^{\text{no count}} - \hat{MV}_{R,R}}{\sigma} \right) (u(x_t^{\text{no count}}, \theta_{m,D}) - u(q_t, \theta_{m,D})) = 0 \tag{C.2}
\]

Defining \( x_t^{\text{no count}} = \frac{MV_t^{\text{no count}} - \hat{MV}_{R,R}}{\sigma} \), we can re-write the first-order condition as:

\[
\frac{1 - \Phi(x_t^{\text{no count}})}{\phi(x_t^{\text{no count}})} = \frac{1}{2\sigma} \frac{u(x_t^{\text{no count}}, \theta_{m,D}) - u(q_t, \theta_{m,D})}{u_x(x_t^{\text{no count}}, \theta_{m,D})} \tag{C.3}
\]

The left-hand side of (C.3) is the inverse hazard rate of a standard normal distribution.
and so is strictly decreasing in \( x^\text{no count}_t \) (and therefore \( x^\text{no count}_t \) since \( x^\text{no count}_t \) strictly increases in \( x^\text{no count}_t \)). The sign of the derivative of the right-hand side with respect to \( x^\text{no count}_t \) is given by

\[
\frac{\partial}{\partial x^\text{no count}_t} \left( u_{xx}(x^\text{no count}_t, \theta_m, \text{D}) \right) - \frac{\partial}{\partial x^\text{no count}_t} \left( u_{x}(x^\text{no count}_t, \theta_m, \text{D}) - u(q_t, \theta_m) \right)
\]

which is strictly positive because \( u_{xx}(x^\text{no count}_t, \theta_m, \text{D}) < 0 \) and \( u(x^\text{no count}_t, \theta_m, \text{D}) > u(q_t, \theta_m) \). Thus, the right-hand side is strictly increasing in \( x^\text{no count}_t \). Together, these facts guarantee a unique solution, \( x^\text{no count}_t \in (q_t, \theta_m, \text{D}) \).

In the case of a whip count and \( q_t < \theta_m, \text{D} \), we can rewrite the party’s expected utility:

\[
EU^\text{count}_D(q_t, x_t) = \Pr(\eta^1_t \geq n^1_t) \left( \Pr(x_t \text{ wins} | \eta^1_t \geq n^1_t) (u(x_t, \theta_m, \text{D}) - u(q_t, \theta_m, \text{D})) + u(q_t, \theta_m, \text{D}) - C_b \right) + \Pr(\eta^1_t < n^1_t) u(q_t, \theta_m, \text{D})
\]

\[
= \Pr(\eta^1_t \geq n^1_t, x_t \text{ wins}) (u(x_t, \theta_m, \text{D}) - u(q_t, \theta_m, \text{D})) - \Pr(\eta^1_t \geq n^1_t) C_b + u(q_t, \theta_m, \text{D})
\]

\[
= \int_{n^1_t}^{\infty} \left( 1 - \Phi \left( \frac{MV_t - \hat{MV}_{R,R} - \eta}{\sigma_2} \right) \right) \frac{1}{\sigma_1} \phi \left( \frac{\eta}{\sigma_1} \right) d\eta \left( u(x_t, \theta_m, \text{D}) - u(q_t, \theta_m, \text{D}) \right) - \left( 1 - \Phi \left( \frac{n^1_t}{\sigma_1} \right) \right) C_b + u(q_t, \theta_m, \text{D})
\]

Taking the derivative with respect to \( x_t \) yields

\[1\]

The second-order condition at \( x^\text{no count}_t \) is also easily checked, but must be satisfied given that marginal expected utility is increasing at \( x_t = q_t \), decreasing at \( x_t = \theta_m, \text{D} \) and the solution is unique.

\[2\]

The necessary conditions for applying the Leibniz Integral Rule with an infinite bound are satisfied. Specifically, the integrand and its partial derivative with respect to \( x_t \) are both continuous functions of \( x_t \) and \( \eta_t \), and it is possible to find integrable functions of \( \eta_t \) that bound the integrand and its partial derivative with respect to \( x_t \)
\begin{align*}
&\frac{d EU_{D}^{\text{count}}(q_t, x_t)}{dx_t} = \frac{-d \eta_0^1}{dx_t} \frac{1}{\sigma_1} \phi \left( \frac{\eta_0^1}{\sigma_1} \right) \left[ 1 - \Phi \left( \frac{MV_t - \hat{MV}_{R,R} - \eta_1}{\sigma_2} \right) \right] (u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D})) \\
&\quad - \frac{1}{2 \sigma_1 \sigma_2} \int_{\eta_0^1}^{\infty} \phi \left( \frac{MV_t - \hat{MV}_{R,R} - \eta}{\sigma_2} \right) \phi \left( \frac{\eta}{\sigma_1} \right) d\eta (u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D})) \\
&\quad + \frac{1}{\sigma_1} \xi_t(x_t, \theta_{m,D}) \int_{\eta_0^1}^{\infty} \left[ 1 - \Phi \left( \frac{MV_t - \hat{MV}_{R,R} - \eta}{\sigma_2} \right) \right] \phi \left( \frac{\eta}{\sigma_1} \right) d\eta \\
&\quad + \frac{1}{\sigma_1} \frac{d \eta_0^1}{dx_t} \phi \left( \frac{\eta_0^1}{\sigma_1} \right) C_b \\
&= \frac{1}{\sigma_1} \xi_t(x_t, \theta_{m,D}) \int_{\eta_0^1}^{\infty} \left[ 1 - \Phi \left( \frac{MV_t - \hat{MV}_{R,R} - \eta}{\sigma_2} \right) \right] \phi \left( \frac{\eta}{\sigma_1} \right) d\eta \\
&\quad - \frac{1}{2 \sigma_1 \sigma_2} \int_{\eta_0^1}^{\infty} \phi \left( \frac{MV_t - \hat{MV}_{R,R} - \eta}{\sigma_2} \right) \phi \left( \frac{\eta}{\sigma_1} \right) d\eta (u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D})) \\
&\quad \text{(C.4)}
\end{align*}

where the second equality uses the fact that $\eta_0^1$ satisfies

\begin{align*}
\left( 1 - \Phi \left( \frac{MV_t - \hat{MV}_{R,R} - \eta_1}{\sigma_2} \right) \right) (u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D})) = C_b \\
&\quad \text{(C.5)}
\end{align*}

Consider the limit as $C_b \to 0$. From (C.5), we can see that, provided $x_t$ is bounded away from $q_t$ so that $u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D}) > 0$ (which we subsequently confirm), we must have $\eta_0^1 \to -\infty$ as $C_b \to 0$. But, as $\eta_0^1 \to -\infty$, the party always continues to pursue the bill after the first aggregate shock. In this case, the optimal alternative policy is identical to the case of no whip count. Formally,

\begin{align*}
\lim_{\eta_0^1 \to -\infty} \frac{d EU_{D}^{\text{count}}(q_t, x_t)}{dx_t} &= \frac{1}{\sigma_1} \xi_t(x_t, \theta_{m,D}) \int_{-\infty}^{\infty} \left[ 1 - \Phi \left( \frac{MV_t - \hat{MV}_{R,R} - \eta}{\sigma_2} \right) \right] \phi \left( \frac{\eta}{\sigma_1} \right) d\eta \\
&\quad - \frac{1}{2 \sigma_1 \sigma_2} \int_{-\infty}^{\infty} \phi \left( \frac{MV_t - \hat{MV}_{R,R} - \eta}{\sigma_2} \right) \phi \left( \frac{\eta}{\sigma_1} \right) d\eta (u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D})) \\
&\quad = \xi_t(x_t, \theta_{m,D}) \left( 1 - \Phi \left( \frac{MV_t - \hat{MV}_{R,R}}{\sigma} \right) \right) \\
&\quad - \frac{1}{2 \sigma} \phi \left( \frac{MV_t - \hat{MV}_{R,R}}{\sigma} \right) (u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D})) \\
&\quad \text{(C.6)}
\end{align*}

where the equality follows from the fact that the convolution of two standard normal distributions is a normal distribution with the sum of the variances and using $\sigma^2 = \sigma^2_1 + \sigma^2_2$. 
Comparing (C.6) with (C.2), we can see immediately that, in the limit, the first-order condition for the whip and no whip cases are identical, and it therefore follows that \( x_t^{\text{count}} \) is unique and interior as in the no whip case. This fact ensures that \( u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D}) > 0 \) in the limit, confirming that we must have \( \eta_1 \rightarrow -\infty \) as \( C_b \rightarrow 0 \).

We now show that \( x_t^{\text{count}} \) is unique and interior for strictly positive \( C_b \). From (C.4), we see that \( \frac{dU_t^{\text{count}}(q_t, \eta_1)}{dx_t} \) is strictly positive at \( x_t = q_t \) and strictly negative at \( x_t = \theta_{m,D} \), ensuring an interior optimum, \( x_t^{\text{count}} \) which must satisfy the first-order condition\(^3\)

\[
\frac{\int_{\eta_1}^{\infty} \left( 1 - \Phi\left( \frac{MV_t^{\text{count}} -MV_{RR,R} - \eta_1}{\sigma_2} \right) \right) \phi\left( \frac{\eta_1}{\sigma_1} \right) d\eta}{\int_{\eta_1}^{\infty} \frac{1}{\sigma_2} \phi\left( \frac{MV_t^{\text{count}} -MV_{RR,R} - \eta_1}{\sigma_2} \right) \phi\left( \frac{\eta_1}{\sigma_1} \right) d\eta} = \frac{(u(x_t^{\text{count}}, \theta_{m,D}) - u(q_t, \theta_{m,D}))}{u_t(x_t^{\text{count}}, \theta_{m,D})} \tag{C.7}
\]

As in the case of no whip count, the right-hand side of (C.7) strictly increases in \( x_t^{\text{count}} \). It remains to show that, in the limit as \( C_b \rightarrow 0 \), the left-hand side of (C.7) strictly decreases in \( x_t^{\text{count}} \), which, by continuity of the left-hand side in \( C_b \), ensures there exists a strictly positive value of \( C_b, \hat{C}_b > 0 \), such that for all \( C_b < \hat{C}_b \), the left-hand side continues to strictly decrease. It then follows that \( x_t^{\text{count}} \) is unique for all \( C_b < \hat{C}_b \). The sign of the derivative of the left-hand side of (C.7) with respect to \( x_t^{\text{count}} \), is determined by\(^4\)

\[
- \frac{d \eta_1}{dx_t^{\text{count}}} \phi\left( \frac{\eta_1}{\sigma_1} \right) \left( 1 - \Phi\left( \frac{MV_t^{\text{count}} - \hat{M}V_{RR,R} - \eta_1}{\sigma_2} \right) \right) \frac{1}{2 \sigma_2} \int_{\eta_1}^{\infty} \phi\left( \frac{MV_t^{\text{count}} - \hat{M}V_{RR,R} - \eta_1}{\sigma_2} \right) \phi\left( \frac{\eta_1}{\sigma_1} \right) d\eta \\
+ \frac{d \eta_1}{dx_t^{\text{count}}} \frac{1}{2 \sigma_2} \phi\left( \frac{MV_t^{\text{count}} - \hat{M}V_{RR,R} - \eta_1}{\sigma_2} \right) \phi\left( \frac{\eta_1}{\sigma_1} \right) \int_{\eta_1}^{\infty} \phi\left( \frac{\eta_1}{\sigma_1} \right) d\eta \\
- \left( \frac{1}{2 \sigma_2} \int_{\eta_1}^{\infty} \phi\left( \frac{MV_t^{\text{count}} - \hat{M}V_{RR,R} - \eta_1}{\sigma_2} \right) \phi\left( \frac{\eta_1}{\sigma_1} \right) d\eta \right)^2 \\
- \frac{1}{4 \sigma_2} \int_{\eta_1}^{\infty} \phi\left( \frac{MV_t^{\text{count}} - \hat{M}V_{RR,R} - \eta_1}{\sigma_2} \right) \phi\left( \frac{\eta_1}{\sigma_1} \right) d\eta \int_{\eta_1}^{\infty} \phi\left( \frac{\eta_1}{\sigma_1} \right) d\eta \\
- \frac{1}{2 \sigma_2} \int_{\eta_1}^{\infty} \phi\left( \frac{MV_t^{\text{count}} - \hat{M}V_{RR,R} - \eta_1}{\sigma_2} \right) \phi\left( \frac{\eta_1}{\sigma_1} \right) d\eta \int_{\eta_1}^{\infty} \left( 1 - \Phi\left( \frac{MV_t^{\text{count}} - \hat{M}V_{RR,R} - \eta_1}{\sigma_2} \right) \right) \phi\left( \frac{\eta_1}{\sigma_1} \right) d\eta \tag{C.8}
\]

By the implicit function theorem, \( \frac{d \eta_1}{dx_t} \) must satisfy (from (C.5))

\(^3\)These statements require \( \eta_1 \rightarrow -\infty \), which, by continuity, is true for \( C_b \) sufficiently small given that \( \eta_1 \rightarrow -\infty \) as \( C_b \rightarrow 0 \).

\(^4\)Again, the necessary conditions for applying the Leibniz Integral Rule with an infinite bound are satisfied.
\[-\phi \left( \frac{MV_i^{count} - \hat{MV}_{R,R} - \eta_1}{\sigma_2} \right) \frac{1}{\sigma_2} \left( \frac{1}{2} - \frac{d\eta_1^i}{d\chi_i^{count}} \right) (u(x_i^{count}, \theta_m, D) - u(q_t, \theta_m)) \]
\[+ \left( 1 - \Phi \left( \frac{MV_i^{count} - \hat{MV}_{R,R} - \eta_1}{\sigma_2} \right) \right) u_i(x_i^{count}, \theta_m, D) = 0 \]

or

\[\frac{d\eta_1^i}{d\chi_i^{count}} = \frac{1}{2} - \sigma_2 \left( 1 - \Phi \left( \frac{MV_i^{count} - \hat{MV}_{R,R} - \eta_1}{\sigma_2} \right) \right) u_i(x_i^{count}, \theta_m, D) \]
\[\phi \left( \frac{MV_i^{count} - \hat{MV}_{R,R} - \eta_1}{\sigma_2} \right) (u(x_i^{count}, \theta_m, D) - u(q_t, \theta_m)) \]  \hspace{1cm} (C.9)

In the limit as \(C_b \to 0\), \(\eta_1 \to -\infty\), in which case the second term of (C.9) approaches zero because \(x_i^{count}\) is bounded away from \(q_t\) and \(\theta_m, D\), and the inverse hazard rate of a standard normal random variable approaches zero as its argument approaches infinity. The limit of (C.8) as \(C_b \to 0\) is then determined by the limit of its second two terms because the first two terms approach zero. Defining \(\varepsilon_i^{count} \equiv \frac{MV_i^{count} - \hat{MV}_{R,R}}{\sigma}\), this limit is given by

\[5 \lim_{x \to -\infty} \frac{1 - \phi(x)}{\phi(x)} = \lim_{x \to -\infty} \frac{-\phi(x)}{\phi(x)} = \lim_{x \to -\infty} \frac{-\phi(x)}{\phi'(x)} = 0 \] where the first equality uses L'Hôpital's rule.
\[
\lim_{n \to -\infty} - \left( \frac{1}{2\sigma^2} \int_{n}^{\infty} \phi \left( \frac{MV_{l, count} - MV_{R, R} - \eta}{\sigma} \right) \phi \left( \frac{\eta}{\sigma_1} \right) d\eta \right)^2 \\
- \frac{1}{4\sigma^2} \int_{-\infty}^{\infty} \phi' \left( \frac{MV_{l, count} - MV_{R, R} - \eta}{\sigma} \right) \phi \left( \frac{\eta}{\sigma_1} \right) d\eta \int_{n}^{\infty} \phi \left( \frac{MV_{l, count} - MV_{R, R} - \eta}{\sigma} \right) \phi \left( \frac{\eta}{\sigma_1} \right) d\eta \\
= - \left( \frac{1}{2\sigma} \Phi \left( \frac{MV_{l, count} - MV_{R, R}}{\sigma} \right) \right)^2 - \frac{1}{4\sigma^2} \phi' \left( \frac{MV_{l, count} - MV_{R, R}}{\sigma} \right) \left( 1 - \Phi \left( \frac{MV_{l, count} - MV_{R, R}}{\sigma} \right) \right) \\
= - \left( \frac{1}{2\sigma} \phi \left( z_{l, count} \right) \right)^2 - \frac{1}{4\sigma^2} \phi' \left( z_{l, count} \right) \left( 1 - \Phi \left( z_{l, count} \right) \right) \\
= - \left( \frac{1}{2\sigma} \phi \left( z_{l, count} \right) \right)^2 + \frac{1}{4\sigma^2} \phi \left( z_{l, count} \right) \left( 1 - \Phi \left( z_{l, count} \right) \right) \\
< - \left( \frac{1}{2\sigma} \phi \left( z_{l, count} \right) \right)^2 + \frac{1}{4\sigma^2} \phi \left( z_{l, count} \right) \left( 1 - \Phi \left( z_{l, count} \right) \right) \left( 1 - \Phi \left( z_{l, count} \right) \right) \\
= 0
\]

where the second equality uses properties of the convolution of normal distributions,
and the inequality follows from the fact that, for a standard normal random variable, \( x \left( 1 - \Phi(x) \right) < \phi(x) \).

For \( q_t > \theta_{m,D} \) so that \( x_t < q_t \), we assume party \( R \) whips against the bill (supports \( q_t \)). In case of no whip count, we can write party \( D \)'s expected utility as

\[
EU_{D \ no \ count}^{q_t, x_t} = \Phi \left( \frac{MV_{l} - \tilde{MV}_{L,R}}{\sigma} \right) \left( u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D}) \right) + u(q_t, \theta_{m,D}) - C_b
\]

With a whip count, it is
Using these expressions, the optimal policy candidates, \( x_i^{\text{count}} \) and \( x_i^{\text{no count}} \), can be shown to be unique (provided \( C_b \) is not too large) as in the previous case. □

To prove Lemma 4.3.5, we first define and prove Lemma C.1.1.

**Lemma C.1.1.** Fix \( C_b < \hat{C}_b \) such that the optimal alternative policies, \( x_i^{\text{count}} \) and \( x_i^{\text{no count}} \), are unique. Then, the alternative policies that satisfy the first-order conditions with and without a whip count (C.7) and (C.3) are such that:

1. For \( q_t \neq \theta_{m,D} \), the optimal alternative policy with a whip count, \( x_i^{\text{count}} \), lies strictly closer to party D’s ideal point, \( \theta_{m,D} \), than that without, \( x_i^{\text{no count}} \).

2. \( MV_i^{\text{count}}(q_t) \) and \( MV_i^{\text{no count}}(q_t) \) strictly increase for \( q_t < \theta_{m,D} \) and strictly increase for \( q_t > \theta_{m,D} \).

**Proof of Lemma C.1.1:**

Part 1. Consider the case of \( q_t < \theta_{m,D} \). We can write the first-order condition in the case of no whip count as an integration over the second aggregate shock (as in the case of the whip count):

\[
EU_{D}^{\text{count}}(q_t, x_t) = \int_{-\infty}^{\eta_1} \Phi \left( \frac{MV_i^{\text{count}} - \hat{MV}_{R,R} - \eta}{\sigma_2} \right) \frac{1}{\sigma_1} \phi \left( \frac{\eta}{\sigma_1} \right) d\eta \left( u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D}) \right) - \Phi \left( \frac{\eta_1}{\sigma_1} \right) C_b + u(q_t, \theta_{m,D})
\]

Consider the left-hand side of this expression, evaluated instead at \( x_i^{\text{count}} \):
for the case of a whip count. Consider the sign of the integrand in (C.10): 

\[
\int_{-\infty}^{\infty} \left[ 1 - \Phi \left( \frac{MV_t^{count} - MV_{\bar{t}}^{count}}{\sigma_2} \right) \right] \phi \left( \frac{\eta}{\sigma_1} \right) d\eta
\]

Thus, the integrand in (C.10) must be strictly negative over \( \eta \), evaluated at the optimal alternative policy for the case of a whip count. To satisfy the first-order condition for the case of a whip count at \( x_t \), the marginal expected utility is positive. The case of no whip count is satisfied (given that \( x_t \) evaluated at \( x_t^{\text{no count}} \) must be strictly negative so that the single zero-crossing is contained in \( \eta \)) so that the integrand approaches 1. As \( \eta \rightarrow \infty \), the integrand evaluated at \( \eta \) must be strictly negative so that the single zero-crossing is contained in \( \eta \) (otherwise the integrand is positive over the whole range and cannot integrate to zero). Thus, the integrand in (C.10) must be strictly negative over \( [-\infty, \eta] \) so that the integral is strictly negative: the marginal expected utility for the case of no whip count must be negative when evaluated at the optimal alternative policy for the case of a whip count. But, then we must have \( x_t^{\text{no count}} < x_t^{\text{count}} \) to ensure that the first-order condition for the case of no whip count is satisfied (given that \( x_t^{\text{no count}} \) is the unique optimum, for every \( x_t < x_t^{\text{no count}} \), the marginal expected utility is positive). The case of \( q_t > \theta_{m,D} \) can be shown similarly.

Part 2. Consider the case of \( q_t < \theta_{m,D} \) when a whip count is conducted. \( MV_t^{count} \) is determined implicitly by the first-order condition, (C.7). Taking its derivative with respect
to \( q_t \), we have

\[
\frac{\partial}{\partial q_t} \left[ \int_{\eta_t}^{\infty} \left( 1 - \Phi \left( \frac{MV_{\text{count}} - MV_{RR, \eta}}{\sigma_2} \right) \phi \left( \frac{\eta}{\sigma_1} \right) d\eta \right) \frac{\phi \left( \frac{\eta}{\sigma_1} \right) d\eta}{\mu_{x_t} \left( x_t, \theta_{m,D} \right)} - \frac{(u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D}))}{u_v \left( x_t, \theta_{m,D} \right)} \right] = 0
\]

\[\begin{align*}
\frac{\partial}{\partial MV_{i, \text{count}}} 
& \left[ \int_{\eta_t}^{\infty} \left( 1 - \Phi \left( \frac{MV_{\text{count}} - MV_{RR, \eta}}{\sigma_2} \right) \phi \left( \frac{\eta}{\sigma_1} \right) d\eta \right) \frac{\phi \left( \frac{\eta}{\sigma_1} \right) d\eta}{\mu_{x_t} \left( x_t, \theta_{m,D} \right)} - \frac{(u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D}))}{u_v \left( x_t, \theta_{m,D} \right)} \right] \frac{\partial MV_{i, \text{count}}}{\partial q_t} = 0 \\
\frac{\partial}{\partial x_t} 
& \left[ \int_{\eta_t}^{\infty} \left( 1 - \Phi \left( \frac{MV_{\text{count}} - MV_{RR, \eta}}{\sigma_2} \right) \phi \left( \frac{\eta}{\sigma_1} \right) d\eta \right) \frac{\phi \left( \frac{\eta}{\sigma_1} \right) d\eta}{\mu_{x_t} \left( x_t, \theta_{m,D} \right)} - \frac{(u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D}))}{u_v \left( x_t, \theta_{m,D} \right)} \right] \frac{\partial MV_{i, \text{count}}}{\partial q_t} = 0
\end{align*}\]

As shown in the proof of Proposition 4.3.4, the term in brackets on the left-hand side is strictly negative for \( C_b < \tilde{C}_b \). But, the term on the right-hand side is also strictly negative so that \( \frac{\partial MV_{i, \text{count}}}{\partial q_t} > 0 \). Similarly, \( \frac{\partial MV_{i, \text{pro} \text{count}}}{\partial q_t} > 0 \). For \( q_t > \theta_{m,D} \), we can similarly establish \( \frac{\partial MV_{i, \text{count}}}{\partial q_t} < 0 \) and \( \frac{\partial MV_{i, \text{pro} \text{count}}}{\partial q_t} < 0 \). □

Proof of Lemma 4.3.5:

\( V_{D, \text{count}}(q_t) > V_{D, \text{pro} \text{count}}(q_t) \) because, for \( C_b \) sufficiently small, \( \eta^1_t < \infty \) and \( \bar{\eta}^1_t > \infty \) (see footnote 3) so that an alternative policy is pursued for a non-zero measure of the support of \( \eta^1_t \). Therefore, for the same alternative policy, party \( D \)’s expected utility with a whip count must strictly exceed that without because over this support of \( \eta^1_t \), the cost, \( C_b \), is avoided and the probability of the alternative passing is the same. If party \( D \) pursues a different alternative policy with a whip count (which it generally does), then it must because it does
even better.

Consider the case of \( q_t < \theta_{m,D} \). We claim both value functions decrease with \( q_t \), but the difference \( V_D^{\text{no count}}(q_t) - V_D^{\text{count}}(q_t) \) increases. By the envelope theorem, the derivative of the value function for the case of no whip count with respect to \( q_t \) is given by

\[
\frac{\partial V_D^{\text{no count}}(q_t)}{\partial q_t} = - \left( 1 - \Phi\left( \frac{MV_{t}^{\text{no count}} - \bar{M}V_{R,R}}{\sigma} \right) \right) u_q(q_t, \theta_{m,D})
- \frac{1}{2\sigma} \phi\left( \frac{MV_{t}^{\text{no count}} - \bar{M}V_{R,R}}{\sigma} \right) \left( u(x_{t}^{\text{no count}}, \theta_{m,D}) - u(q_t, \theta_{m,D}) \right)
- \left( 1 - \Phi\left( \frac{MV_{t}^{\text{no count}} - \bar{M}V_{R,R}}{\sigma} \right) \right) u_x(x_{t}^{\text{no count}}, \theta_{m,D})
= - \left( 1 - \Phi\left( \frac{MV_{t}^{\text{no count}} - \bar{M}V_{R,R}}{\sigma} \right) \right) (u_q(q_t, \theta_{m,D}) + u_x(x_{t}^{\text{no count}}, \theta_{m,D}))
\]

where the first equality follows from applying the first-order condition. With unbounded aggregate shocks and \( q_t, x_t^{\text{no count}} < \theta_{m,D} \), this derivative is strictly negative so that the value of pursuing an alternate policy strictly decreases with \( q_t \).

In a similar manner, for the case of a whip count, we have

\[
\frac{\partial V_D^{\text{count}}(q_t)}{\partial q_t} = - \frac{1}{2\sigma_1 \sigma_2} \int_{\eta_1}^{\infty} \phi\left( \frac{MV_{t}^{\text{count}} - \bar{M}V_{R,R} - \eta}{\sigma_1} \right) \phi\left( \frac{\eta}{\sigma_1} \right) d\eta \left( u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D}) \right)
- \frac{1}{\sigma_1} u_q(q_t, \theta_{m,D}) \int_{\eta_1}^{\infty} \left( 1 - \Phi\left( \frac{MV_{t}^{\text{count}} - \bar{M}V_{R,R} - \eta}{\sigma_2} \right) \right) \phi\left( \frac{\eta}{\sigma_1} \right) d\eta
= - \frac{1}{\sigma_1} (u_q(q_t, \theta_{m,D}) + u_x(x_t^{\text{count}}, \theta_{m,D})) \int_{\eta_1}^{\infty} \left( 1 - \Phi\left( \frac{MV_{t}^{\text{count}} - \bar{M}V_{R,R} - \eta}{\sigma_2} \right) \right) \phi\left( \frac{\eta}{\sigma_1} \right) d\eta
\]

which is also strictly negative, given \( \eta_1 < \infty \).

Finally, consider the marginal difference in the value functions:

\[
\frac{\partial V_D^{\text{count}}(q_t)}{\partial q_t} = - \frac{1}{2\sigma_1 \sigma_2} \int_{\eta_1}^{\infty} \phi\left( \frac{MV_{t}^{\text{count}} - \bar{M}V_{R,R} - \eta}{\sigma_1} \right) \phi\left( \frac{\eta}{\sigma_1} \right) d\eta \left( u(x_t, \theta_{m,D}) - u(q_t, \theta_{m,D}) \right)
- \frac{1}{\sigma_1} u_q(q_t, \theta_{m,D}) \int_{\eta_1}^{\infty} \left( 1 - \Phi\left( \frac{MV_{t}^{\text{count}} - \bar{M}V_{R,R} - \eta}{\sigma_2} \right) \right) \phi\left( \frac{\eta}{\sigma_1} \right) d\eta
= - \frac{1}{\sigma_1} (u_q(q_t, \theta_{m,D}) + u_x(x_t^{\text{count}}, \theta_{m,D})) \int_{\eta_1}^{\infty} \left( 1 - \Phi\left( \frac{MV_{t}^{\text{count}} - \bar{M}V_{R,R} - \eta}{\sigma_2} \right) \right) \phi\left( \frac{\eta}{\sigma_1} \right) d\eta
\]
\[
\frac{\partial}{\partial q_t} (V^\text{count}_D(q_t) - V^\text{no count}_D(q_t)) \\
= \frac{-1}{\sigma_1} \left( u_q(q_t, \theta_{m,D}) + u_x(x^\text{count}_t, \theta_{m,D}) \right) \int_{-\infty}^{\infty} \left( 1 - \Phi\left( \frac{MV^\text{count}_t - MV_{R,R} - \eta}{\sigma_2} \right) \right) \phi\left( \frac{\eta}{\sigma_1} \right) d\eta \\
+ \left( u_q(q_t, \theta_{m,D}) + u_x(x^{\text{no count}}_t, \theta_{m,D}) \right) \left( 1 - \Phi\left( \frac{MV^{\text{no count}}_t - MV_{R,R}}{\sigma} \right) \right)
\]

From the first part of Lemma C.1.1, \(x^{\text{no count}}_t < x^\text{count}_t\), which ensures \(u_x(x^{\text{no count}}_t, \theta_{m,D}) > u_x(x^\text{count}_t, \theta_{m,D})\). Furthermore,

\[
1 - \Phi\left( \frac{MV^{\text{no count}}_t - MV_{R,R}}{\sigma} \right) \\
> 1 - \Phi\left( \frac{MV^\text{count}_t - MV_{R,R}}{\sigma} \right) \\
= \frac{1}{\sigma_1} \int_{-\infty}^{\infty} \left( 1 - \Phi\left( \frac{MV^\text{count}_t - MV_{R,R} - \eta}{\sigma_2} \right) \right) \phi\left( \frac{\eta}{\sigma_1} \right) d\eta \\
> \frac{1}{\sigma_1} \int_{-\infty}^{\infty} \left( 1 - \Phi\left( \frac{MV^{\text{no count}}_t - MV_{R,R} - \eta}{\sigma_2} \right) \right) \phi\left( \frac{\eta}{\sigma_1} \right) d\eta \\
> 0
\]
given \(\eta^1_t < \infty\). Therefore, the difference in expected utility strictly increases with \(q_t\).

For \(q_t > \theta_{m,D}\), we can establish that both value functions increase in \(q_t\), but their difference decreases, in an identical manner. \(\square\)

**Proof of Proposition 4.3.6**

Assume \(C_b < \hat{C}_b\) so that, from Proposition 4.3.4, \(x^\text{count}_t\) is unique. Consider \(q_t < \theta_{m,D}\). We first show that as \(q_t \to \theta_{m,D}\), \(V^\text{no count}_D(q_t) \to -C_b\) and \(V^\text{count}_D(q_t) \to 0\). The first follows from simple inspection of \(EU^\text{no count}_D(q_t, x_t)\), noting that \(x^{\text{no count}}_t\) must approach \(\theta_{m,D}\) as \(q_t \to \theta_{m,D}\) because it is contained in the interval, \((q_t, \theta_{m,D})\), by Proposition 4.3.4. Similarly, inspecting \(EU^\text{count}_D(q_t, x_t)\), we see that \(V^\text{count}_D(q_t) \to - \left( 1 - \Phi\left( \frac{\eta^1_t}{\sigma_1} \right) \right) C_b\). But, as \(q_t \to \theta_{m,D}\),

we can see from (C.5) that \(\eta^1_t\) must approach infinity such that \(\Phi\left( \frac{\eta^1_t}{\sigma_1} \right) \to 1\).

Given these facts, strictly positive costs, and the result of Lemma 4.3.5 that both value functions strictly decrease with \(|q_t - \theta_{m,D}|\), there exists a status quo cutoff, \(\overline{q}_t < \theta_{m,D}\), such that for all \(q_t \in (\overline{q}_t, \theta_{m,D})\), no alternative policy is pursued. Specifically, \(\overline{q}_t\) is given by the larger of the two policies, \(q_1\) and \(q_2\) which satisfy \(V^\text{no count}_D(q_1) = 0\) and \(V^\text{count}_D(q_2) = C_w\).
respectively.

For $q_l < \overline{q}_l$, there are two possibilities. If $q_1 > q_2$, then set $\underline{q}_l = \overline{q}_l = q_1$ with $V_D^{\text{count}}(q_1) < C_w$ and $V_D^{\text{no count}}(q_1) = 0$. In this case, for any $q_l < q_1$, an alternative policy is pursued without a whip count: by Lemma 4.3.5, over this range, $V_D^{\text{no count}}(q_1) > 0$ so that an alternative policy without a whip count is preferred over not pursuing an alternative policy and, as $q_t$ decreases from $q_1$, $V_D^{\text{count}}(q_t) - V_D^{\text{no count}}(q_t)$ decreases so that not conducting a whip count remains more valuable than conducting one.

If $q_1 < q_2$, then set $\overline{q}_l = q_2$ and define $\underline{q}_l < \overline{q}_l$ to be the policy for which $V_D^{\text{count}}(q_2) - C_w = V_D^{\text{no count}}(q_2)$. Such a point must exist because, by Lemma 4.3.5 as $q_t$ decreases from $\overline{q}_l$, $V_D^{\text{count}}(q_t) - V_D^{\text{no count}}(q_t)$ decreases and so must eventually approach zero. Thus, for $q_t$ sufficiently small, $V_D^{\text{count}}(q_t) - C_w < V_D^{\text{no count}}(q_t)$. With these cutoffs, for $q_t \in (-\infty, \underline{q}_l]$, an alternative policy is pursued without a whip count because $V_D^{\text{no count}}(q_t) > V_D^{\text{count}}(q_t) - C_w > 0$ for all $q_t < \underline{q}_l$. For $q_t \in (\underline{q}_l, \overline{q}_l]$, an alternative policy is pursued with a whip count because $V_D^{\text{count}}(q_t) - C_w > 0$ and, by Lemma 4.3.5, $V_D^{\text{count}}(q_t) - V_D^{\text{no count}}(q_t)$ increases with $q_t$ over this range so that $V_D^{\text{count}}(q_t) - C_w > V_D^{\text{no count}}(q_t)$.

Symmetric arguments establish cutoffs, $\underline{q}_r$ and $\overline{q}_r$, for the bill pursuit decisions over the range $q_l > \theta_{m, D}$. □

C.2 The choices from Party “R”

The model in the main text focuses on the decision of partymembers and whips from party “D”, since the analogous results hold for those from party “R”. The same is true with the identification results. However, for completeness and for the estimation procedure, it is useful to write out the inequalities used for “R”, as they differ slightly different from equations (4.1) and (4.5). In the associated likelihood functions, we just substitute by the appropriate probabilities of a yes vote at the whip count and roll call stages.

The difference in the equations comes from party “R” being on the right, while “D” is on the left. This implies that, for “R”, those that vote “Nay” (and need to be whipped) are to the left of the Marginal Voter, while for “D”, those that vote “Nay” to their policies are to the right of the marginal voter.

Hence, at the whip count, a politician from party “R” votes “Yea” if:

$$\delta_{1,t} + \theta_t \geq MV_t - \eta_{1,t}, \quad (C.11)$$

The probability of a yes at the whip count stage is given by:

218
\[ P(Y_{i,t}^{\text{wc}} = 1) = P(\delta_{1,i,t} + \theta_i \geq MV_t - \eta_{1,i}) = P(\delta_{1,i,t} \geq MV_t - \eta_{1,i} - \theta_i) = P(\delta_{1,i,t} \geq \gamma_{1,i} - \theta_i) = 1 - G(\gamma_{1,i} - \theta_i) = 1 - \Phi(\gamma_{1,i} - \theta_i). \] (C.12)

At the Roll Call stage, they vote yes if:

\[ \delta_{1,i,t} + \delta_{2,i,t} + \eta_{1,i} + \eta_{2,i} + \theta_i \geq MV_t - y_{max}^R. \] (C.13)

where \( y_{max}^R \) is allowed to be different than the “D” one in the main section.

Hence, the probability of a yes at the roll count stage is given by:

\[ P(Y_{i,t}^{\text{rc}} = 1) = P(\delta_{1,i,t} + \delta_{2,i,t} + \theta_i \geq MV_t - \eta_{1,i} - \eta_{2,i} - y_{max}^R) = P(\delta_{1,i,t} + \delta_{2,i,t} \geq MV_t - \eta_{1,i} - \eta_{2,i} - \theta_i - y_{max}^R) = P(\delta_{1,i,t} + \delta_{2,i,t} \geq \gamma_{2,i} - \theta_i - y_{max}^R) = 1 - G_1(\gamma_{2,i} - \theta_i - y_{max}^R) = 1 - \Phi(\gamma_{2,i} - \theta_i - y_{max}^R \sqrt{2}), \] (C.14)

where the last line uses the parametric assumptions on \( G_{1+2} \).

For the likelihood of this model for this party, we just replace the above equations into (4.17).

**C.2.1 Agenda Setting**

We focus on illustrating the Republicans’ problem in agenda setting in the case of no whip count, with status quo’s between the party medians. In the case of no whip count, we can rewrite party \( R \)’s expected utility as

\[ EU_{R}^{\text{no count}}(q_{t}, x_{t}) = \left( 1 - \Phi \left( \frac{MV_t - \tilde{MV}_{L,R}}{\sigma} \right) \right) \left( u(x_{t}, \theta_{m,R}) - u(q_{t}, \theta_{m,R}) \right) + u(q_{t}, \theta_{m,R}) - C_b \]
The derivative with respect to \(x_t\) is given by

\[
\left( 1 - \Phi \left( \frac{MV_t - \hat{MV}_{L,R}}{\sigma} \right) \right) u_x(x_t, \theta_{m,R}) - \frac{1}{2\sigma} \phi \left( \frac{MV_t - \hat{MV}_{L,R}}{\sigma} \right) \left( u(x_t, \theta_{m,R}) - u(q_t, \theta_{m,R}) \right)
\]

where \(\phi()\) denotes the pdf of the standard normal distribution. At \(x_t = q_t\), the derivative is strictly positive given \(q_t < \theta_{m,R}\) and the fact that \(\hat{MV}\) is finite. At \(x_t = \theta_{m,R}\), it is strictly decreasing given \(u(q_t, \theta_{m,R}) < 0\). Together these facts ensure an interior solution which is unique, following similar arguments to that for Party “D”. Any interior solution satisfies the first-order condition,

\[
\frac{\left( 1 - \Phi \left( \frac{MV_t^{\text{no cont}} - MV_{KL}}{\sigma} \right) \right)}{\phi \left( \frac{MV_t^{\text{no cont}} - MV_{KL}}{\sigma} \right)} = \frac{1}{2\sigma} \frac{u(x_t^{\text{no cont}}, \theta_{m,R}) - u(q_t, \theta_{m,R})}{u_x(x_t^{\text{no cont}}, \theta_{m,R})}. \tag{C.15}
\]
Appendix D

Appendix to Chapter 5

D.1 Additional Tables and Figures
Notes: The figures present the constructed distribution of votes in the sample, by main political party and by timing of decision. On the y-axis, I show the fraction of voters (from the total amount of votes) who decided for that party at that sub-period. The final vote share for a party is the sum of the bars across all periods. One can see that the late deciders, deciding in the last 2-3 days and in the last day, were fundamental to deciding elections in 2006, 2009 and 2015.
### Table D.1: Summary statistics

#### Quantitative Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (Female = 1)</td>
<td>1919</td>
<td>0.52</td>
<td>0.50</td>
</tr>
<tr>
<td>Age</td>
<td>1898</td>
<td>44.63</td>
<td>17.62</td>
</tr>
<tr>
<td>Education</td>
<td>1897</td>
<td>14.00</td>
<td>3.24</td>
</tr>
<tr>
<td>Ideology</td>
<td>1832</td>
<td>5.57</td>
<td>2.75</td>
</tr>
<tr>
<td>Rooms per Member of the Household</td>
<td>1853</td>
<td>1.36</td>
<td>0.79</td>
</tr>
<tr>
<td>Politically Knowledgeable</td>
<td>1878</td>
<td>0.24</td>
<td>0.48</td>
</tr>
</tbody>
</table>

#### Qualitative Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>Percentage of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watches News on TV:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almost Never</td>
<td>256</td>
<td>13.42</td>
</tr>
<tr>
<td>Not so Often</td>
<td>137</td>
<td>7.18</td>
</tr>
<tr>
<td>Once A Week or So</td>
<td>66</td>
<td>3.46</td>
</tr>
<tr>
<td>2-3 times a week</td>
<td>260</td>
<td>13.63</td>
</tr>
<tr>
<td>At least once a day</td>
<td>789</td>
<td>41.35</td>
</tr>
<tr>
<td>More than once a day</td>
<td>400</td>
<td>20.96</td>
</tr>
<tr>
<td>Total</td>
<td>1908</td>
<td>100</td>
</tr>
<tr>
<td>Language of the Interview:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hebrew</td>
<td>1325</td>
<td>69.05</td>
</tr>
<tr>
<td>Arabic</td>
<td>289</td>
<td>15.06</td>
</tr>
<tr>
<td>Russian</td>
<td>305</td>
<td>15.89</td>
</tr>
<tr>
<td>Total</td>
<td>1919</td>
<td>100</td>
</tr>
<tr>
<td>Religious Observance:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Not at all”</td>
<td>410</td>
<td>21.80</td>
</tr>
<tr>
<td>“A little bit”</td>
<td>890</td>
<td>47.32</td>
</tr>
<tr>
<td>“A lot”</td>
<td>372</td>
<td>19.78</td>
</tr>
<tr>
<td>“I observe all of it”</td>
<td>209</td>
<td>11.11</td>
</tr>
<tr>
<td>Total</td>
<td>1919</td>
<td>100</td>
</tr>
</tbody>
</table>

Notes: The table presents Summary Statistics for the Main Variables that will be used. Education is measured in years of schooling. Ideology is measured in the self-reported value on the scale of Left (0) to Right (10). Rooms in the dwelling per Household member will be used as a proxy for income. In the bottom half of the table, we see descriptive evidence of two variables: Language of the Interview (used as a proxy for Ethnicity), exposure to the media (captured by how often does one watch News on TV) and Religious Observance. Knowledgeable is defined by the correctly knowing the minimum threshold for a party to join the Knesset (2% in 2006), and who the speaker of the Knesset was (R. Rivlin, in 2006). We can see that there is quite a lot of dispersion, as many individuals know both, and many know neither.
Table D.2: Distribution of Votes and Seats in the Knesset, 2006 Elections

<table>
<thead>
<tr>
<th>Party/List</th>
<th>Number of Votes</th>
<th>Number of Seats</th>
<th>% of Votes</th>
<th>% Votes in Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kadima</td>
<td>690,901</td>
<td>29</td>
<td>22</td>
<td>21.59</td>
</tr>
<tr>
<td>Labor</td>
<td>472,366</td>
<td>19</td>
<td>15.1</td>
<td>14.52</td>
</tr>
<tr>
<td>Shas</td>
<td>299,054</td>
<td>12</td>
<td>9.5</td>
<td>4.92</td>
</tr>
<tr>
<td>Likud</td>
<td>281,996</td>
<td>12</td>
<td>9</td>
<td>8.41</td>
</tr>
<tr>
<td>Israel Beitenu</td>
<td>281,880</td>
<td>11</td>
<td>9</td>
<td>12.3</td>
</tr>
<tr>
<td>Ichud Leumi/Mafdal</td>
<td>224,083</td>
<td>9</td>
<td>7.1</td>
<td>6.51</td>
</tr>
<tr>
<td>Torah and Shabbat Judaism(UTJ)</td>
<td>147,091</td>
<td>6</td>
<td>4.7</td>
<td>3.49</td>
</tr>
<tr>
<td>Meretz</td>
<td>118,302</td>
<td>5</td>
<td>3.8</td>
<td>4.37</td>
</tr>
</tbody>
</table>

Number of Eligible Voters (Total) | 5,014,622       |
Valid votes (Total)               | 3,137,064       |
Qualifying threshold (2%)         | 62,742          |
Votes per seat                    | 24,619          |

Notes: The table presents the results of the 2006 Election to the Israeli Parliament (Knesset). Turnout was 63.55%. The first four columns are extracted from the official election results, available online\(^1\). The final column is extracted from our data, from the Panel Surveys conducted by the Israel National Election Studies, Tel Aviv University. They are the empirical shares, as calculated from the survey (post-elections) answer to “list voted in the last election”. They are shown to be close to the true outcome. For further information on the data, see the Data section. I present in this table only the parties and lists that I use in the empirical section. Small parties that did not cross the threshold, or the lists United Arab List (4 seats), Hadash (3 seats), National Democratic Assembly (3 seats) and the Pensioners Party (7 seats); as I do not have measures for their policy vectors.
Table D.3: How Voters Are Deciding

(What will have the) Greatest Effect on your Voting (Pre-Elections Answers)

<table>
<thead>
<tr>
<th>Issue</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Political process with the Palestinians</td>
<td>81</td>
<td>13</td>
</tr>
<tr>
<td>Defence situation</td>
<td>185</td>
<td>29</td>
</tr>
<tr>
<td>Socio-economic situation</td>
<td>197</td>
<td>31</td>
</tr>
<tr>
<td>Relationships between religious and nonreligious</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>Corruption and the rule of law</td>
<td>41</td>
<td>6</td>
</tr>
<tr>
<td>The unity of the people</td>
<td>58</td>
<td>9</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>Do not know</td>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td>More than one answer</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>638</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Main Opinions on what the Elections were About (Post-Elections Answers)

<table>
<thead>
<tr>
<th>Issue</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Political topic (disengagement, negotiations with the Palestinians)</td>
<td>257</td>
<td>18</td>
</tr>
<tr>
<td>The security topic</td>
<td>280</td>
<td>20</td>
</tr>
<tr>
<td>The economic-social topic</td>
<td>687</td>
<td>49</td>
</tr>
<tr>
<td>Corruption</td>
<td>126</td>
<td>9</td>
</tr>
<tr>
<td>Other</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>All the topics</td>
<td>26</td>
<td>2</td>
</tr>
<tr>
<td>None</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,404</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Notes: The table presents the distribution of Issues that voters considered were important for them (in the first survey), and what they thought was significant in the elections (post-election). We can see that these mostly revolve around policy issues, such as economy and security. Most voters consider that security and economy issues are the most important ones for their voting decision and expect that to be the case from others. This motivates our theoretical approach using a state of nature which citizens wish to learn about.
**Table D.4:** Israeli Civilian Fatalities in Terrorist Attacks, 2003-2005

<table>
<thead>
<tr>
<th>City</th>
<th>Number of Israeli Civilian Fatalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afula</td>
<td>3</td>
</tr>
<tr>
<td>Ashdod</td>
<td>10</td>
</tr>
<tr>
<td>Baqah al-Gharbiyah</td>
<td>1</td>
</tr>
<tr>
<td>Beersheva</td>
<td>16</td>
</tr>
<tr>
<td>Erez (Industrial Zone)</td>
<td>1</td>
</tr>
<tr>
<td>Gadish</td>
<td>1</td>
</tr>
<tr>
<td>Hadera</td>
<td>7</td>
</tr>
<tr>
<td>Haifa</td>
<td>36</td>
</tr>
<tr>
<td>Jerusalem</td>
<td>59</td>
</tr>
<tr>
<td>Kfar Sava</td>
<td>1</td>
</tr>
<tr>
<td>Kfar Ya’bez</td>
<td>1</td>
</tr>
<tr>
<td>Kibbutz Beit Govrin</td>
<td>1</td>
</tr>
<tr>
<td>Kibbutz Eyal</td>
<td>1</td>
</tr>
<tr>
<td>Lahav</td>
<td>2</td>
</tr>
<tr>
<td>Moshav Nehusha</td>
<td>1</td>
</tr>
<tr>
<td>Netanya</td>
<td>9</td>
</tr>
<tr>
<td>Netiv Ha’asara</td>
<td>1</td>
</tr>
<tr>
<td>Petah Tikva</td>
<td>1</td>
</tr>
<tr>
<td>Rosh Ha’ayin</td>
<td>1</td>
</tr>
<tr>
<td>Sde Trumot</td>
<td>1</td>
</tr>
<tr>
<td>Sderot</td>
<td>5</td>
</tr>
<tr>
<td>Tel Aviv-Yafo</td>
<td>23</td>
</tr>
</tbody>
</table>

Notes: The table shows the summary statistics for the data used from the NGO B’Tselem, about the number of fatalities of Israeli civilians between January 2003 - December 2005 (just before the campaign begins). This data is then merged with the survey data used in the rest of the model by the voters’ cities of residence.
### Table D.5: Who stops earlier?

<table>
<thead>
<tr>
<th></th>
<th>(1) Full Sample</th>
<th>(2) Voters Deciding During the Campaign</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ordered Logit</td>
<td>Last Week</td>
</tr>
<tr>
<td>(Ideology-5)</td>
<td>0.0304</td>
<td>0.00763</td>
</tr>
<tr>
<td></td>
<td>(0.0231)</td>
<td>(0.00562)</td>
</tr>
<tr>
<td>(Ideology − 5)^2</td>
<td>-0.0162**</td>
<td>-0.00313**</td>
</tr>
<tr>
<td></td>
<td>(0.00640)</td>
<td>(0.00154)</td>
</tr>
<tr>
<td>Education</td>
<td>-0.0217</td>
<td>-0.00812*</td>
</tr>
<tr>
<td></td>
<td>(0.0181)</td>
<td>(0.00462)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0170**</td>
<td>-0.00408**</td>
</tr>
<tr>
<td></td>
<td>(0.0377)</td>
<td>(0.00947)</td>
</tr>
<tr>
<td>Gender (Female)</td>
<td>-0.119</td>
<td>-0.00449</td>
</tr>
<tr>
<td></td>
<td>(0.113)</td>
<td>(0.0286)</td>
</tr>
<tr>
<td>Religiosity</td>
<td>-0.0493</td>
<td>-0.00495</td>
</tr>
<tr>
<td>Observes “A little bit”</td>
<td>(0.134)</td>
<td>(0.0358)</td>
</tr>
<tr>
<td>Religiosity</td>
<td>-0.241</td>
<td>-0.0488</td>
</tr>
<tr>
<td>Observes “A lot”</td>
<td>(0.181)</td>
<td>(0.0463)</td>
</tr>
<tr>
<td>Religiosity</td>
<td>-1.123***</td>
<td>-0.204***</td>
</tr>
<tr>
<td>Observes “All of it”</td>
<td>(0.301)</td>
<td>(0.0649)</td>
</tr>
<tr>
<td>Language</td>
<td>-0.928***</td>
<td>-0.160***</td>
</tr>
<tr>
<td>(Arabic)</td>
<td>(0.221)</td>
<td>(0.0496)</td>
</tr>
<tr>
<td>Language</td>
<td>-1.073***</td>
<td>-0.208***</td>
</tr>
<tr>
<td>(Russian)</td>
<td>(0.171)</td>
<td>(0.0423)</td>
</tr>
<tr>
<td>Rooms Per</td>
<td>0.0268</td>
<td>0.0209</td>
</tr>
<tr>
<td>Household Member</td>
<td>(0.0734)</td>
<td>(0.0183)</td>
</tr>
<tr>
<td>Knowledgeable</td>
<td>-0.352***</td>
<td>-0.120***</td>
</tr>
<tr>
<td></td>
<td>(0.123)</td>
<td>(0.0319)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>1206</th>
<th>1206</th>
<th>1206</th>
<th>677</th>
<th>677</th>
<th>677</th>
</tr>
</thead>
<tbody>
<tr>
<td>R^2</td>
<td>0.080</td>
<td>0.077</td>
<td>0.077</td>
<td>0.088</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Robust Standard errors in parentheses.

* p < 0.10, ** p < 0.05, *** p < 0.01

Notes: The table presents Ordered Logit (Columns (1), (4)) and OLS results (remaining columns) on the characteristics of voters who decide at different points in time. For Columns (1) and (4), I present ordered logit regressions, where the outcome is “when the individual decided on whom to vote”, as shown in Figure D.1a. The higher values refer to deciding later (i.e. last day is the highest group, with 2-3 days before being the second highest; and “knew all along” is the lowest ordered category). The other columns use the dependent variable Last Week, defined as 1 if the voter decided in the last week (and 0 if not), and Last Day (1 if the voter decided in the last day). Columns (4)-(6) condition on those who have acquired some information according to the model (i.e. did not answer “knew all along”), according to our extension. The variable Rooms per Household member refers to Rooms in the dwelling divided by members of the household, which is a proxy for income since income is not stated. The omitted categories for the categorical variables are: “not at all” (for Religious observance), Hebrew (for Language) and Male (for Gender). The regressions also control for Media Exposure, as defined by how often they watch news on TV. These are not significant and are not shown for parsimony.
Proof. **Proof of Lemmas 5.2.1, 5.2.3:** For the expressions in both lemmas, it follows from [59], Section 11.

The comparative statics for 5.2.1 are immediate, (with the last one being strict for \( m_i > 0 \)):

\[
\frac{\partial \text{Var}(x | H_i, t)}{\partial m_i} = -\frac{\sigma^2}{(m_i + \tau \sigma^2)^2} < 0 \tag{D.1}
\]

\[
\frac{\partial \text{Var}(x | H_i, t)}{\partial \tau} = -\frac{(\sigma^2)^2}{(m_i + \tau \sigma^2)^2} < 0 \tag{D.2}
\]

\[
\frac{\partial \text{Var}(x | H_i, t)}{\partial \sigma^2} = \frac{m_i + \tau \sigma^2 - \tau \sigma^2}{(m_i + \tau \sigma^2)^2} > 0 \tag{D.3}
\]

The comparative statics for 5.2.3 have unclear sign, because they are all multiples of \( \sum_{i=1}^{m_i} e_{i,t} \) which has an unclear sign.

\( \square \)

**Proof. Proof of Lemma 5.2.2**

I use that \( \mathbb{E}(x | H_{i,t}) \) cannot be a function of \( t \). Indeed, if that was the case, then the voter would know which direction the expectation would be at \( t+1 \), even at period \( t \). However, since \( e_{i,t+1} \) is unknown at \( t \) and is i.i.d., from the expression in Lemma 5.2.3, that cannot be the case.

By the timing of the model, the first term is the state at \( t \) (since voting will only occur at \( T \)), so only the term:

\[-\text{Var}_i[x | H_{i,t}] - c_i y_{i,t} \]

matters in the decision of accumulation.

From [59], Section 12, Theorem 1 (on p. 285), we have that since the risk (of a sequential decision procedure with a Normal distribution process) does not depend on the values of the observations (as the variance does not, from 5.2.1), the optimal sequential decision procedure is given by a procedure in which a fixed number of decisions will be taken.

Hence, she will choose to acquire a signal at \( t \) if and only if:

\[
0 < -\text{Var}_i[x | H_{i,t} \cup \{e_{i,t+1}\}] - c_i - (-\text{Var}_i[x | H_{i,t}]) \tag{D.4}
\]

\[
c_i < \text{Var}_i[x | H_{i,t}] - \text{Var}_i[x | H_{i,t} \cup \{e_{i,t+1}\}] \tag{D.5}
\]
In the case of a normal distribution, with a quadratic loss function, this is given by the equation stated in our Lemma (see p. 260, replacing \( r \) by \( 1/\sigma^2 \)). \( \square \)

**Proof. Proof of Corollary 5.2.2.1**

First note that, from Lemma 5.2.1, the value of the variance of the beliefs in \( x \) is strictly decreasing in the number of signals. It follows that, if a voter stops acquiring information at some \( t^* \), it will not be worth it again to acquire at some \( t > t^* \). We drop the index \( i \) in \( \tau_i \) for simplicity.

From (D.5), a voter will acquire a signal at \( 0 < t < T - 1 \), but not at \( t + 1 \) if and only if:

\[ c_i < \text{Var}_i [x | \mathcal{H}_{i,t} - \text{Var}_i [x | \mathcal{H}_{i,t} \cup \{e_{i,t+1}\}] \]

and

\[ c_i > \text{Var}_i [x | \mathcal{H}_{i,t+1} - \text{Var}_i [x | \mathcal{H}_{i,t} \cup \{e_{i,t+2}\}] \]

Replacing the values from Lemma 5.2.1:

\[ c_i < \frac{\sigma^2}{t - 1 + \tau \sigma^2} - \frac{\sigma^2}{t + \tau \sigma^2} \]

and

\[ c_i > \frac{\sigma^2}{t + \tau \sigma^2} - \frac{\sigma^2}{t + 1 + \tau \sigma^2} \]

and hence, the stopping time would be \( m = t \).

She will stop at \( t = 0 \) if it is not worth to acquire a signal at \( t = 0 \). That means the cost must be higher than the gain at \( 0 \) (the difference between the prior and the variance of beliefs with one signal, which leads to:

\[ c_i > \frac{1}{\tau} - \frac{\sigma^2}{1 + \tau \sigma^2} \]

Similarly, the voter will acquire signals at every period (and hence, stop at \( T \)) if it is worth it to acquire signals at \( T - 1 \). Since the gains of a signal are strictly decreasing over time, this would mean that signals are acquired at every period. For the citizen to acquire at \( T - 1 \), it must be that, from (D.5):

\[ c_i < \frac{\sigma^2}{T - 1 + \tau \sigma^2} - \frac{\sigma^2}{T + \tau \sigma^2} \]

\( \square \)

**Proof. Proof of Lemmas 5.2.4:** From Lemma 5.2.2 we see that if \( c_i \) increases, then the
voter cannot choose more information; since the right hand side is strictly decreasing in the number of signals. For the second part, I will focus on the case in which \( b_i > 0 \). Let \( q \) be such that \( a_q = \max\{a_1, \ldots, a_J\} \). Let \( b_i \) be such that the agent chooses “In”. Then as \( b_i \to \infty \), it is clear that \( -2b_i(a_q - a_i) \to -\infty \). This, in turn, implies that the product over \( k \) goes to 0, as \( \Phi(-\infty) = 0 \), and hence the whole term converges to \( -\kappa < 0 \), so the agent will always be out.

Note that, conditional on being “In”, the acquisition problem does not depend upon \( b_i \), due to (5.2.2).

\[\text{Lemma D.2.1.} \text{ We now find the probabilities of stopping at a period and voting for a given candidate (given an information history). These are used in our Maximum Likelihood estimation and are based on the parametrization} \ (5.10) \text{, together with the distribution of the information signals.} \]

\[
P(t_i = t \mid I_n, z_i, x; \theta) = \begin{cases} 
1 - \Phi\left(\frac{1}{\sigma_i} (\ln(\frac{1}{1+\xi} - \frac{\sigma_i^2}{\tau - 1 + 5 \sigma_i^2}) - \frac{\sigma_i^2}{\tau - 1 + 5 \sigma_i^2}) - \frac{\sigma_i^2}{\tau - 1 + 5 \sigma_i^2} - \frac{\sigma_i^2}{\tau - 1 + 5 \sigma_i^2} - \frac{\sigma_i^2}{\tau - 1 + 5 \sigma_i^2}\right) & \text{if } t = 0 \\
\Phi\left(\frac{1}{\sigma_i} (\ln(\frac{\sigma_i^2}{\tau - 1 + 5 \sigma_i^2} - \frac{\sigma_i^2}{\tau - 1 + 5 \sigma_i^2}) - \frac{\sigma_i^2}{\tau - 1 + 5 \sigma_i^2} - \frac{\sigma_i^2}{\tau - 1 + 5 \sigma_i^2} - \frac{\sigma_i^2}{\tau - 1 + 5 \sigma_i^2}\right) & \text{if } t = T, \end{cases} \quad 0 < t < T
\]

where \( \Phi(\cdot) \) is the CDF of \( \mathcal{N}(0, 1) \), and:

\[
P(v_i = j \mid t_i = t, z_i, x; \theta) = \begin{cases} 
1, & \text{if } t = 0, j = \arg\max_{k \in \{1, \ldots, J\}} (a_k - b_i - \mu_i)^2 \\
0, & \text{if } t = 0, j \neq \arg\max_{k \in \{1, \ldots, J\}} (a_k - b_i - \mu_i)^2 \\
\prod_{k \neq j} \Phi\left(\frac{1}{\sigma_i} (a_k^2 - a_j^2 - 2b_i(a_k - a_j)) - \frac{(\tau^2 + 5 \mu_i^2)(a_k^2 - a_j^2)}{\mu_i^2 \sqrt{\tau^2 + 5 \mu_i^2}}\right), & \text{if } t > 0, j \neq J \\
1 - \sum_{j = 1}^{J} \prod_{k \neq j} \Phi\left(\frac{1}{\sigma_i} (a_k^2 - a_j^2 - 2b_i(a_k - a_j)) - \frac{(\tau^2 + 5 \mu_i^2)(a_k^2 - a_j^2)}{\mu_i^2 \sqrt{\tau^2 + 5 \mu_i^2}}\right), & \text{if else} \end{cases}
\]

In the second part of the Lemma above, we construct the solutions for the probability of choosing a party. In the simplest case, when no information is acquired and no randomness from signals arrives, then the choice is deterministic given the ideology. For the other cases, it must be the case that the utility from one party is higher than from all the others ones. For that, we use an approximation. That is warranted for the maximum likelihood since the model is (i) identified without this, (ii) the dependency between choices is a nuisance parameter compared to our main analysis.

\[\text{Proof. Proof of Lemmas D.2.1:} \text{ An immediate substitution of} \ (5.10) \text{ in} \ (5.2.2.1) \text{ yields the expression above for} \ P(t_i^* = t \mid z_i, x; \theta). \text{ This determines the second term of the likelihood. Note that this is similar to a hazard model, with the continuation being that one has} \]

231
P(v_j^* = j \mid t^*_j = t, t^*_i > 0, x, z_i; \theta) = P(-\mathbb{E}_i [x \mid \mathcal{H}_i] - b_i)^2 > \max_{k \neq j} (a_k - b_i - \mathbb{E}_i [x \mid \mathcal{H}_i])^2)
\approx \prod_{k \neq j} P(-\mathbb{E}_i [x \mid \mathcal{H}_i] - b_i)^2 > (a_k - b_i - \mathbb{E}_i [x \mid \mathcal{H}_i])^2
= \prod_{k \neq j} P \left( \frac{1}{2} (a_k^2 - a_j^2 - 2b_i(a_k - a_j)) > (a_k - a_j)\mathbb{E}_i [x \mid \mathcal{H}_i] \right)

where the second line uses that \{a_k\} are fixed and the signals are i.i.d.

Using (5.2.3), we know that:

E_i (x \mid \mathcal{H}_i) (a_k - a_j) = \frac{(\sum_{m=1}^{t^*_i} e_{i,m} + \tau_i \mu_i \sigma^2)}{t^*_i + \tau_i \sigma^2} (a_k - a_j) \sim \mathcal{N} \left( \frac{(t^*_i x + \tau_i \mu_i \sigma^2) (a_k - a_j)}{t^*_i + \tau_i \sigma^2}, \frac{(a_k - a_j)^2 t^*_i \sigma^2}{(t^*_i + \tau_i \sigma^2)^2} \right)

where I have used that e_{i,m} \sim \mathcal{N}(x, \sigma^2) \quad \forall m, and it is i.i.d. over time. Since the probabilities must sum up to 1, I take j = J as the final option, and that the probability for voting J is given by one minus the others, when t^*_j > 0.

It follows that:

P(v_i^* = j \mid t_i^* = t, x, \theta) \approx \begin{cases} \prod_{k \neq j} \Phi \left( \frac{1}{2} (a_k^2 - a_j^2 - 2b_i(a_k - a_j)) \frac{t + \tau_i \sigma^2}{|a_k - a_j| \sqrt{\tau_i \sigma^2}} \right) & \text{if } t^*_i > 0, j \neq J \\ 1 & \text{if } t^*_i = 0 \quad \text{and } j = \arg \max_{k \in \{1, \ldots, J\}} \\ 0 & \text{if } t^*_i = 0 \quad \text{and } j \neq \arg \max_{k \in \{1, \ldots, J\}} \end{cases}

(D.8)

If t^*_i = 0, then there is no uncertainty (as no signals are received); and hence the probability of choosing j is given by 1 if j is closest to the ideology; as she solves:

\max_{a \in \{a_1, \ldots, a_J\}} -(a - b_i - \mu_i)^2

(D.9)

Hence, (D.7) and (D.6) in (5.12) gives us the (quasi) likelihood function.
D.3 Israeli Political System

The empirical section of this paper will focus on data from Israel, in particular data from 2006. I briefly introduce the Israeli political system which is the environment for my empirical study. Israel is a multiparty parliamentary democracy, with one chamber (the Knesset). The Knesset has 120 seats that are distributed according to proportional representation, although subject to a threshold to join Congress (which, between 2006 and 2015, went from 2% to 3.25%).

The Prime Minister is the leader of government, and the President the Head of State. Voters choose parties nationally when they vote, as parties present a closed national list. The Israeli system is known to present a fractured party system, which makes coalition formation a bargaining process between many parties, to guarantee the minimum majority of 61 seats in government (see [56]).

Major parties are currently the right-wing Likud, and the left leaning Labor Party; complemented by other parties along ethnic or religious lines. These can be Arab parties, parties from the large Russian immigrant population, or Jewish Orthodox parties. In 2006, Kadima was also a large centrist party, which had been formed from moderates from Likud in 2005. Table D.2 shows the distribution of seats in the Knesset, in 2006. It also shows how our sampled data matches the population results.

D.3.1 Israel in 2006

It is important to understand the context in which elections and campaigning were being held in 2006. The model and the identification will fundamentally use the ideas that this was a national campaign, with (pragmatic) voters that cared about issues and learning about the state of the country’s security and economy. The context upon which this happens, with variation in exposure of information about security to the population spread around the country will also be used in Section 7. It will also be important that the timing of the election did not reflect strategic decisions about terrorism ([6]) or about other policy concerns. I now review some key facts from that time period. [14] provides a more extensive review.

In 2005, the government was led by the right-wing Likud and its leader, Prime Minister Ariel Sharon. The government was held by a coalition of its two largest parties, Labor and Likud, following a series of coalition changes due to budget or policy differences in 2004. At the top of the agenda was Sharon’s decision to unilaterally withdraw settlers from the Gaza Strip and from (some) of the settlements in the West Bank, which split the Likud party.

233
This followed the end of the Second Palestinian Intifada (uprising), which had started in 2000 and seemed to end in late 2004/early 2005, following the death of Yasser Arafat and the Sharm-el-Sheikh conference in February 2005. It had caused large losses of life and economic losses, both on the Palestinian and Israeli sides.\footnote{According to the NGO B’Tselem (whose data will be described later), there were 3295 Palestinians killed by Israeli security forces in the occupied territories, while there were 686 Israeli civilians killed by Palestinians (228 in the occupied territories, 458 within Israeli territory) and (226 in the occupied territories, 84 within Israel).}

Meanwhile, Labor chose a new leader in November, 2005. Amir Peretz very narrowly beat Shimon Peres, on a platform of pulling the Labor Party from the coalition government, and returning it to its socialist economic roots and policies. The result was 41.5% to 40.7% and unexpected, as viewed by media around the world at the time (for e.g., [163]).

As [137] discusses, with no more government, Sharon then said he would no longer deal with the rebels at Likud and created a new centrist party (Kadima). New elections were then called for March 28, 2006. In the beginning of the campaign, however, Sharon suffered a stroke and the leadership of Kadima was transitioned to Ehud Olmert.

D.4 First Stage Estimation of the Priors

I use equation (5.16) in a first step to estimate the priors. This greatly simplifies the problem, while resulting in only minor efficiency losses. To do so, I use the first principal component of the matrix of observed characteristics $z_{it}$, which are the same variables as those used in the cost function. This first component in the data corresponds to over 95% of the variation.

A naive estimation of $1/\tau_i$ could be from $z_i^\gamma$. In finite samples, it can be the case that this value is negative. However, note from equation (5.13) that:

$$\frac{\mu'_i - \mu_i}{\mu_i} = \frac{1}{\tau_i} t_i,$$

which can be positive or negative, depending upon the direction that the updates go towards (to or away from the original prior). In that case, I wish to capture both sides of the variation. Hence, I use

$$\frac{1}{\tau_i} = |z_i^\gamma|$$

as the estimate.

D.5 Optimization Routine

To find the results presented, I implement the following routine, based upon the use of a global optimizer (Genetic Algorithm), supplemented by local optimization algorithms.

As described, I begin with the first stage estimate \( \{\mu_i\}_{i=1}^N \), based upon the standard
deviation of the answers in the sample to the voters’ beliefs about the current situation of Israel. I then estimate the values for \( \{\tau_i\}_{i=1}^N \) given our treatment procedure, using terrorist attacks and updating beliefs prior to the campaign, given in Section 7.

Given the priors, I construct the likelihood given in (5.17), with the data \((v^*_i, t^*_i, z'_i, b_i)\).

The Genetic Algorithm is used to find a candidate global maximum to the likelihood, with candidate estimates \((\hat{\sigma}_{GA}, \hat{\sigma}_{\eta,GA}, \hat{\beta}_{GA})\), where the underscore denotes it is the first stage estimates from the Genetic Algorithm.

I follow that with a local optimizer, constraining \( \sigma \geq 0, \sigma_{\eta} \geq 0 \) and using the first stage estimates as starting values. Then I consider the estimates of our parameters to be \( \hat{\theta} = (\hat{\sigma}_L, \hat{\sigma}_{\eta,L}, \hat{\beta}_L, \hat{\mu}_i) \), where the underscore denotes that they are from the local optimizer, having used the Genetic Algorithm estimates as starting values.