Rally The Vote: Electoral Competition With Direct Campaign Communication

Click here for most recent version

Anubhav Jha*

Abstract

Political rallies have formed a large part of U.S. electoral campaigns since the 19th century and remain relevant today. This paper models candidates' rally decisions as an empirical dynamic game of electoral competition and applies it to estimate rally effectiveness for the 2012 and 2016 U.S. presidential elections. The model supports three empirical patterns. As the election approaches, candidates rally more, concentrate on tight state-level races, and within those tight races, they hold more rallies in states with more electoral college votes. Model parameter estimates uncover that rallies by presidential candidates were effective in increasing their poll margin lead over their opponent. The estimates also reveal that a rally by a presidential candidate is more persuasive than a television ad. I construct and execute model selection tests that infer whether candidates are strategic and forward-looking to validate model assumptions. Counterfactual exercises show that Trump's rallies were electorally pivotal, while rallies by other candidates had no effect on their chances of winning. The effects of short-term campaign silences (i.e., forbidding political campaigning) are limited since candidates can gain sufficient support from the electorate before campaign silences begin.

*Department of Politics, Princeton University. Email: anubhavpcjha@gmail.com I am grateful and indebted to my PhD advisors Francesco Trebbi, Matilde Bombardini, Vitor Farinha Luz and Paul Schrimpf for their constant support and guidance. I would also like to thank the Dev/PE, Econometrics, and Theory group at Vancouver School of Economics, University of British Columbia and the BPER lab at University of California Berkeley. In particular I thank Siwan Anderson, Ernesto Dal Bo, Claudio Ferraz, Frederico Finan, Patrick Francois, Arkadev Ghosh, Sudipta Ghosh, Sam Hwang, Hiro Kasahara, Wei Li, Andrew Little, Matt Lowe, Vadim Marmer, Adlai Newson, Nathan Nunn, Mike Peters, Juan Felipe Riaño, Federico Ricca, Gerard Rolland, Sergei Severinov, Kevin Song and Munir Squires for useful suggestions and comments. All mistakes are my own. Researcher's own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business., The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

1 Introduction

Among all methods of persuasion used by politicians, few are as old as political rallies. Their origin can be traced back to *oratory* and *rhetoric* in ancient democracies. Career politicians in the Roman Republic, such as Cicero, regularly performed *oratories* at *contios*[1]— informal public meetings where Roman magistrates addressed the people (van der Blom, 2016). In ancient Greece, *rhetorics* were delivered at special sites called *bouleutêria*[2]— Ancient Greek counterparts of massive auditoriums (Johnstone and Graff, 2018). It was not until the late 19th century that political rallies became an important electioneering tool at a large-scale. William Jennings Bryan used the railway network to travel 18,000 miles across the U.S. to give speeches and make other appearances to the public in 1896 (Buggle and Vlachos, 2022; Bryan, 1909). This practice was later utilized by Harry Truman and Thomas Dewey in their 1948 U.S. presidential campaigns (Heersink and Peterson, 2017; Donaldson, 1999).

In the internet age, Donald Trump's rallies had an average attendance of 5,505 during the 2016 fall campaign.[3] Nine of these rallies had more than 10,000 attendees. In the fall campaigns of 2012 and 2016, political rallies constituted 44.5%[4] of all campaign activities involving presidential candidates (fundraisers followed at 17.4%). Political rallies are also prevalent in the developing world. For instance, a rally in the Indian city of Kolkata in South Asia had half a million attendees (Al Jazeera, 2019). In Tanzania, rallies are a more commonly used campaigning instrument than canvassing (Paget, 2019). In Latin America, specifically Ecuador and Argentina, rallies form essential features of campaigns (De la Torre and Conaghan, 2009; Szwarcberg, 2012).

Even though rallies are a favored campaigning instrument and a direct form of political communication, systemic evidence on their importance is limited. The lack of evidence on political rallies dramatically contrasts with the work on the efficacy of political advertising (Gordon and Hartmann, 2013; Hill et al., 2013; Gerber et al., 2011; Spenkuch et al., 2018), strategic advertising allocations (Erikson and Palfrey, 2000; Gordon and Hartmann, 2016; Snyder, 1989), and also dynamic inter and intra-electoral spending (Acharya et al., 2022; de Roos and Sarafidis, 2018; Kawai and Sunada, 2022). Empirical work on political rallies has proven challenging due to endogenous rally decisions, measurement error, candidate level heterogeneity, and small sample sizes. These concerns are complex to address.[5] Theoretical

[1]Of all oratorical occasions, these provided the most flexible access to the Roman citizenry, which had an ultimate say in elections and legislation. For more details see van der Blom (2016).

[2]See Johnstone and Graff (2018) for images of 3D reconstruction and description of these historical sites. I would also like to quote the article's main results:

"...we are able to provide empirically grounded accounts of how various settings for boulê meetings actually functioned as auditoriums ... in most cases, Classical and Hellenistic bouleutêria served as excellent venues for oratorical performances before audiences ranging from a few hundred to over a thousand."

[3]This figure is calculated using news reports on individual rallies from multiple news providers. Complete details on sources of each Trump's rally can be provided upon request.

[4]I used candidate calendars made available by Appleman (2012, 2016) for calculating this figure.

[5]For instance, Shaw and Gimpel (2012) randomized a gubernatorial candidate's visit locations in Texas but not the opponent's visit locations. More recently, Snyder and Yousaf (2020) did an event study at the media market level by using Cooperative Election Study surveys. The authors find significant effectiveness for Trump's rallies but not for others. Due to a

work is also challenging due to multiple equilibria arising from the finite time horizon in these settings.

This paper makes four contributions that improve our understanding of political rallies. The first contribution is to provide an economic model of intra-electoral competition where politicians campaign by holding rallies. In this model, campaigning effects decay over time, ensuring that earlier rallies are less effective than those held closer to election day. The model supports a perfect information structure, which implies that an equilibrium exists and it is essentially unique (i.e., multiplicity in the model is of probability zero).

The second contribution is to provide estimates of rally effectiveness. The identification problem at the core of most of the reduced-form literature is that the estimator of rally effectiveness may be biased downward because candidates may be more likely to rally in states where they need to boost their popularity. This selection would underbias estimates, making rallies appear ineffective, which is a common finding in this literature. In this empirical game of dynamic electoral competition, factors like the contemporaneous rally decision of opponents, net popularity gains due to candidates' past choices, time to the election, and the relative popularity of candidates across the different states, all enter the rallying decisions of candidates. So, for example, popularity shocks and past actions in other regions provide identifying variation for a candidate's actions in a given state.

The third contribution of the paper is to estimate the electoral effects of political rallies by executing counterfactual experiments that can uncover the effect of *total rallying* on vote shares and the winning probability of a given candidate. The fourth contribution is to execute counterfactual experiments that show government interventions, such as campaign silence laws, may not always succeed in regulating the use of political rallies.

Constructing empirical games possessing a finite time horizon is challenging. One vital issue is the existence of multiple equilibria.[6] So far, most dynamic campaigning models have considered inter-electoral (or infinite horizon) settings (Kawai and Sunada, 2022; Polborn and Yi, 2006; Gul and Pesendorfer, 2012), rather than intra-electoral settings. Models that consider campaigning within an election have had to settle for a unique normalization over equilibrium strategies (Acharya et al., 2022), as a unique equilibrium is harder to support. As a result, these models lack predictions of candidate-specific strategies, which is critical for understanding candidate level differences and counterfactual analysis.

One of the main objectives of this paper is to construct an empirical game that is sufficiently tractable to permit estimation and inference. Moreover, we are interested in studying each election separately, and therefore we observe only one game. I deviate from the approaches used in firm entry/exit games (Aguirregabiria and Mira, 2007; Arcidiacono et al., 2016)[7] and use stage games as the unit of observation. In this approach, the number of observations grows with the time horizon of the game.

To this end, this paper presents a dynamic game of perfect information with a finite time horizon.

low number of respondents in CES surveys at the day×media market level, their measures of intention to vote for a candidate carry additional noise, thereby increasing the underlying variance, which is carried into their estimates. Moreover, SUTVA is harder to maintain at media market level analysis as there can be geographical spillovers.

[6]For instance, Watanabe and Yamashita (2017) showed that to obtain a tractable set of Markov perfect equilibria, one requires strict assumptions that are not feasible for a flexible empirical approach.

[7]Firm entry/exit games rely on many games for inference and are unsuitable for studying each election separately.

In this game, office-seeking candidates are subject to electoral competition, while facing regional differences and dynamic uncertainty in their popularity. Dynamic uncertainty accommodates unforeseen circumstances in electoral races that lead to a candidate jumping ahead or falling behind his opponent. Regional differences address state-specific factors, such as a state's natural inclination towards a party or a regional popularity shock. On election day, if a candidate's popularity in a state is positive, they receive a payoff proportional to the state's electoral college votes. A candidate's electoral payoff in the election is the sum of these state-specific payoffs. In a given period, candidates can hold a rally in a state and increase their current popularity in that state. This state-specific popularity, which I will call local popularity, is modeled as an AR(1) process. The autocorrelation of this process allows current rallies to affect future popularity. However, the effects of rallies dissipate over time, and the magnitude of this dissipation (decay)[8] maps one-to-one with the autocorrelation parameter, which I call persistence in popularity.

In this game, rallies are costly indivisible goods, such that the costs of these goods vary across candidates and states to address heterogeneity along these two dimensions. In each period, I assume that candidates move in a stochastic order and are equally likely to be the first or the second mover. The first assumption provides a perfect information structure and gives us rally choice probabilities, which are uniquely solvable using backward induction. The second assumption implies that no candidate has an ex-ante first or second mover advantage.

The model provides comparative statics and the intuition for identifying model parameters. For instance, an increase in persistence in popularity also increases the likelihood of earlier rallies by candidates. Intuitively, the induced lower decay rate allows campaigning effects to last longer (Hill et al., 2013; Gerber et al., 2011; Acharya et al., 2022), which incentivizes candidates to hold a higher number of initial rallies. An increase in the cost of rallies introduces a downward level shift in the probability of rallying. The effectiveness parameter exhibits a non-monotonic relationship with probability of rallying. With higher effectiveness, candidates can maintain a sufficient level of popularity with fewer rallies and, therefore, optimally choose to reduce the number of rallies.

For my empirical application, I use two data sources. I use candidate calendars provided in Appleman (2012) and Appleman (2016) for rally locations and dates. I use state-specific poll margins provided by FiveThirtyEight for local popularity. I document that politicians increase political rallies in areas where competition is neck and neck as elections approach. This pattern holds individually for all candidates. I also show that within swing states, the correlation of rally choices with electoral college votes increases as the election approaches. These two patterns highlight that candidates gradually prioritize electoral size and poll margin leads while holding rallies. The model also supports these patterns.

To estimate the model parameters, I leverage the Markov Property obeyed by local popularity to prove that daily rally decisions and poll margins must also obey the Markov Property. I further characterize the transition density of daily observations. This transition density is time inhomogeneous (Pouzo et al., 2022; Ailliot and Pene, 2015) because the equilibrium choice probabilities depend on the number of days left before the election. The transition density provides the means to construct the likelihood function,

[8] Decay in campaigning effects has been documented by Hill et al. (2013); Gerber et al. (2011); Acharya et al. (2022).

which is used in estimation.

This paper finds that Trump's rallies increased his poll margin lead by 0.084 pp[9] in a state, while Clinton's rallies increased her poll margin lead by 0.075 pp. For the 2012 election, I find that Romney and Obama's rallies increased poll margin lead by 0.073 pp, and 0.065 pp pp, respectively. I also find that in both elections, the combined effect of rallies by both candidates is insignificant. This is an assumption made in Strömberg (2008) for the 2000 and 2004 elections, which I confirm empirically for both the 2012 and 2016 elections. I also estimate the weekly decay rate and find it to be 28%. Which is higher than the perceived decay rate in Acharya et al. (2022) and lower than decay rates in Hill et al. (2013) and Gerber et al. (2011). I also find that Clinton had the highest rally cost, while Trump had the lowest.

I then compare the persuasive effects of political rallies with those of other tools of political persuasion. I discover that the persuasion rate of one rally far exceeds the persuasion rate of a T.V. ad. The persuasion rate for a Trump's rally is 0.167%, and that of a Republican T.V. ad was 0.01% (Spenkuch and Toniatti, 2018). This implies that to compensate for one MAGA rally, Trump would have required 17 ad spots in a media market. However, since rallies are harder to scale,[10] the cumulative effect of rallies is lower than that of a T.V. ad.[11] Rallies are less persuasive than *slanted news* (DellaVigna and Kaplan, 2007). In particular, I find that a republican presidential candidate would need 67 additional rallies to compensate for the absence of Fox News in a media market.

I further validate the model and the estimated model parameters by executing three exercises. In the first exercise, I analyze the out-of-sample performance of the model and find that the model correctly predicts at least 76% rally decisions for all candidates in the validation sample. In the second exercise, I show that estimated rally effects are robust after relaxing multiple model and data assumptions. In the third exercise, I execute model selection tests to infer if candidates are strategic and forward-looking. We fail to reject these two assumptions made about candidates in this paper.

Next, I execute two counterfactual experiments. The first counterfactual experiment focuses on the cumulative effect of rallies on electoral outcomes. These differ from contemporaneous effects on popularity due to decay. Furthermore, since candidates hold multiple rallies in the same location, there is also cumulation. These two forces are in tension, and the effect on election results is ambiguous. To isolate the effect of a candidate's *total rallies*, I compare electoral outcomes under (i) "None Rally" with (ii)"Only one candidate rallies". In this exercise, I find that Trump's rallies increased his chances of winning by 40%. However, other candidates did not increase their chances of winning significantly.

There is little evidence that shows political rallies are an informative form of political communication. For instance, Snyder and Yousaf (2020) found that Trump and Clinton's rallies did not change issue salience and issue preference for the most critical issues in 2016.[12] The authors also did not find any

[9] Percentage points of votes. Poll margins are constructed from FiveThirtyEight, which aggregates polls from various pollsters. Their objective is to predict vote shares; therefore, these numbers directly translate into vote shares.

[10] A candidate can be physically present at one location at a given time, while ads can be run at multiple locations simultaneously.

[11] I find that in 2016 and 2012, rallies changed the decision of $576.6K$ and $388K$ voters, while T.V. ads changed the decision of $2.1M$ voters in 2012.

[12] Test of rallies on these outcomes provided only for 2016.

change in voters' perception of candidates' valence post rally in 2012.[13] The nature of rallies themselves is concerning. A rally, unlike presidential debates, provides the candidate with an uncontested platform where leaders can make factually inaccurate claims. For instance, Former President Donald Trump made 131 factually incorrect remarks at a rally in Wisconsin (New York Times, 2020).[14]

Motivated by these empirical findings and anecdotes, I explore whether government intervention can regulate rallies. For this purpose, I execute a counterfactual experiment that uses campaign silence as a government policy. Campaign silence, also known as election silence, is an intervention that bans political campaigning for a given number of days and is generally imposed right before the election. Campaign silence policies vary in length across countries. Some countries, such as France, impose a campaign silence that lasts one day (Pickles, 1960), while countries such as Cyprus, Indonesia, and Brazil impose campaign silences that last two or more days (Knews, 2022; IFES, 2012; Globo, 2020). It is unclear what length of campaign silence effectively reduces the influence of campaigning on election results. I provide the minimal effectual campaign silence length that can change electoral outcomes. For the 2012 election, campaign silence of any duration has no effect. In 2016, when Trump's rallies were pivotal, campaign silences would be effective if they lasted more than four days before the election. Any shorter duration, does not substantially dissipate the effects of Trump's rallies.

This paper contributes to different strands of literature stretching across several disciplines. Firstly, the model contributes to the literature on political campaigning (Kawai and Sunada, 2022; Erikson and Palfrey, 2000; de Roos and Sarafidis, 2018; Meirowitz, 2008; Polborn and Yi, 2006; Garcia-Jimeno and Yildirim, 2017; Gul and Pesendorfer, 2012; Strömberg, 2008) by constructing a dynamic framework where candidates choose to rally. Strömberg (2008) studies campaign state visits and builds a model where candidates allocate time across states, but his model is static, has identical strategies, and does not incorporate decay. I provide a dynamic model with candidate-specific strategies where campaign effects decay. Acharya et al. (2022) study political spending within an election and identify the perceived decay rate associated with campaigning. The authors characterize optimal spending ratios rather than candidate-specific spending strategies due to multiple equilibria, while I use a perfect information structure that allows one to study candidate-specific strategies. Kawai and Sunada (2022) study spending across elections, whereas I study rallying decisions within an election. Garcia-Jimeno and Yildirim (2017) study strategic interaction between candidates in bipartisan races and media in the context of U.S. Senate races. In their paper, there is only one location where candidates can campaign. I analyze presidential races and allow for campaigning across multiple locations.

This paper also contributes to the literature on the effectiveness of political campaigning events (Wood, 2016; Shaw, 1999; Shaw and Roberts, 2000; Shaw and Gimpel, 2012). I contribute to this literature by estimating effects of rallies on poll margins and also electoral outcomes. The literature finds mixed evidence on effectiveness of rallies and related events on polls, vote shares, and other outcomes of interest. Moreover, these estimates also vary with the identification strategy used by the authors. In the past, authors have ignored the heterogeneity of effectiveness across candidates and attempted to

[13] Test of rallies on perceived candidate valence is provided only for 2012.

[14] Also see (Politico, 2020) and (New York Times, 2021).

provide an average estimate. Recently, Snyder and Yousaf (2020) studied political rallies and showed that Trump significantly affected intention to vote, while other non-populist candidates did not. Where authors in this study used difference-in-difference specification at the media market level to address the selection bias,[15] I directly address the selection by modeling these decisions.

I also contribute to the literature on empirical dynamic games. Traditionally, empirical dynamic games use a discrete choice set up to study firm entry/exit decisions (Aguirregabiria and Mira, 2007; Arcidiacono et al., 2016). The estimation and inference procedures here rely on observing many games. I contribute by constructing a framework where inference can be made by observing one game. For this purpose, I exploit the use of stage games as a unit of observation and the Markovian dependence that every consecutive stage game possesses to estimate and infer parameters.

Lastly, this paper touches the literature on analyzing and estimating stochastic goodwill models in marketing and operations research (Kwon and Zhang, 2015; Grosset and Viscolani, 2004; Marinelli, 2007; Doganoglu and Klapper, 2006; Chintagunta and Vilcassim, 1992) by extending the stochastic goodwill framework to a dynamic discrete game framework. Traditionally these models have studied dynamic advertising for firms that wish to maintain/increase their goodwill among their consumers. I provide a model where the advertising level can only be 0 or 1 in a given market. Moreover, there is a period-specific capacity constraint on the level of advertising where a marketeer can only advertise in one market at a given time. The contribution here is the capability of the model to provide predictions on advertising across multiple markets at once, that may or may not be horizontally differentiated. Moreover, the time horizon is finite, and the model appeals to situations where the marketer faces a specified product launch deadline.

The paper proceeds as follows, Section 2, discusses the model, equilibrium, and comparative statics. Section 3 discusses data sources, summary statistics, and the three empirical patterns. Section 4, discusses parameterization and estimation procedure. Section 5, discusses the estimates, persuasion rates, in-sample model fit, and out-of-sample model fit. Section 6 discusses robustness tests. Section 7, discusses the model selection tests. Section 8, discusses counterfactual experiments. Finally, Section 9 concludes.

2 Model

The model analyzes the interaction between candidate rally choices and their popularity level. In this model, we have K states, $T+1$ periods, and two candidates, $\{R, D\}$. I assume one popularity measure per state that holds information on the relative popularity of candidates. This popularity measure can take values in $\mathbb{R}$. Here popularity is interpreted as R's poll margin lead over D. Naturally, if popularity is positive, then R is leading in the polls in the state. If negative, then D is leading the polls. Popularity follows an AR(1) process.

The game is played over T periods, and a sequential move stage game is played in each period. In this

[15]The assumption of SUTVA is more challenging to justify at this geographic level due to spontaneous news coverage of rallies in geographically closer media markets.

stage game, the order of play among the candidates is random. A candidate, at their turn, must choose at most one state out of K states. The candidate can choose not to rally as well. If the candidate chooses to rally, then they pay a cost.

The decision-making stops at period T, and in the election period $T+1$, every state chooses the popular candidate as the winner. The winning candidate in a state gains a payoff proportional to the number of electoral college votes associated with the state. Their total payoff is the sum of payoffs received from each state.

2.1 Preliminary

I denote the set of states and the number of states by $K = \{1, 2, \ldots, K\}$. An arbitrary candidate will be denoted by $i \in \{R, D\}$. Arbitrary period is denoted by $t \in \{1, 2, \ldots, T\}$. Recall that decisions are made in periods $1, 2, \ldots, T$. Each state has a popularity measure p_{kt} that denotes the relative popularity level of R and D in state k. If the game ends with a candidate being popular in-state k, then they win all electoral college votes.

Candidates can allocate a unit of perishable indivisible goods to at most one state at a given time. The indivisible good is a political rally. This good can not be saved and has a constant candidate-specific cost c_i for candidate i.

Let a_{ikt} be a binary variable indicating if candidate i chose to rally in state k at period t or not. Recall that the following holds $\sum_{k=1}^{K} a_{ikt} \leq 1$, i.e., candidate i can hold at most one rally in a given period. Period $t+1$ popularity, $p_{k,t+1}$, is given by the following AR(1) process:

$$p_{k,t+1} = \alpha_R a_{Rk,t} + \alpha_D a_{Dk,t} + \rho p_{kt} + \delta_k + \nu_{k,t+1} \tag{2.1}$$

Here, α_i is defined as candidate i's effectiveness in influencing popularity, where $i \in \{R, D\}$. The parameter ρ is the persistence in popularity and is one of the key parameters of the game that determines the interplay of candidate choices and the popularity level. $\nu_{k,t}$ is a random variable indicating a generic popularity shock. I make following assumption on popularity shocks across all states, $(\nu_{1t}, \nu_{2t}, \ldots, \nu_{Kt})$.

Assumption 2.1 (Popularity Shocks) *The popularity shocks* $(\nu_{1,t}, \nu_{2,t}, \ldots, \nu_{K,t})$ *are distributed according to a multivariate normal distribution.*

$$(\nu_{1,t}, \nu_{2,t}, \ldots, \nu_{K,t}) \sim N(0, \sigma_\nu^2 I_K) \tag{2.2}$$

Note that $\sigma_\nu^2 I_K$ *is a positive definite matrix.*

The term σ_ν is referred to as the volatility in popularity. For the baseline model, I assume that popularity shocks are normally distributed and are uncorrelated across states. This assumption is relaxed in Section 6, where two distinct types of correlations between the states are allowed.[16]

[16]See Assumption A.1 for a more generalized popularity shock distribution that can be supported.

Let the density of popularity in period $t+1$ given period t primitives be denoted by $f(p_{t+1}|a_{R,t},a_{D,t},p_t)$. Here $p_s = (p_{1,s}, p_{2,s}, \ldots, p_{K,s})$ and $a_{it} = (a_{i,1t}, a_{i,2t}, \ldots, a_{i,Kt})$ for $i \in \{R, D\}$. Let $\delta = (\delta_1, \delta_2, \ldots, \delta_K)$ then by assumption 2.1 this density is given by:

$$f(p_{t+1}|a_{R,t},a_{D,t},p_t) = \frac{1}{\sigma_\nu^K}\prod_{k=1}^{K}\phi\left(\frac{p_{k,t+1} - \alpha_R a_{Rk,t} - \alpha_D a_{Dk,t} - \rho p_{k,t} - \delta_k}{\sigma_\nu}\right) \tag{2.3}$$

Where $\phi(.)$ denotes the p.d.f. of the standard normal distribution. This popularity evolution equation can also be statistically founded by considering a mean reverting process similar to Acharya et al. (2022).[17]

Every state k has a payoff that is proportional to the number of electoral college votes, e_k, the state has. In period $T+1$ if the game terminates with $p_{k,T+1} > 0$, candidate R receives $e_k E$, where E denotes the maximal payoff a candidate can receive. Candidate R's total payoff will be aggregate of payoffs received from each state and it is given by:

$$\begin{aligned}
V_{R,T+1}(p_{T+1}) &= \sum_{k=1}^{K} e_k E \times \mathbb{1}\left\{p_{k,T+1} > 0\right\} \\
\Rightarrow \mathbb{E}_{p_{T+1}}\left[V_{R,T+1}(p_{T+1})\middle|p_T, a_{RKT}, a_{DKT}\right] &= \sum_{k=1}^{K} e_k E \times \mathbb{P}\left[p_{k,T+1} > 0\middle|p_T, a_{RKT}, a_{DKT}\right]
\end{aligned} \tag{2.4}$$

Where $\mathbb{E}_{p_{T+1}}$ is an expectation operator, which takes expectation with respect to p_{T+1}. The symbol $\mathbb{P}$ denotes the probability. For the baseline case $\mathbb{P}\left[p_{k,T+1} > 0\middle|p_T, a_{RKT}, a_{DKT}\right] = \Phi\left(\frac{\alpha_R a_{Rk,t} + \alpha_D a_{Dk,t} + \rho p_{k,t} + \delta_k}{\sigma_\nu}\right)$, where $\Phi(.)$ denotes the standard normal c.d.f.

For candidate D, their electoral payoff is defined as $V_{D,T+1} = E - V_{R,T+1}$. Therefore, if R looses in a state then D simultaneously wins there. Moreover, by using these payoffs, one can see that D's popularity can be denoted by $-p_{k,t}$. Therefore, if R is popular (i.e. $p_{k,T+1} > 0$) then D is unpopular and vice-e-versa.

2.2 Timing of Decisions and Information

Game begins at period $t = 1$ and ends at $t = T+1$, where decisions are only made in periods $1, 2, \ldots, T$. A stage game, as described in Figure 1, is played in each period $t \in \{1, 2, \ldots, T\}$. The timing of decisions and revelation of information for the stage game played at period t (refer to Figure 1) are described in sub-periods $\tau_1, \tau_2, \ldots, \tau_6$. In sub-period τ_1 popularity vector p_t is observed. In sub-period τ_2, nature

[17] Authors also micro-founded the mean reversion process by considering a set of impressionable voters (Andonie and Diermeier, 2019) who vote on the basis of good will. In Marketing and Operations Research, Kwon and Zhang (2015) consider market shares directly and model them as a general Brownian motion. If we have $K = 1$ then this is a special case of that where we impose further restrictions to obtain a mean reverting nature in the process. Moreover, for this setting market share is analogous to vote share or polls. I reduce the state variable by considering poll margin lead, and call it popularity. This is a deviation from Doganoglu and Klapper (2006), who model the consumer behavior directly. A consumer's utility in their set up depend on a brand's goodwill in addition to brand-specific characteristics. This set up can be extended to consider general votes however presence of multiple states can make the model computationally infeasible.

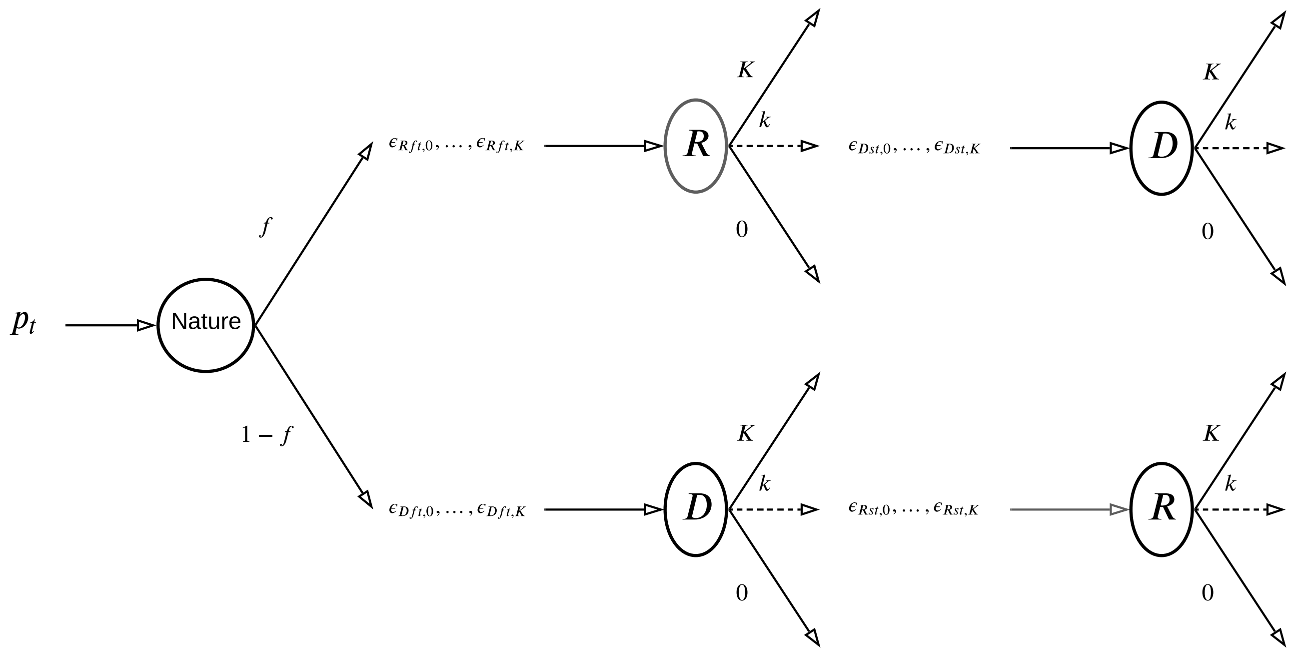


Figure 1: Stage Game

The stage game for each period $t = 1, 2, \ldots, T$ is provided in this figure. In each period, both candidates observe the popularity level p_t. Then nature chooses a first mover and a second mover. Then cost shocks for the first mover are drawn and are observed by both candidates. The first mover decides where to rally. Then cost shocks for the second mover are drawn and observed by both candidates. Second mover makes their decision and then the game proceeds to period $t+1$.

chooses a first mover. In sub-period τ_3, the first mover draws their cost shocks, and in sub-period τ_4 they make their decision. In sub-period τ_5 second mover cost shocks are drawn, and in sub-period τ_6 second mover will make their decision. After this sub-period, the game proceeds to period $t+1$. I give more details on each sub-period below.

τ_1 At sub-period τ_1 the popularity level vector $p_t = \left(p_{1t}, p_{2t}, \ldots, p_{Kt}\right)$ is realized and observed by the candidates. Here p_{kt} indicates the current popularity level of the candidates.

τ_2 Nature makes a draw to choose a first mover and a second mover for the stage game. The probability i is chosen as the first mover is denoted by f_i. Let $f_R = f$ and $f_D = 1 - f$.

τ_3 The first mover i's cost shocks, $\epsilon_{ift} = \left(\epsilon_{if,t,0}, \epsilon_{if,t,1}, \ldots, \epsilon_{if,t,K}\right)$[18], are realized in this sub-period. Here $\epsilon_{if,t,0}$ is the cost shock for not rallying, while $\epsilon_{if,t,k}$ is the cost shock for rallying in state k. Moreover, they are realized immediately before a decision is made.

τ_4 The first mover, i, in period t solves the following Bellman equation after observing the current popularity p_t and cost shocks $\epsilon_{ift} = (\epsilon_{if,t,0}, \ldots, \epsilon_{if,t,K})$:

[18] These shocks are part of the random utility specification for the candidates. For more details please refer to McFadden (1973) and McFadden (1978). The interpretation of these cost shocks is unforeseen events that may increase or decrease the difficulty of rallying in a state. For instance, Hurricane Sandy made campaigning on the Atlantic seaboard very difficult in the 2012 presidential election.

$$V_{ift}(p_t, \epsilon_{if,t}) = \max_{k\in\{0,1,\ldots,K\}} \Bigg\{ - c_i \times \mathbb{1}\{k \neq 0\} - \epsilon_{if,t,k} + \beta \sum_{l=0}^{K} \mathbb{E}_p \Big[V_{i,t+1}(p) \Big| a_{it} = k, a_{jt} = l, p_t \Big] \times \sigma_{js,t}\Big(l; k, p_t\Big) \Bigg\} \tag{2.5}$$

In order to choose option k, i must pay the cost c_i if they choose to rally, and the random cost shock $\epsilon_{if,t,k}$. These two components form the flow costs for candidate i. Moreover, while making this decision, i also considers the continuation value associated with each option. This continuation value consists of a nested conditional expectation of i's value in the next period. The inner expectation is taken with respect to popularity in the next period given that action, $a_{it} = k, a_{jt} = l$ and current period popularity p_t. The outer expectation is with respect to j's actions given i chose k and the current popularity p_t. The probability of j choosing an action l is denoted by $\sigma_{js,t}\,(l; k, p_t)$ and it is an equilibrium object. Let $a_{ift}(p_t, \epsilon_{if,t})$ be the associated policy function with $V_{ift}(p_t, \epsilon_{if,t})$.

τ_5 The second mover, denoted by s, draws cost shocks, $\epsilon_{is,t} = \Big(\epsilon_{is,t,0}, \epsilon_{is,t,1}, \ldots, \epsilon_{is,t,K}\Big)$.

τ_6 The second mover i's decisions are made in this period. In addition to observing the popularity level and their cost shock, the second mover also observes the decision made by the first mover. Therefore, $a_{jf,t}$ is also a state variable for the first mover. The second mover solves the following interim Bellman equation after observing p_t, first mover action $a_{jft} = l$ and cost shocks ϵ_{ist} :

$$V_{ist}(l, p_t, \epsilon_{is,t}) = \max_{k\in\{0,1,\ldots,K\}} \Bigg\{ - c_i \times \mathbb{1}\{k \neq 0\} - \epsilon_{is,t,k} + \beta \mathbb{E}_p \Big[V_{i,t+1}(p) \Big| a_{it} = k, a_{jt} = l, p_t \Big] \Bigg\} \tag{2.6}$$

Here c_i and $\epsilon_{is,t,k}$ are the flow-costs from choosing option k. The continuation value is expectation of i value in the next period given a_{it}, a_{jt}, p_t. Let a_{ist} be the associated policy function.

Here I define the value function at popularity vector p prior to the order of play in period t:

$$V_{i,t}(p_t) = f_i \times \mathbb{E}_{\epsilon_{if,t}}\Big(V_{ift}(p_t, \epsilon_{ift})\Big) + (1 - f_i) \sum_{k=0}^{K} \Big[\sigma_{jft}(k; p_t) \times \mathbb{E}_{\epsilon_{is,t}}\Big(V_{ist}(k, p_t, \epsilon_{ist})\Big) \Big] \tag{2.7}$$

The above Bellman equation uses the expected value i receives as the first mover and as the second mover, given p_t, to calculate the value of entering a period when popularity is p_t. With probability f_i, i is chosen as the first mover and the term $\mathbb{E}_{\epsilon_{if,t}}\Big(V_{ift}(p_t, \epsilon_{ift})\Big)$ is the value of i becoming the first mover prior to the realizing of cost-shocks. The operator $\mathbb{E}_{\epsilon_{if,t}}$ is an expectation operator, which calculates the expectation of $V_{ift}(p_t, \epsilon_{ift})$ with respect to the cost shocks ϵ_{ift}. The second term is the expected payoff of being the second mover. When i is the second mover, there are total $K + 1$ possibilities that can take place, and each possibility corresponds to the decision made by j as the first mover. The expectation is with respect to the conditional choice probability (CCPs from here on) with which j rallies in state k. Once an action k is chosen the expected payoff of being the second mover when action k was chosen by first mover is given by the term $\mathbb{E}_{\epsilon_{is,t}}\Big(V_{ist}(k, p_t, \epsilon_{ist})\Big)$.

2.3 Equilibrium

In this game, only one candidate decides at a time. Moreover, past actions, popularity levels, and cost shocks are observed at the time of decision making. Candidates are forward looking and have a rational expectation of future payoffs and actions. Therefore, the game at hand is a perfect information game, and it can be solved using backward induction.

Below I state the first assumption on cost shocks

Assumption 2.2 (Independent Cost Shocks) *The cost shocks are independent across all information nodes and actions. That is the following holds:*

$$\epsilon_{im,t,k}|p_t \perp \epsilon_{i'm',t',k'}|p'_{t'} \quad \forall \ (i,m,t,k,p_t) \neq (i',m',t',k',p'_{t'}) \tag{2.8}$$

In the previous subsection, the Bellman equations assumed that current cost shocks are payoff relevant, but past cost shocks were not. The above assumption, along with the definition of the popularity process and the timing of the game, implies that the Bellman equations depend on current popularity shocks, current cost shocks, and first mover decisions (in the case of the second mover). The same holds for the associated policy functions or, in this case, the best responses of candidates.

I also assume that cost shocks are drawn from Type-1 Extreme Value distribution. The assumption is stated below:

Assumption 2.3 (Distribution of Cost Shocks) *Cost shocks are drawn from Type-1 Extreme Value distribution:*

$$\epsilon_{im,t,k}|p_t \sim T1EV \quad \forall \ (i,m,t,k,p_t) \tag{2.9}$$

Assumptions 2.3 and 2.2 ensure that the Subgame Perfect Equilibrium will exists and it will be unique—that is multiplicity will exist with probability zero.[19]

This provides us with unique candidate specific campaigning strategies in terms of conditional choice probability (CCP, hereafter).[20] This is one of the key contributions of this paper. Prior to this, researchers working in structural electoral games have focused on static models, Strömberg (2008) and Gordon and Hartmann (2016). The literature has also considered dynamic games that allow for a unique normalization of candidate equilibrium strategies, Acharya et al. (2022). Here I provide a dynamic game of

[19]Since only one candidate makes a decision at a given time we do not need to search for any fixed point in order to find an equilibrium. Therefore, the only way multiple equilibria would exist is when a candidate is indifferent between two actions. However, under these assumptions the convolutions formed by adding ϵ_{imtk} to $-\epsilon_{imtl}$ for each i, m, t ,k, l is a continuous random variable. Therefore these indifference equalities occurs with probability zero. Note that in order to support this we do not truly need that the distribution is extreme value and cost shocks are always independent. In fact as long as such convolutions are continuous random variables this form of uniqueness will take place. In section A.1, I provide a more general set up that can incorporate correlated cost shocks or dependence on past actions.

[20]Another advantage of using type-1 extreme value distribution is to have lower computation burden while evaluating CCPs. Moreover, the use of this distribution is very common in literature that relies on using Discrete Choice models, Dynamic Discrete Games and etc.

electoral competition with a finite time horizon that can support unique equilibrium strategies in probability space.[21]

Under assumptions 2.2 and 2.3, we can show that best responses, in probability space, are functions[22] of current popularity, cost shocks, and, in the case of the second mover, first mover action. Now, I characterize the equilibrium conditional choice probabilities. The Proposition 2.1 lays out this characterization. If one knows electoral payoffs then it is possible to evaluate all period specific value functions and CCPs of both candidates at each possible level of popularity. Eq. 2.4 defines electoral payoffs and by using this equation, Proposition 2.1 uses backward induction to characterize period-specific value functions and CCPs.

Proposition 2.1 (Characterization of Value Functions and CCPs) *Given eq. 2.4, which defines electoral payoff, Assumptions 2.2 and 2.3 the following holds for all* $t = 1, 2, ..., T$

- *The value function* $V_{i,t}$ *takes the following functional form:*

$$V_{i,t}(p_t) = f_i \times \ln\left(\sum_{k=0}^{K} \exp\left\{u_{if,t}(k;p_t)\right\}\right) + (1-f_i) \times \sum_{k=0}^{K}\left[\sigma_{jf,t}(k;p_t)\ln\left(\sum_{l=0}^{K}\exp\left\{u_{is,t}(l;k,p_t))\right\}\right)\right] \tag{2.10}$$

- *The expected probability of* i *choosing action* k *as the first mover is given by:*

$$\sigma_{if,t}(k;p_t) = P\Big(k = a^*_{if,t}(p_t,\epsilon_{if,t})\Big) = \frac{\exp\Big(u_{if,t}(k;p_t) - u_{if,t}(0;p_t)\Big)}{1+\sum_{l=1}^{K}\exp\Big(u_{if,t}(l;p_t) - u_{if,t}(0;p_t)\Big)} \tag{2.11}$$

- *The probability of* i *choosing action* k *as the second mover is given by:*

$$\sigma_{is,t}(k;l,p_t) = P\Big(k = a^*_{is,t}(a_{jft} = l, p_t, \epsilon_{is,t})\Big) = \frac{\exp\Big(u_{is,t}(k;l,p_t) - u_{is,t}(0;l,p_t)\Big)}{1+\sum_{q=1}^{K}\exp\Big(u_{is,t}(q;l,p_t) - u_{is,t}(0;l,p_t)\Big)} \tag{2.12}$$

  Where, the option specific value function, $u_{if,t}(k;p_t)$, *for* i *when they are the first mover at period* t *at popularity level* p_t *satisfies the following:*

$$u_{if,t}(k;p_t) = \sum_{l=0}^{K} u_{is,t}(k;l,p_t) \times \sigma_{js,t}(l;k,p_t) \tag{2.13}$$

  The option specific value function, $u_{is,t}(k;p_t)$, *for* i *when they are the second mover at period* t *at popularity level* p_t *and the first mover chose* l *satisfies the following eq. 2.14:*

$$u_{is,t}(k;l,p_t) = -c_i \times \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[V_{i,t+1}(p)\Big|\, a_{it} = k, a_{jt} = l, p' = p_t\right] \tag{2.14}$$

[21] Similar to Milgrom and Weber (1985) and Aguirregabiria and Mira (2007), we can define the Subgame Perfect Nash Equilibrium in probability space or distributional strategies. Authors focused on Markov strategies. In this paper, due the assumptions A.1(or 2.1) and 2.2 the equilibrium strategies are also Markov. Therefore the Subgame Perfect Nash Equilibrium and the Markov Perfect Equilibrium are the same in this game.

[22] These are functions and not best response correspondences because multiplicity is of probability zero and only the aforementioned terms are part of relevant state variables.

The proof for Proposition 2.1 is given in Section A.2. The proof involves an application of the generalized model described and solved in Section A.1. The statement regarding observed choices have been omitted here for brevity but nonetheless can easily be accommodated. The idea of the proof is quite simple. Since the game is of perfect information one can show that the proposition holds for $t = T$. First I show that the optimal actions given the cost shocks are unique for the second mover with probability 1. Then I characterize its conditional choice probabilities and value function for the second mover. Given this I move to the first mover and then repeat these steps. Then in the induction step I assume it holds for a given t along with the characterization of conditional choice probabilities and show that it has to hold for $t-1$ by using the same steps as for period T.

The proposition recursively characterizes the equilibrium choice probabilities and value functions for a candidate i. Within any stage game, the second mover value functions ($u_{ist}(k; l, .)$ for all $k, l = 0, 1, \ldots, K$) directly depend on expectation of the next period value function. The first mover option value functions ($u_{ift}(k; .)$ for all $k = 0, 1, \ldots, K$) also depend on expectation of the next period value function, as their values are given by second mover value functions. These first and second mover value functions solely determine rallying probabilities in the model.

Given the recursive nature of equations in proposition 2.1, obtaining predictions over chosen actions is not trivial. These equations do not have a reduced form due to the presence of multinomial logistic choice probabilities. The logistic functional forms period by period introduce an additional layer of complex relation between endogenous variables and the parameters that are hard to study using analytic methods. Therefore, I have to rely on simulations for analyzing equilibrium behavior. In order to simulate the choice probabilities, I approximate the value functions as the state space is a continuum here. I follow Judd et al. (2014) for constructing a sparse a grid and the accompanying Chebyshev polynomial to approximate the value functions, which include the period level value function and also the option specific value functions for each mover and player.[23]

2.4 Equilibrium Behavior

Studying and analyzing the first and second mover probabilities individually is a cumbersome task analytically and computationally. A total of $K + 1 + (K + 1)^2$ choice probabilities predict how a candidate would rally given the appropriate information about state variables. We reduce the number of choice probabilities by considering the probability of choosing k prior to the order of play. To derive these probabilities first consider the probability of observing $a_{Rt} = k$ and $a_{Dt} = l$ given p_t, i.e. probability of

[23]For computing the next period expectation I follow Judd et al. (2017) and compute integration only once. To calculate these conditional expectations I use Konrad-Patterson quadrature (it can be obtained from `http://www.sparse-grids.de/`). This is an accompanying website for Heiss and Winschel (2008). Appendix B provides every detail about the algorithm that I construct and use for this purpose. I also assess and present the performance of numerical approximation by checking residuals of equations in proposition 2.1 on empirically relevant[24] non-collocation points (Judd, 1992). The mean of the distribution errors of approximation are in the order of 10^{-5}. The maximal error of approximation is obtained in the order of 10^{-2} with a very thin right tail (R's first mover probability in the Figure 10).

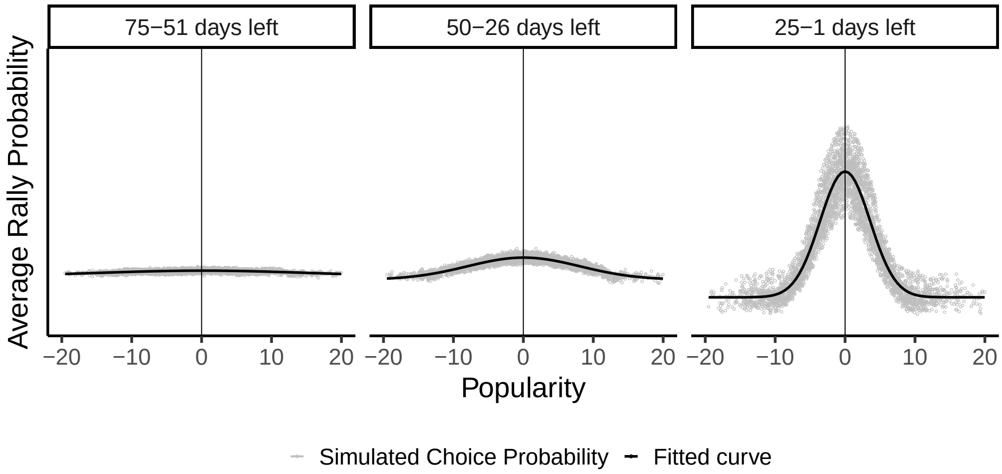


Figure 2: *The figure plots the dynamic relationship a candidate's probability of rallying in a state ($\sigma_{it}(k, p_t)$ defined in equation 2.16) and their popularity in that state (p_t). The figure shows that candidates choose to rally more if competition is neck and neck as the election approaches. This relationship holds irrespective of how many states we consider in the game. Here I show the pattern holds for four states with no rally option case. The pattern also holds for one state and no rally option. Refer to Figure 12.*

observing action profile k, l given current popularity across states.

$$\sigma_t(k,l;p_t) = f \cdot \sigma_{Rft}(k;p_t) \cdot \sigma_{Dst}(l;k,p_t) + (1-f) \cdot \sigma_{Dft}(l;p_t) \cdot \sigma_{Rst}(k;l,p_t) \tag{2.15}$$

We can obtain choice probabilities prior to the order of play by considering the marginal distributions of $\sigma_t(k,l;p_t)$. Here R's choice probabilities are given by the marginal distribution corresponding to k, and for D it is given by the marginal distribution of l.[25] The following equation defines these choice probabilities.

$$\begin{aligned} \sigma_{Rt}(k;p_t) &= \sum_{l=0}^{K} \sigma_t(k,l;p_t) \\ \sigma_{Dt}(l;p_t) &= \sum_{k=0}^{K} \sigma_t(k,l;p_t) \end{aligned} \tag{2.16}$$

Three remarks are necessary regarding how $\sigma_{it}(k; p_{1t}, \ldots, p_{kt}, \ldots, p_{Kt})$ relates with p_{kt} (refer to Figure 2 for an illustration).[26] Firstly, the choice probabilities have a weak relationship with popularity in the

[25] Note, the choices are correlated and therefore the probability of an action profile is not the product of its marginals.

[26] This involves calculating simulated version of conditional choice probabilities, as defined in equation 2.16, across 1,000 possible popularity values in each period. For each draw of popularity I draw 1,000 cost shocks from extreme value distribution and pick the best option for each candidate given their respective state variables values. Then I divide the set of periods into three bins and calculate average probability of rally for each popularity draw: (i) beginning of the election phase consisting of periods 1-100 (or 75-51 days before election —periods are not the same as days as in the data candidates can visit multiple days); (ii) middle of the election phase consisting of period 101-200 (or 50-26 days before the election); (iii) end of election phase 201-300 (or 25-1 days before the election). The results of this exercise are depicted in the Figure 2.

respective states in earlier periods. The explanation behind this hinges on two channels. The first channel is that of discounting. At any given period the corresponding value functions satisfy the following inequality:

$$\frac{\left(1-\beta^{T-t+1}\right)(\ln(1+Ke^{-c_i})+\gamma)}{1-\beta} \leq V_{it+1}(p_t) \leq \frac{\left(1-\beta^{T-t+1}\right)(\ln(1+Ke^{-c_i})+\gamma)}{1-\beta} + \beta^{T-t-1}\sum_{k=1}^{K} e_k E \qquad (2.17)$$

The derivation of this inequality is provided in Section A.3. For earlier periods (i.e. small t), the upper bound is closer to the lower one, therefore making V_{it+1} independent of electoral payoffs. This implies that the probability of rallying depends only on flow value, which is the cost of rallying, and not on the continuation value. Note that the flow payoff is independent of current popularity in the model. Therefore corresponding choice probabilities also reflect this nature.

The second channel is decay. In the initial periods, rallies will have negligible effects on election day popularity because these effects decay exponentially with time. If a candidate rallies in a period t, then its *direct* effect on p_{t+1+j} will be $\rho^j \alpha_R$. Effect on election day popularity will be $\rho^{T-t}\alpha_R$.[27] This term is smaller for small t and higher for larger t. Therefore in the initial periods, return from a rally is small.

A second remark highlights what happens when we are closer to elections. There is a higher probability of a rally in states where a candidate's popularity is close to zero. This is true because of the nature of electoral payoffs at the state level. Recall that a candidate wins all electoral payoff from a state if they are popular in that state. This ensures that the change in expected payoff due to a rally is maximal when popularity is closer to zero. Areas with popularity far away from the cut-off are more challenging to swing either way than areas closer to the cut-off. Therefore in equilibrium, candidates prioritize areas that are easier to swing and allocate higher amounts of rallies in areas where popularity is closer to the cut-off.

The third remark focuses on the transition shown in Figure 2. As elections approach, candidates prioritize areas where competition is tight over where it is not. It is explained by the conjunction of remarks 1 and 2. As elections approach, the decay and discounting channel weakens, making rallying a more profitable option. From remark 2, recall that areas where competition is tight, are easier to swing than areas where popularity is lopsided. Therefore, we see a higher gradual increase in rallies where competition is neck and neck than in areas where it is not.

2.5 Comparative Statics

In this section, I will discuss four comparative statics concerning rally effectiveness, cost of rallying, persistence, and volatility in popularity.

Increase in persistence in popularity, $\rho \uparrow$: One direct implication of an increase in persistence in popularity is a higher observed auto-correlation. However, persistence also has an interesting implication for rally choices. With an increase in this parameter, earlier rallies have more lasting effects on future

[27] There is also an indirect effect as rally today will change tomorrow's popularity, and hence it will change the behavior of the candidate and their opponent.

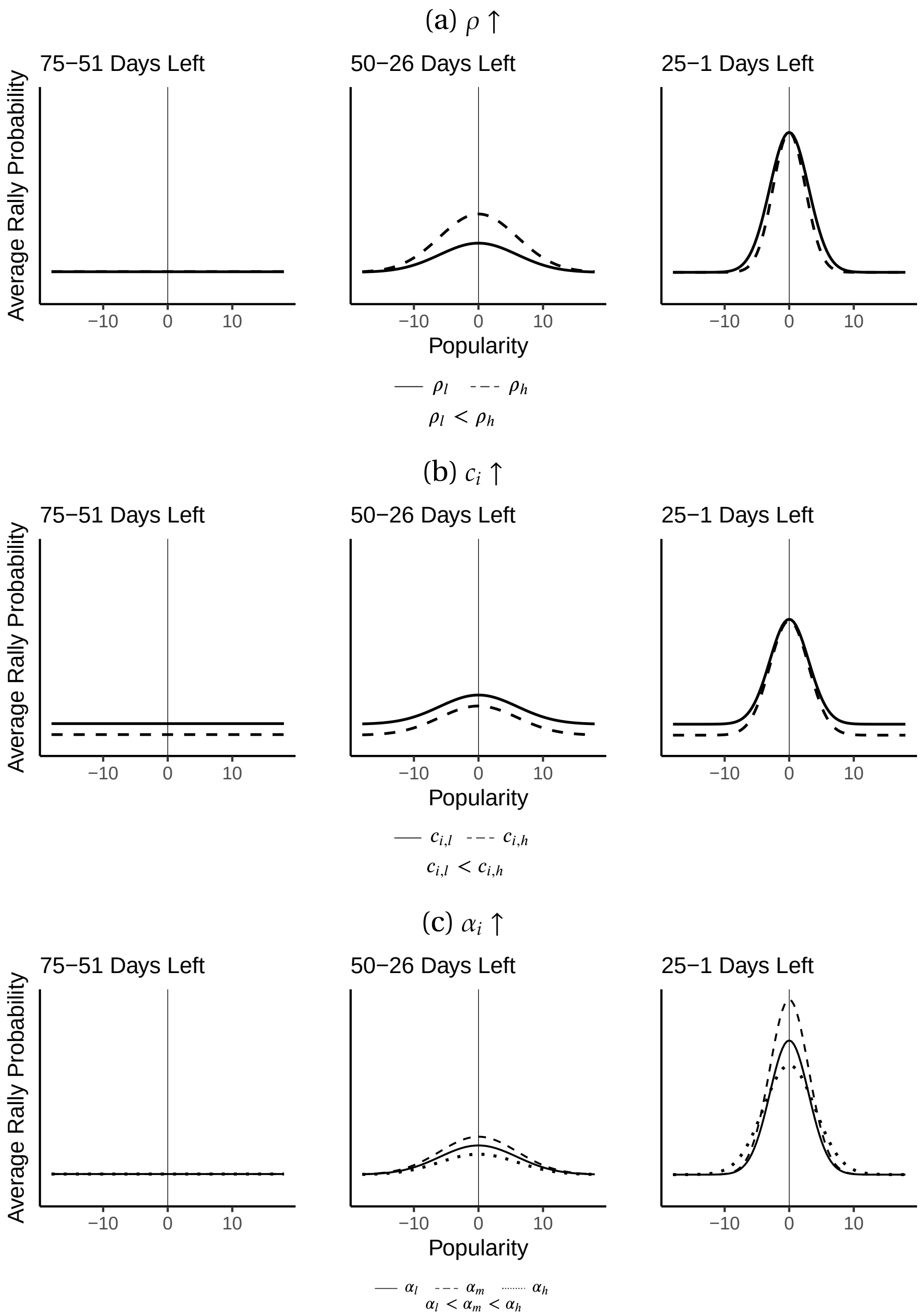


Figure 3: *This figure illustrates how probability of rally changes when the key parameters ρ, α_i and c_i are increased. Panel 3a illustrates the comparative statics with respect to ρ. Panel 3b illustrates the comparative statics with respect to c_i. Panel 3c illustrates the comparative statics with respect to α_i.*

popularity values and, therefore, higher chances to swing election day results. This can be seen by considering the *direct* effect period t rally has on election day popularity. Recall this is given by $\rho^{T-t}\alpha_i$, which increases with ρ. The higher ρ is, the higher would be the magnitude of this effect. This increases the returns to rallying in period t and therefore incentivizes rallying in earlier periods. As a result, the probability of rallying in earlier periods will increase weakly with ρ.[28] The comparative statics is demonstrated

[28]Another implication would be lower rallying in later periods, which is driven by the fact that past popularity has a higher weight and can counter-act the effect of rallying.

in the Figure 3a.

Increase in cost of rallying, c_i ↑: An increase in the cost of a rally parameter decreases the probability of rallying as it becomes a costlier option. One interesting feature is that the decrease in the probability of rallying is similar to a level shift, particularly at the beginning of the election. Recall that returns from rallying are negligible in the beginning due to the decay and intertemporal discounting channels. As a result, the level of rallying is solely determined by the cost of rallying. To see this, consider proposition A.4. This proposition shows that the benefit from rallying in the early phase of the election is close to zero. As a result, the probability of rallying only depends on its cost, which stays constant throughout the election. Therefore, a change in cost induces a parallel shift in the probability of rallying. The comparative static is demonstrated in the Figure 3b.

Increase in rally effectiveness, α_i ↑: The probability of rallying has a non-monotonic relation with effectiveness. With moderate increases, the probability of a rally will naturally increase with effectiveness. However, if there is a large increase in the effectiveness of rallies, then a lower level of rallying can maintain a sufficient amount of popularity and save the cost of rallying later. This comparative static is demonstrated in the Figure 3c.

Increase in volatility in popularity, σ_v ↑: An increase in volatility in popularity will lead to more uncertainty in future popularity. This increase in uncertainty is detrimental to earlier campaigning, making rallying a less attractive option in the initial phases of the campaign. This leads to a decrease in rallying probability in each period t and in the popularity level p_t.

3 Data

3.1 Sources

This paper uses two primary data sources. The first one is *Democracy in Action*, which provides me with rally choices. The second source is FiveThirtyEight for obtaining state-specific poll margins (interpretation of popularity in the model).

Rally choices: The data on rally choices is obtained from *Democracy in Action*, a website created by Eric M. Appleman. In particular for 2012 and 2016 presidential elections I used Appleman (2012) and Appleman (2016) respectively. The website has information on the calendars of presidential candidates. For each day, the website provides activities candidates undertook in chronological order.[29] The website provides information not only on rallies but also on various other activities. I classify these activities into groups, one of which is political rallies. The calendar provides information on multiple phases of

[29]The events or activities that were performed at earlier times precede the ones which are undertaken later. This can be seen by considering entries for which such information is provided. The earlier events precede later events in these entries. For instance, consider Hilary R. Clinton's schedule on November 7, 2016, the entry '*Afternoon Get out to vote rally... Grand Rapids, MI.*' precedes the entry '*Final midnight "Get Out the Vote"... Raleigh, NC*'. Consider Donald J. Trump's schedule on October 13, 2016, the entry '*Mid-day rally... Palm Beach, FL*' precedes the entry '*Evening rally... Cincinnati, OH*'. The order is consistent for each day where times of the day have been provided.

the election. However, I use information from 100 days before the election. Therefore, everything on and after July 29 and July 31 for the 2012 and 2016 elections, respectively.

The group of activities I am interested in involves a candidate (i) holding a rally, (ii) giving a speech, or (iii) organizing a special event. I call these activities a political rally. In the model, candidates hold one rally in a period, while in the data, candidates can hold multiple rallies in a day. I define periods of the model as a quarter of a day and allocate these periods based on chronological information provided in "Democracy In Action". I also ignored rallies held in stronghold states and counted consecutive rallies in a state as one to ensure that there were at most four rallies in a day. In Appendix C, I provide details on data cleaning and allocation of periods to rally decisions.

Poll Margins: I use FiveThirtyEight's poll repository for obtaining aggregate poll margins at the state level. It is an organization that focuses on opinion poll analysis, economics, politics, and sports blogging. Since its creation, FiveThirtyEight has focused on producing reliable forecasts for presidential general elections, primaries, house elections, and gubernatorial elections. In 2016, the organization produced one of the most accurate forecasts for the presidential general elections.

As a poll aggregator, FiveThirtyEight[30] collects polls from multiple pollsters to generate reliable forecasts. It uses individual polls to produce polling averages after correcting for partisan biases that make individual polls unsuitable for a comprehensive study. Their forecasts are probabilistic, allowing for historical uncertainty that individuals' polls do not report. Their forecast model prioritizes state-level polls over national-level polls as the former are better predictors of results within the state. Moreover, forecasts in one state utilize information from states similar to itself for better predictability. Their polls are more conservative in early summer than closer to elections as individual poll fluctuations are more informative towards the end than in the beginning.

3.2 Summary Statistics

Summary statistics for R's poll margin lead across all states appear in the Figure 16. I have removed District of Columbia, NE-1, NE-2, NE-3, ME-1, ME-2 for brevity. I also show the aggregate number of activities in the raw data obtained after classifying the set of all activities. These aggregate numbers are provided in Table 9. The table also shows how many rallies were removed specifically after the cleaning process for each candidate. For 2016 I do drop a larger share of rallies than for 2012. This is because various rallies are either in a stronghold state, which will not be part of the quantitative analysis of the data and the model.

I provide a more detailed summary statistics for the states that had two or more rallies by a candidate in Table 1. It should be noted that these states provide ample cross-sectional variation in a republican candidate's poll margin lead, ranging from -5.26 pp to 8.59 pp in 2012 and from -7.1 pp to 14.3 pp. However, within-state variation in poll margins is smaller. Most standard errors are between 0.6 to 1.9,

[30]For more details on the methodology that Fivethirtyeight's forecast model uses for general elections, visit the following: https://fivethirtyeight.com/features/a-users-guide-to-fivethirtyeights-2016-general-election-forecast/

Table 1: Summary Statistics for Swing States

State	Rallies Per Day Obama'12	Romney'12	Clinton'16	Trump'16	R's Poll Margin 2012	2016	Electoral College Votes
Arizona	0	0	0	0.03	8.59	1.58	11
	(0)	(0)	(0)	(0.171)	(0.785)	(1.1)	
Colorado	0.11	0.08	0	0.08	-0.394	-4.91	9
	(0.373)	(0.339)	(0)	(0.339)	(1.2)	(1.5)	
Florida	0.06	0.2	0.15	0.23	-0.0386	-2.47	29
	(0.239)	(0.55)	(0.5)	(0.548)	(1.38)	(0.977)	
Iowa	0.15	0.1	0.04	0.05	-1.32	1.16	6
	(0.5)	(0.333)	(0.243)	(0.219)	(1.23)	(1.41)	
Michigan	0	0	0.04	0.05	-4.78	-7.17	16
	(0)	(0)	(0.197)	(0.261)	(1.82)	(1.3)	
Nevada	0.06	0.04	0.04	0.04	-2.98	-0.731	6
	(0.239)	(0.197)	(0.197)	(0.243)	(0.755)	(1.1)	
New Hampshire	0.06	0.03	0.02	0.08	-2.59	-5.91	4
	(0.278)	(0.171)	(0.141)	(0.273)	(1.4)	(1.72)	
North Carolina	0	0.03	0.09	0.15	2.4	-1.66	15
	(0)	(0.223)	(0.379)	(0.458)	(1.06)	(0.828)	
Ohio	0.2	0.26	0.1	0.13	-2.58	-0.849	18
	(0.532)	(0.613)	(0.389)	(0.442)	(1.2)	(1.69)	
Pennsylvania	0	0.02	0.09	0.16	-5.26	-6.26	20
	(0)	(0.141)	(0.351)	(0.443)	(1.21)	(1.09)	
Virginia	0.1	0.2	0	0.06	-0.996	-6.87	13
	(0.302)	(0.586)	(0)	(0.239)	(1.13)	(1.91)	
Wisconsin	0.05	0	0	0.05	-3.88	-7.1	10
	(0.219)	(0)	(0)	(0.219)	(1.67)	(1.62)	

**Standard deviations in parenthesis wherever applicable.*

[a] Note: The table shows summary stats for number of rallies in a day across states that had two or more rallies. These statistics are given for last 100 days before election. To obtain the total rallies in a state in a given election multiply the numbers by 100.

suggesting that within-state uncertainty faced by candidates while campaigning is moderate.

Summary statistics for average rallies per day are also provided in Table 1. States like Florida witnessed a significant number of rallies consistently in both elections (26 in 2012 and 38 in 2016). States like Arizona had no rallies in 2012, and in 2016 only 3 by Trump. Ohio is a state that witnessed a moderately high number of rallies in both years, as it had an average of 0.28 rallies per day in 2012 and 0.23 in 2016. On the other hand, Pennsylvania had an average of 0.02 rallies per day in 2012 but 0.25 rallies per day in 2016. Pennsylvania is an example of a state whose relative importance changed from one election to another. This pattern is reversed for Virginia, which had 30 rallies in 2012 but only 6 in 2016.

3.3 Dynamic Patterns of Political Rallies

3.3.1 Rally Ramp-up

For each candidate, rally intensity increases as election day comes close. This pattern is provided in Figure 4. To produce these plots, I consider daily rallies across all states. Then, I create 15-day bins for 90-1 days before the election and a 10-day bin for 100-91 days before an election and calculate average rallies per day, standard deviation, and the corresponding 95% confidence interval.

The pattern in Figure 4 is similar to dynamic spending patterns documented and thoroughly studied in Acharya et al. (2022) and, therefore, can be explained by the decay rate of the popularity process. In my model, the persistence parameter has a one-to-one relation with the decay rate. The critical difference

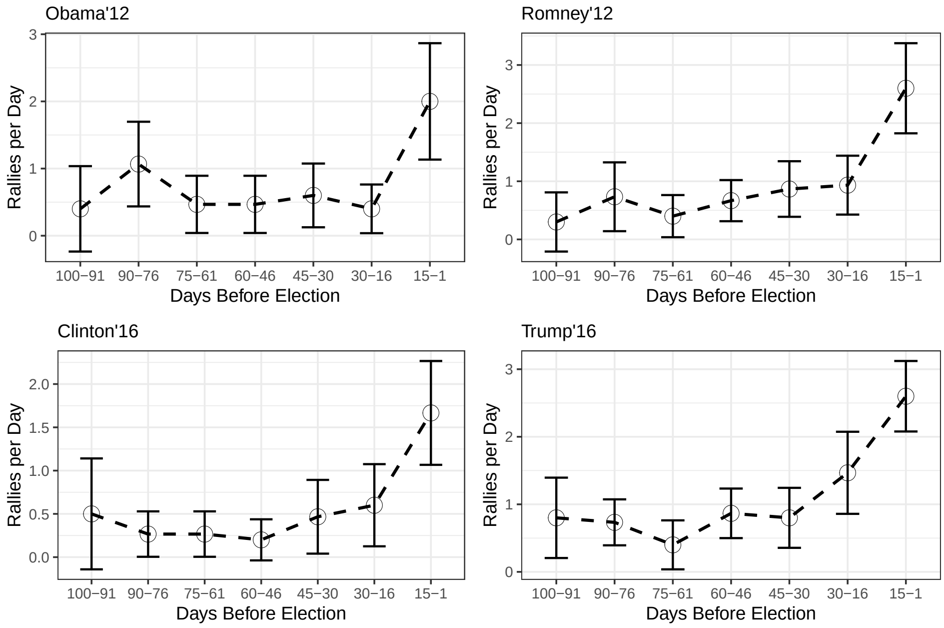


Figure 4: *This Figure shows average rallies per day for 15 day bins (10 day bin for* $100-91$ *days before election). For each of these bins the corresponding confidence interval for average rallies per day is also provided. The Figure shows that for all candidates rally intensity increases as one get closer to the election. Moreover, it is highest when there are only 15 days left before election. This can be explained by the fact that later rallies have more lasting effect on election day popularity than earlier rallies. Therefore, candidates invest more time on later rallies than earlier rallies.*

here is the distinct ramp-up patterns corresponding to each candidate within an election. These can be explained by critical parameters like the cost of rallying and rally effectiveness.[31]

3.3.2 Rallies and Poll Margin

Candidates rally in highly contested states as elections come close. This pattern relates to the qualitative prediction discussed in Section 2.4(refer to the Figure 2). To document this pattern, I create 25-day bins and analyze candidates' rallies per day in a state against their lagged poll margin lead. In Figure 5, along with a generalized additive model fit for 2012 and 2016 presidential election.

From Figure 5, it is evident that as the election comes close, candidates rally more intensely in states where candidate polls are neck and neck. A cross-sectional pattern for television advertising and vote share lead has been documented in Gordon and Hartmann (2016) and, for campaign activity in general, Strömberg (2008). However, the gradual emergence of this bell-shaped relation, especially in the context of rallies, is novel.

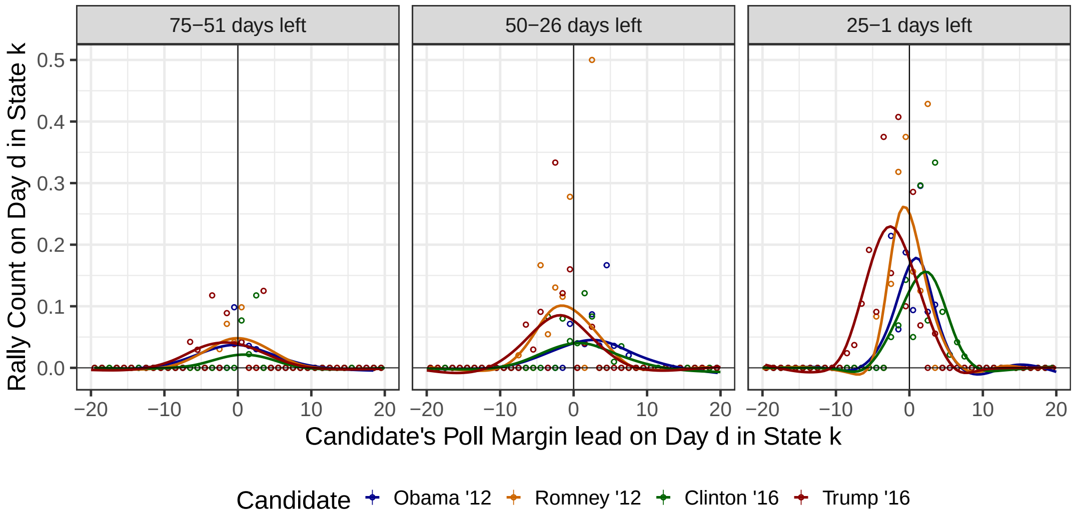


Figure 5: *This Figure shows a bin scatter of a candidate's number of rallies in a day and poll margin lead along with a generalized additive model fit. The points are bin scatter points. For each candidate, it can be seen that the fit transitions from having a weak relationship, when there are 76-51 days are left before the election, to a bell-shaped relationship when there are 25-1 days left before the election.*

Table 2: Estimates for Regression 3.1

Dependent Variable:	Rally Count ($A_{i,d,k,y}$)				
	Full Sample	Obama'12	Romney'12	Clinton'16	Trump'16
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
$\mathbb{1}\{-100 \leq d \leq -76\} \times E_k$	0.001	-0.005**	0.001	0.003	0.006**
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
$\mathbb{1}\{-75 \leq d \leq -51\} \times E_k$	0.001	-0.004*	0.006**	0.0006	0.002
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
$\mathbb{1}\{-50 \leq d \leq -26\} \times E_k$	0.005**	0.002	0.007**	0.006***	0.007**
	(0.001)	(0.002)	(0.003)	(0.002)	(0.003)
$\mathbb{1}\{-25 \leq d \leq -1\} \times E_k$	0.009**	0.002	0.008***	0.012***	0.013***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Fixed-effects					
$i \times y$	Yes	-	-	-	-
Day-Bin	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	4,800	1,200	1,200	1,200	1,200
R^2	0.03063	0.01653	0.03032	0.04647	0.04728
Within R^2	0.02790				

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

3.3.3 Rallies and Electoral College Votes

Lastly, I document how rally intensity correlates with electoral college votes within the states where competition is neck and neck. More specifically, I consider the states listed in Table 1 for this exercise.[32]

[31] In Acharya et al. (2022), these differences can not be thoroughly analyzed as the authors focus on spending ratios while I use individual choices for identifying and estimating my model.

[32] These states for 2012 have more states than the swing states used in Snyder and Yousaf (2020). For 2016, if Maine is also included, these states will be the same as the swing states used in Snyder and Yousaf (2020).

Within this set of states, candidates prioritize states with higher electoral college votes as elections come close over those with lower electoral college votes. To see this, I first divide the days into four bins. The bins are given by $B_1(d) = \mathbb{1}\{-100 \leq d \leq -76\}$, $B_2(d) = \mathbb{1}\{-76 \leq d \leq -51\}$, $B_3(d) = \mathbb{1}\{-50 \leq d \leq -26\}$ and $B_4(d) = \mathbb{1}\{-25 \leq d \leq -1\}$. I estimate the following regression:

$$A_{i,d,k,y} = \sum_{s=1}^{4} \beta_{0,s} B_s(d) + \sum_{s=1}^{4} \beta_{1,s} B_s(d) E_k + \gamma_{iy} + \epsilon_{i,d,k,y} \tag{3.1}$$

In the above regression, $A_{i,d,k,y}$ is number of political rallies candidate i held on day d in state k in election y. The variable E_k denotes the electoral college votes state k has. Since we are interested in the coefficient of E_k, the state-level fixed effects have been omitted, although we maintain candidate-level fixed effects. This regression is estimated for the whole sample and each candidate separately. Table 2 presents the results.[33]

From Table 2, for all candidates, we see that $\beta_{1,s}$ increases as elections come close; that is, as the election approaches, candidates prioritize states with a higher electorate when choosing within swing states. The candidate level analysis reveals this pattern for Trump, Clinton, and Romney. In the case of Obama, the coefficient $\beta_{1,1}$ starts with a negative and significant value and gradually becomes positive, but insignificant. Here the positive correlation still increases as elections come close.

4 Identification and Estimation

In this section I first discuss the parameterization for the model and the variation that helps to identify each parameter. Then I proceed to derive the true but infeasible likelihood function. Lastly, I describe the feasible likelihood, which is a simulated version of the infeasible likelihood function.

4.1 Parameterization and Identification

For the parameterization of rally cost parameters, I add state-level fixed costs to the existing candidate-specific parameters, which allows for cost heterogeneity across states. The cost parameters are given by $c_R, c_D, c_1, \dots, c_K$. Controlling for all state fixed effects along with the candidate specific cost of rallying leads to a multicollinearity problem in this setting. To avoid this problem I normalize one state's fixed cost to 0. I choose $c_K = 0$ and then the parameter c_R is interpreted as R's cost of rallying in state K. Moreover, parameters c_R and c_D are identified by the initial level of rallying in state K.[34] Similarly, c_k is identified by the average probability of rallying in state k for both candidates.

[33] I consider the swing states with at least two rallies by any candidate for this analysis. So the remaining set of states that we have is Pennsylvania, Michigan, Wisconsin, Nevada, New Hampshire, Ohio, Iowa, Virginia, Colorado, Florida, North Carolina, and Arizona.

[34] This is easy to see, the following equation can be derived using a similar inequality as in equation 2.17.

$$V_{it}(p_t) = \frac{\left(1 - \beta^{T-t+1}\right)\left(\ln(1 + \sum_{k=1}^{K} e^{-c_i + c_k}) + \gamma\right)}{1 - \beta} + o(T - t) = \Delta + o(T - t) \tag{4.1}$$

The second set of parameters that we are interested in is the set of parameters that govern the popularity process. These include $\alpha_R, \alpha_D, \rho, \sigma_v, \delta_1, \ldots, \delta_K$. The identification for α_i relies on candidate i's strategy and the changes in popularity, P_{ikt}, post a rally in state k at time $t-1$. The parameter ρ is identified jointly by the auto-correlation in popularity data and the gradual increase in the level of rallying with the approaching election day. The dispersion in popularity given the lagged period primitives and similar to ρ, the gradual increase in rallies jointly help in identifying σ_v. State-specific drifts, δ_k are identified by long-run means of popularity once the rallies are controlled for and by the popularity value at which the probability of rallying is highest given the period t and opponent strategy.

The data used in this paper cannot identify the parameters f, probability of R moving first, and E, maximal electoral payoff. To identify f, one would need observations on who moved first, which is unavailable. I calibrate f to a value of 0.5, as this value eliminates any ex-ante first mover or second mover advantage in the game. The parameter E, maximal electoral payoff, is not identified. Its identification relies on differences in payoff from rallying in a state and not rallying. However, multiple parameters influence this difference. Effectiveness, cost parameters, and also persistence parameter rely on this variation. Due to this, parameter E can not be identified. I calibrate its value to 538, the total number of electoral college votes in the United States. I also consider the alternative value of $E = 157$, which is the total electoral college votes states used for estimation.

4.2 Likelihood

This subsection presents the log likelihood that I use to estimate the model. Given the assumptions 2.2, 2.3 and 2.1 (or A.1) we can characterize the transition density for random vectors $\tilde{X}_t = (A_t, P_{t+1})$ for $t = 1, 2, \ldots, T$. This definition of random vectors states that $\tilde{X}_t$ is the vector containing rally decisions in period t and the popularity vector in the follow up period, which is indexed as t+1. The transition density that governs the random vectors, $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_T$, is instrumental in deriving the likelihood function. The Lemma 4.1 defines this transition density for us.

Lemma 4.1 *Given assumptions 2.2, 2.3 and 2.1 (or A.1) the random vectors $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_T$ obey the Markov property. Moreover the transition density $\psi_t(X_t|X_{t-1})$ for $t \geq 1$ is given by:*

$$\psi_t(\tilde{X}_t|\tilde{X}_{t-1}) = f(P_t|A_{R,t}, A_{D,t}, P_t)\sigma_t\left(A_t; P_t\right) \tag{4.3}$$

where $\sigma_t(.;.)$ is defined in equation 2.15 and $f(.|.,.,.)$ is defined in equation 2.3 (or A.4).

The proof for Lemma 4.1 follows straight from the equilibrium choice probabilities and the popularity process. This transition density would be the ideal choice of the likelihood construction if all popularity values were observed. I have one observation for poll margins for a day, which limits us with one observation of poll margins for every four periods. I assume that poll margins I observe at d are realized at the

Given the above holds then the difference in payoffs of rallying in state K and not rallying is given by

$$u_{ist}(K;l,p_t) - u_{ist}(0;l,p_t) = c_i + \mathbb{E}\left[V_{it}(p)|K,l,p_t\right] - \mathbb{E}\left[V_{it}(p)|K,l,p_t\right] = c_i + \delta - \delta + o(T-t) \approx c_i \tag{4.2}$$

The same also holds for the first mover choice probabilities. Note that this payoff difference directly maps to log of odds ratio of rallying in state K with respect to not rallying at all.

beginning of the day $d+1$. In other words, the poll margin I observe on the day d is isomorphic to the popularity candidates would observe in period $t = 4d+1$, which one may also call the first sub period of day $d+1$. The remaining popularity values for periods $4d+2$, $4d+3$ and $4d+4$ are missing. The idea behind this index is to map any period t with a pair (d,l), where d is a day, and l is a sub period of the day. There will be four sub-periods in each day. Therefore, for any period t, there exists a day d and sub period, l such that $t = 4(d-1)+l$.[35]

Let X_d be the observations for day d. For a day d, I observe all chosen rally choices taken by candidates. These choices are denoted by $\{A_{4d-3}, A_{4d-2}, A_{4d-1}, A_{4d}\}$ where $A_{4(d-1)+l} = (A_{R,4(d-1)+l}, A_{D,4(d-1)+l})$ for $l = 1,2,3,4$. Note that $A_{i,4(d-1)+l}$ is the rally decision taken by candidate i on day d and sub period l (or period $4(d-1)+l$). I also observe popularity, or poll margin, for day d, which I assume to be realized in sub period 1. Therefore, for a given day d, P_{4d-3} is observed, but P_{4d-2}, P_{4d-1} and P_{4d} are not. Moreover note that $P_{4(d-1)+l} = (P_{4(d-1)+l,1}, P_{4(d-1)+l,2} \dots, P_{4(d-1)+l,K})$ where $P_{4(d-1)+l,k} \in \mathbb{R}$. For day d observation I consider $d+1$ observed popularity. Hence for day d the observation is given by $X_d = \{A_{4d-3}, A_{4d-2}, A_{4d-1}, A_{4d}, P_{4d+1}\}$.

Proposition 4.1 shows that the random vectors $\{X_1, X_2, \dots, X_{\bar{D}}\}$ obey the Markov property and their density for the day to day transitions of these observations, denoted by $\lambda_d^\theta(X_d|X_{d-1})$, is also given

Proposition 4.1 *Given assumptions 2.2, 2.3 and 2.1 (or A.1) the random vectors $\{X_1, X_2, \dots, X_{\bar{D}}\}$ obeys the Markov property and its governed by the transition density $\lambda_d^\theta(X_d|X_{d-1})$, which is given by:*

$$\lambda_d^\theta(X_d|X_{d-1}) = \int_{(p_2,p_3,p_4)\in\mathbb{R}^{3K}} \left(\prod_{l=1}^{4} \sigma_{4(d-1)+l}(A_{4(d-1)+l}; p_l)\right) \times \left(\prod_{l=1}^{4} f\left(p_{l+1}|A_{4(d-1)+l}, p_l\right)\right) dp_2 dp_3 dp_4 \tag{4.4}$$

Where $p_1 = P_{4d-3}$ and $p_5 = P_{4d+1}$.

Proposition 4.1 is proved in appendix. The proof follows from applying Lemma 4.1 recursively. Based on this proposition I can formulate the likelihood of observing $X_1, X_2, \dots, X_{\bar{D}}$. Also, do note this Markov process is not time homogeneous, as the densities vary with day d. The key driving factor for making this density to vary with day d are the candidate's equilibrium choice profiles. As seen in the data and also the model predictions rally intensity increases as election comes close and therefore a Markov process that is time homogeneous can not support these features.

Moreover, the integration in Proposition 4.1 is not feasible analytically and therefore I rely on Monte Carlo methods, which are discussed in the next subsection. Note that it is a $3K$ dimensional integration, and it imposes a dimensional constraint on us. We can use $\lambda_d^\theta(X_d|X_{d-1})$ to find the log likelihood, which is given by:

$$\ell\ell\Big(\theta; X_1, X_2, \dots, X_{\bar{D}}\Big) = \sum_{d=1}^{\bar{D}} \log\left(\lambda_d^\theta\Big(X_d\Big|X_{d-1}\Big)\right) \tag{4.5}$$

[35]The uniqueness of this guaranteed as any natural number t can be expressed as an expansion in the following sense. $t = 4\times(d-1)+l$ where $d-1$ is the quotient one would obtain when t is divided by 4 and l is the corresponding remainder of this operation.

4.3 Simulated Likelihood

In this section, I will briefly describe the procedure that I used to evaluate the log-likelihood.[36] I will assume that $f(.|.)$ is given by equation 2.3. I use Quasi-Monte-Carlo method, in particular I generate $2^{10} \times 3K$ points from $3K$ dimensional Sobol sequence. Then I construct its probability integral transform to obtain the corresponding standard normal shocks. Let these shocks be denoted by $\zeta = \{\zeta^m = (\zeta^m_{1,1}, \ldots, \zeta^m_{1,K}, \ldots, \zeta^m_{3,1} \ldots, \zeta^m_{3,K})\}_{m=1}^M$. For each draw m and day d, I construct the corresponding popularity path. Let this path be denoted by $\hat{p}^{m,d}_{1,k}, \hat{p}^{m,d}_{2,k}, \ldots, \hat{p}^{m,d}_{4,k}$. We observe $\hat{p}^{m,d}_{1,k} = P_{d,1,k}$, which is the popularity value realized at the beginning of day d. Then values $\hat{p}^{m,d}_{2,k}, \ldots, \hat{p}^{m,d}_{4,k}$ are not observed and therefore are obtained by:

$$\hat{p}^{m,d}_{l+1,k} = \alpha_R \mathbb{1}\{A_{R,d,l} == k\} + \alpha_D \mathbb{1}\{A_{D,d,l} == k\} + \tilde{\alpha}\mathbb{1}\{A_{R,d,l} == k, A_{D,d,l} == k\} + \rho \hat{p}^{m,d}_{l,k} + \delta_k + \sigma_\nu \zeta^m_{l,k} \tag{4.6}$$

Each paths gives me the mean for observed popularity the next day, as shown below:

$$\hat{p}^{m,d}_{5,k} = \alpha_R \mathbb{1}\{A_{R,d,4} == k\} + \alpha_D \mathbb{1}\{A_{D,d,4} == k\} + \tilde{\alpha}\mathbb{1}\{A_{R,d,4l} == k, A_{D,d,4} == k\} + \rho \hat{p}^{m,d}_{4,k} + \delta_k \tag{4.7}$$

I construct the set of popularity sequences, $\hat{p}^{m,d}_{1,k}, \hat{p}^{m,d}_{2,k}, \ldots, \hat{p}^{m,d}_{5,k}$ for all k and m. Let this set be denoted as $\mathcal{P}_d = \{\hat{p}^{m,d} = (\hat{p}^{m,d}_{1,1}, \ldots, \hat{p}^{m,d}_{1,K}, \ldots, \hat{p}^{m,d}_{5,1}, \ldots, \hat{p}^{m,d}_{5,K})\}_{m=1}^M$.

Now, recall from Section 2 that the equilibrium does not provide a reduced form prediction and therefore I rely on numerical methods to approximate the equilibrium. Let $\hat{\sigma}_t\,(a_R, a_D; p)$ be the approximated probability that candidate R and D chose a_R and a_D conditional on current popularity p (formally defined in equation B.15). Given these two objects, I construct the approximated transition density as followed:

$$\begin{aligned} \lambda^\theta_d\left(X_d \middle| X_{d-1}\right) &\approx \hat{\lambda}^\theta_d\left(X_d \middle| X_{d-1}\right) \\ &\approx \frac{1}{M}\sum_{m=1}^{M}\left\{\left(\prod_{l=1}^{4}\hat{\sigma}_{4(d-1)+l}\left(A_{d,l}; \hat{p}^{m,d}_l\right)\right) \times \frac{1}{\sigma_\nu^K}\left(\prod_{k=1}^{K}\phi\left(\frac{P_{d+1,1,k} - \hat{p}^{m,d}_{5,k}}{\sigma_\nu}\right)\right)\right\} \end{aligned} \tag{4.8}$$

The density $\hat{\lambda}^\theta\left(X_d \middle| X_{d-1}\right)$ provides a close approximation of $\lambda^\theta\left(X_d \middle| X_{d-1}\right)$. If ζ were drawn from a standard normal distribution instead, call this density $(\hat{\lambda}^\theta\left(X_d \middle| X_{d-1}\right))$ then it is not hard to see that $\hat{\lambda}^\theta\left(X_d \middle| X_{d-1}\right) \rightarrow \lambda^\theta\left(X_d \middle| X_{d-1}\right)$ as $M \rightarrow \infty$. The error of this integral would vanish to zero with a rate of $\sqrt{M}$. However, I am using QMC, which in practice is known to provide better convergence rate if the variation of $\lambda^\theta_d(.|.)$ is finite. This is true under the condition $\sigma_\nu > \frac{1}{\Delta}$ and $1 - \rho > \frac{1}{\Delta}$ for a large $\Delta >> 0$. Therefore, the approximate log-likelihood is given by:

$$\ell\ell(\theta; X_0, X_1, \ldots, X_{\bar{D}}) \approx \hat{\ell\ell}\,(\theta; X_0, X_1, \ldots, X_{\bar{D}}) = \frac{1}{\bar{D}}\sum_{d=1}^{\bar{D}} \log\left(\hat{\lambda}^\theta_d\left(X_d \middle| X_{d-1}\right)\right) \tag{4.9}$$

I conduct an extensive set of Monte Carlo experiments to assess the finite sample bias that arises due to the two layers of approximation that I use. The first layer refers to the equilibrium approximation, and the second layer is the QMC procedure. For $K = 2$, I find that there is negligible bias. For $K = 4$, the bias is still contained within 10% of the parameter value. These Monte Carlo results are shown in Section D, specifically sub-section D.3.

[36] For a more detailed description of the numerical approximation of the equilibrium refer to Section B and for the estimation procedure that uses this numerical approximation refer to Section D.

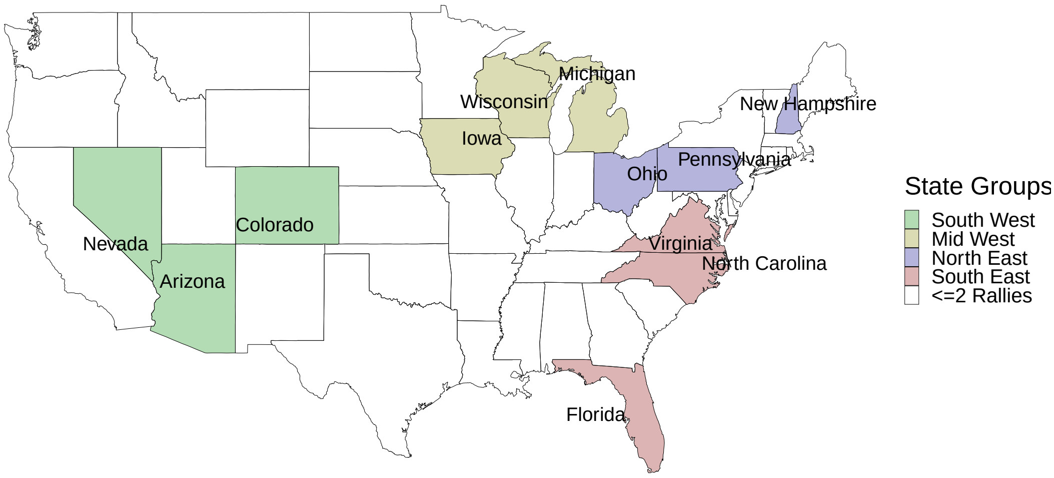


Figure 6: *The figure displays states considered in the analysis. The unlabeled states, which are omitted, had less than 2 rallies by either candidates. The states considered in the analysis are swing states. To keep dimensionality of the model under control I group these swing states into 4 groups.*

Table 3: Summary Stats for States Groups

	Rallies Per Day				R's Poll Margin		
State	Romney'12	Obama'12	Trump'16	Clinton'16	2012	2016	Electoral College Votes
South West	0.05	0.05	0.08	0.00	2.81	-1.18	26
	(0.22)	(0.22)	(0.27)	(0.00)	(0.56)	(0.96)	
Mid West	0.06	0.07	0.07	0.04	-3.85	-5.52	32
	(0.24)	(0.25)	(0.25)	(0.19)	(1.49)	(1.36)	
North East	0.17	0.14	0.18	0.11	-3.84	-3.84	42
	(0.37)	(0.34)	(0.39)	(0.31)	(1.07)	(1.28)	
South East	0.18	0.09	0.21	0.15	0.40	-3.24	57
	(0.39)	(0.28)	(0.41)	(0.35)	(1.11)	(0.96)	

**Standard deviations in parenthesis wherever applicable.*

[a] Note: The table shows summary stats for number of daily rallies and average Republican poll margin lead across states groups. These statistics are given for last 100 days before election. In 2012 Mid West and North East within swing states were more Democrat leaning while South West states were Republican leaning. South East states, however, were close to the margin. In 2016 most all states were Democrat leaning based on aggregate polls.

4.4 State Groups

Estimating the model on the whole set of U.S. states is infeasible, as it introduces roughly 50 state variables for the dynamic programming problem that each candidate solves. Even if one considers the states with at least two rallies by a candidate, the achieved dimension reduction is not sufficient for reliable estimation. As a result, I construct groups of states that allow me to estimate the model with adequate accuracy. I construct the same state groups across the two elections so that our estimates are comparable across the two elections. Therefore, I will include states like Arizona in the 2012 estimation, which had no rallies in 2012, but three rallies in 2016.

I create four groups of states, where each group is the intersection of a US region and the swing states[37] in that region. For instance, this intersection is given by swing states Nevada, Arizona, and Colorado for the South West group. South East group consists of Florida, Virginia, and North Carolina. The Mid West group consists of Michigan, Wisconsin, and Iowa. Finally, the group North East group consists of New Hampshire, Pennsylvania, and Ohio. The misplaced state is Ohio as it is a Midwest state. This is to obtain a state group with a similar number of electoral college votes as the southwest states.[38]

For calculating poll margin leads for each state group, I consider the weighted mean of poll margins for each state belonging to the group. The weights are a state's proportional electoral college votes within the state group. The summary statistics for rallies and final poll margins are provided in Table 3. For estimation I consider deviation of poll margins from the mean across all states and days, that is $\frac{1}{DK}\sum_{k=1}^{K}\sum_{d=1}^{D} P_{k,d}$ where $P_{d,k}$ is the weighted average poll margin within a state group.

5 Results

5.1 Estimates

Table 4 presents the results from the estimation exercise. Columns (1) and (2) correspond to the main parameters. Columns (3) and (4) correspond to the fixed effects used in the model. I estimate that the effectiveness of Trump's rallies, α_R, was 0.0838 pp in a state group, which is significantly larger than 0. This estimate tells us that Trump's rallies successfully gained support from the electorate. To be precise, Trump gained a lead of 0.0838 percent of votes over Clinton after a rally. This discovery is in line with what Snyder and Yousaf (2020) find in their event study regarding Trump's rallies.

In addition to Trump's rallies, I find that Clinton, Romney, and Obama rallies also successfully improved their lead in the polls. The effectiveness estimates for their rallies are 0.0745 pp, 0.0732 pp, and 0.0653 pp, which are also significant.[39] As mentioned before, the literature has found mixed evidence on whether campaign visits increase support from the electorate (Shaw, 1999; Shaw and Roberts, 2000; Shaw and Gimpel, 2012; Wood, 2016). The results in this paper support the claim that it does. This paper's findings qualitatively agree with articles that have considered candidate-specific effects (Shaw and Gimpel, 2012) and/or allowed media coverage (Shaw and Roberts, 2000) as they document that candidate campaign visits effectively gain the electorate's support.

The combined effect of both candidates rallying is indistinguishable from 0 for both presidential elections. This was particularly assumed and argued in Strömberg (2008) for the case of the 2004 and 2008 U.S. presidential elections. Strömberg (2008), did not estimate the effectiveness of state visits but assumed it to be the same across both candidates. Here, I empirically show that simultaneous rallies held by competing candidates in the 2012 and 2016 U.S. presidential elections canceled each other.

[37] Here, the swing states, for the lack of a better name, are defined as states with at least two rallies in either election.

[38] I consider two alternate definitions of states groups for robustness tests. One considers the states used in Snyder and Yousaf (2020), and the other treats Florida as an individual state group by combining Virginia and North Carolina with the North East group as Florida is geographically isolated. The results from this exercise are discussed in Section 6.

[39] Here I have dropped the signs, and the estimates are interpreted as effects of a candidate rally on their poll margin lead

Table 4: Parameter Estimates

Main Parameters			Fixed Effects		
Parameters	2012	2016	Parameters	2012	2016
Popularity					
α_R	0.0732	0.0838	δ_1	0.032	0.023
	0.0176	0.0155		0.0097	0.008
α_D	-0.0653	-0.0745	δ_2	-0.040	-0.02
	0.0180	0.0152		0.0099	0.009
ρ	0.989	0.991	δ_3	-0.035	-0.0069
	0.002	0.001		0.0084	0.0073
σ	0.147	0.16	δ_4	0.023	-0.0074
	0.0141	0.0148		0.0063	0.0073
Costs					
c_R	2.68	2.36	c_1	0.44	0.943
	0.262	0.208		0.315	0.411
c_D	2.89	3.26	c_2	0.53	0.788
	0.19	0.259		0.270	0.308
	-	-	c_3	0.094	-0.0447
	-	-		0.270	0.22

	2012	2016
Observations	100	100
Log Likelihood	-658.74	-654.54

[a] Note: The table shows estimates for the model parameters. Here the standard errors have been computed by using observation wise gradient and likelihood hessian. I use HAC estimates for this purpose to take care of correlations in gradient values. For computing the gradient and hessian I used Auto-differentiation in Julia.

The estimates of persistence in popularity are also provided in Table 4. The persistence is approximately 0.99 for both years. The weekly decay rate supported by the estimate of the persistence parameter is 28%.[40] This decay rate lies in the far right tail of the perceived decay rate distribution estimated in Acharya et al. (2022). However, the decay rate is lower than Hill et al. (2013), which is at 52.4%. The third-degree lagged polynomial specification considered in Gerber et al. (2011) estimated a decay rate of 25%, which is quite close to our estimate of 28%.

The cost estimates reflect the expected benefit threshold, measured in electoral college votes, beyond which a candidate chooses to rally in a state. The estimate of Trump's threshold is significantly lower than the estimated threshold for Clinton. These estimates reveal that Trump was more likely to hold a rally even if it had a much smaller chance of contributing to his overall success. On the other hand, Clinton was more cautious in holding rallies. In her case, a rally was held if it had a much larger chance of contributing to her overall success. This estimate captures not only an asymmetry in observed campaigning strategies between the opponents, despite having similar levels of campaign effectiveness,

[40] If λ is decay rate for Δ periods then the following relation holds $\rho = e^{-\frac{\lambda}{\Delta}}$. For weekly decay rate, $\Delta = 7 \times 4$ and therefore $\lambda = -28 \times \log(\rho)$

Table 5: Persuasion Rates

	2012		2016		Spenkuch and Toniatti (2018)	
	Romney	Obama	Trump	Clinton	Rep. Ads	Dem. Ads
Pers. Rate of **1** Rally/T.V. ad (%)	0.147	0.130	0.167	0.150	0.01	0.03
	0.0357	0.0361	0.031	0.030	0.005	0.004
Agg. Switched Decisions (in Millions)	0.388		0.577		2.2	
	0.057		0.097		-	

[a] Note: This table compares persuasion rates of rallies with those of advertising. Here I am considering persuasion rate of one rally on the election day with those estimated for advertising by Spenkuch and Toniatti (2018). Persuasion rates of a rally in my setting is defined and given by equation 5.1. The standard errors are calculated using the delta method. The aggregate number of switched decisions for rallies is given by changes in vote margins in the counterfactual regarding electoral effects of rallies. For T.V. ads the numbers are taken from Spenkuch and Toniatti (2018).

but also the asymmetry in their attitudes towards risky investments.

Meanwhile, this asymmetry is not found for the 2012 election, as the cost estimates of rallies are not significantly different for Obama and Romney.[41] However, there are differences in voter attitudes across states, specifically larger state groups. States groups such as the North East (Maine, Pennsylvania, and Ohio) and Mid West (Iowa, Michigan, and Wisconsin) were more Obama-leaning while South West states were Romney-leaning. MW and NE state groups (or SW state group) have considerable amount of electoral college votes and therefore, their inclination towards D (or R) limits the dependence of electoral success on candidate rallies, Section 8 demonstrates this.

To put rally effectiveness estimates into perspective, consider the Table 5. The table displays the persuasion rates for rallies and advertising. The persuasion rates on advertising, estimated in Spenkuch and Toniatti (2018), are also provided. I follow the definition of persuasion rates in DellaVigna and Kaplan (2007) and Spenkuch and Toniatti (2018), to derive these objects for this setting. The persuasion rate on the election day is given in equation 5.1, where V_k denotes the proportion of the voting age population in state group k.

$$f_i^{\text{Rally}} = \frac{2}{100 + (-1)^{\mathbb{1}\{i=R\}} \sum_{k=1}^{K} \frac{\delta_k * V_k}{1-\rho}} \times \left| \frac{\Delta\text{Poll Margin}}{\Delta\text{Rallies}} \right| \tag{5.1}$$

I use α_i in place of $\frac{\Delta\text{Poll Margin}}{\Delta\text{Rallies}}$ to provide persuasion rate of a rally by a candidate. The persuasion rate in this set measures the share of voters that changed their behavior in response to a political rally. While comparing these rates with advertising, I find that rallies' persuasion rates are higher than advertising's. Moreover, one would require 17 T.V. ad spots in each DMA of a state group to compensate for one Trump's rally. The same compensation ratio for Romney, Obama, and Clinton is given by 15, 4, and 5, respectively. This indicates that on average Republican rallies carried a much larger persuasive effect

[41] I acknowledge the existence of some asymmetry in campaigning strategies used by Obama and Romney, which the model does not capture. This asymmetry is in the observed correlation between electoral college votes and the number of rallies on a day within swing states. See Table 2

than their T.V. ads.

The persuasion rates for rallies are still lower than that of T.V. news. For instance the same for FOX News is $f = 11.6$ (DellaVigna and Kaplan, 2007). Assuming no decay, Trump would require to expose a population with 67 MAGA rallies to compensate for the absence of FOX News. For advertising, this trade-off amounts to 500 spots on cable T.V. (Spenkuch and Toniatti, 2018). Contrary to the hypothesis that rallies are less important than T.V. ads in an election, these numbers do raise a fair bit of ambiguity as to whether this is true or not.

While these numbers show that the persuasive effect of a rally far outweighs that of a T.V. ad, T.V. ads possess scaling capabilities that political rallies lack. Recall that political rallies are naturally constrained while, arguably, one can run T.V. ads across many media markets simultaneously.[42] Therefore, the cumulative effect of T.V. ads can outweigh that of rallies. For instance, the number of voters that switched their decisions due to rallies in 2016 was 576.6 K and in 2012 was 387.6 K.[43] For T.V. ads, the number of voters is 2.2M, roughly 4 times what it was for rallies in 2016.[44]

5.2 Model Fit

In this section, I examine the in-sample and the out-of-sample performance of the model. The model is capable of fitting the dynamic moments to a large degree. In Figure 17, I compare the model's probability of rallying in a state for each candidate with that observed in the data. I also compare the average number of daily rallies in a state between the model and the data. These comparisons are provided in Table 16. The model's predicted average number of daily rallies lies in the 95% confidence intervals of the observed average number of daily rallies.

Table 16 also presents the correlation between rally decisions predicted by the model and observed in the data. For Trump, this correlation is the lowest at 69% and the correlation is highest for Clinton at 84%. I also show the proportion of rally decisions that are correctly predicted by the model. Prediction is defined as the rally decision that has the maximum probability. By comparing these predictions with the observed decisions I find that the worst proportion of correct predictions is 73% for Trump. The highest correct prediction is for Clinton at 86%.

The model supports the correlation pattern between electoral college votes and rally count, this was the third pattern in Section 3.3. For this purpose, I divide the election periods into four bins, which are 100-76 days before the election (denoted as -4), 75-51 days before the election (denoted as -3), 50-26 days before the election (denoted as -2) and 25-1 days before the election (denoted as -1). For the

[42] Assuming prices do not change and candidates have deep pockets.

[43] For rallies I rely on the counterfactual experiment on the cumulative effect of rallies

[44] The number of rallies held in 2020 was much lower than in 2016. Due to the concern for covid-19 and voters practicing social distancing, very few rallies were held by Democrats. However, at the same time, spending on T.V. ads was much higher than ever. I conjecture that a significant proportion of this unforeseen increase in spending on Cable T.V. ads can be explained by the decline in rallies. Exactly what proportion is explained by the decline in rallies will depend on the degree of substitutability between these campaigning instruments. Estimating the trade-off between spending money on T.V. ads and holding political rallies is an open question.

Table 6: Out-of-Sample Fit

Panel (A): Comparison of Means

	Romney		Obama		Trump		Clinton	
	Model	Data	Model	Data	Model	Data	Model	Data
South West	0.171	0.1	0.184	0.15	0.13	0.2	0.0639	0
		0.14		0.18		0.2		0
Mid West	0.168	0.05	0.169	0.2	0.116	0.1	0.06	0.05
		0.1		0.2		0.14		0.1
North East	0.271	0.15	0.238	0.1	0.309	0.45	0.156	0.25
		0.18		0.14		0.29		0.22
South East	0.39	0.2	0.272	0.1	0.353	0.2	0.178	0.3
		0.2		0.14		0.2		0.24

Panel (B): Measures of Fit

	Romney	Obama	Trump	Clinton
Correlation	0.8430	0.8320	0.7019	0.8162
Mean Squared Error	0.242	0.254	0.4061	0.2684
Correct Predictions	0.8750	0.8625	0.7625	0.8500

[a] This table shows the out-of-sample model fit. Here I divide the data into two parts, where I randomly select (without replacement) 20% of the observations, call this the validation sample. I estimate the model on the remaining 80% of the data, the training sample, and then calculate model fit metrics on the validation sample. The model's predicted average number of rallies in a day lie within 1 s.d. from the observed counterparts in the validation sample. The worst correlation is 0.70. For each period, I define prediction as the option with the highest probability of choosing. I compare these predictions with the data and calculate the proportion of correct predictions. Using this metric for prediction I find that worst correct predictions is 76%.

model's prediction and the observed sample, I plot the fitted line with a bin scatter plot in Figure 19 for each day bin. The increasing correlation as the election comes close is supported by the model.

I show the out-of-sample model fit in Table 6. For this purpose, I divide the data into two sub samples, training and validation. I randomly select (without replacement) 20% of the observations for the validation sample. I estimate the model on the remaining 80% observations, the training sample, and then calculate model fit metrics on the validation sample.

The majority of the average number of daily rallies predicted by the model (for each candidate and state) lies within one standard deviation from its observed counterpart in the validation sample. The worst correlation is 0.70 corresponding to Trump's rally decisions. I also calculate the correct predictions made by the model, which ranges from 76% for Trump to 87% for Romney.

6 Robustness

In this section, I examine and show the robustness of parameter estimates from my main specification. First, I consider an alternative definition of state groups and swing states. I isolate Florida as an individual state group in the first alternative definition. In the second alternative definition, I consider the swing states used in Snyder and Yousaf (2020). Then I allow for geographic serial correlation in popularity shocks. I also consider aggregate shocks. I also test robustness to forecasting error in the polls. Apart from these, additional robustness tests in Appendix E consider a wide range of concerns.[45]

6.1 Robustness to State Group Definitions

6.1.1 Florida as Individual State Group

Florida is farther from North Carolina and Virginia than Pennsylvania, Ohio, and New Hampshire. The North East States are Pennsylvania, Ohio, New Hampshire, North Carolina, and Virginia in this test. For the South East States, I only include Florida. I estimate the model, and the results from this exercise are given in columns (1) and (2) in Table 7.

Here for 2016, I do not find significant differences in effectiveness estimates. For 2012, I find estimates do decrease, but they remain significant at the 1% level of significance.

6.1.2 Swing States used in Snyder and Yousaf (2020)

In my baseline specification, I use states that had at least two rallies by a candidate. In this robustness test, I use the swing states Snyder and Yousaf (2020) used in their event study of political rallies. I apply the same rule of state groups as before. Here for 2012, the resultant South West states are Colorado and Nevada; Mid West state group consists of Iowa and Wisconsin; North East state groups are New Hampshire and Ohio. The South East state group is the same as the baseline. For 2016 this gives me the same state groups as in the baseline for South West, Mid West, and South East. The North East states have Maine in addition to the baseline states.

I estimate the model, and the results from this exercise are given in columns (3) and (4) in Table 7. I do not find significant changes from the baseline effectiveness estimates for 2016. Estimates for 2012 do decrease but remain significant.

6.2 Robustness to Correlated Popularity Shocks

In the baseline specification, I assume that popularity shocks are uncorrelated across states; however, it could be the case that this is not true. I consider two model extensions and test how robust the base-

[45]Robustness to alternate modes of campaigning have been tested using Nielsen Ad Intel data from Kilts Center for Marketing at University of Chicago Booth and the Weselyn Media Project from University of Wisconsin. The estimates do not change significantly. These results will be released once the approval from Kilts Marketing Center has been obtained.

line estimates are. Specifically, I consider spatial correlation and the presence of aggregate shocks and estimate the model for both.

Correlated popularity shocks introduce few changes in equilibrium conditions, approximation algorithm, and also the simulated likelihood. The first change is in the expectation operators in the Proposition 2.1. These operators now account for the resulting correlated structure of popularity and they govern under Assumption A.1. A second change is in the equilibrium approximation; here, I use Cholesky factorization while evaluating expectations of value functions wherever applicable. The simulated log-likelihood in this setting is given by:

$$\ell\ell(\dots) \approx \hat{\ell\ell}\left(\theta; X_0, X_1, \dots, X_{\bar{D}}\right)$$
$$\approx \frac{1}{\bar{D}} \sum_{d=1}^{\bar{D}} \log \left[\frac{1}{M} \sum_{m=1}^{M} \left\{ \left(\prod_{l=1}^{4} \hat{\sigma}_{4(d-1)+l} \left(A_{4(d-1)+l}; \hat{p}_l^{m,d} \right) \right) \times \frac{\exp\left[-\frac{1}{2} \left(P_{d+1,1,k} - \hat{p}_{5,k}^{m,d} \right)^T \Omega^{-1} \left(P_{d+1,1,k} - \hat{p}_{5,k}^{m,d} \right) \right]}{\sqrt{2\pi \det(\Omega)}} \right\} \right] \tag{6.1}$$

Here Ω is the resulting variance-covariance matrix. This matrix will differ for each case of spatial correlation and aggregate shocks. I parameterize Ω for each case and estimate the underlying parameters of Ω along with the baseline parameters.

6.2.1 Spatial Correlation

I examine how robust the estimates are if one accounts for spatial correlation. Geographically closer state groups may exhibit a higher degree of correlation in popularity shocks than state groups farther away. This can be supported by a variance-covariance matrix where off-diagonal elements are inversely proportional to the distance between states representing the column and row. I assume variance across the states is constant, and the proportionality constant for state pairs is also constant. The variation in correlation across state pairs depends on the distance between the states constituting the pair.

To achieve spatial serial correlation, I introduce the following changes. First, the popularity equation follows a similar equation given by:

$$p_{k,t+1} = \alpha_R a_{Rkt} + \alpha_D a_{Dkt} + \rho p_{kt} + \delta_k + \nu_{kt+1} \tag{6.2}$$

Except for the shocks, every other term in the equation is the same as the baseline specification. Here the shocks $(\nu_{1,t+1}, \nu_{2,t+1}, \dots, \nu_{K,t+1})$ follow:

$$\begin{bmatrix} \nu_{1,t+1} \\ \nu_{2,t+1} \\ \vdots \\ \nu_{K,t+1} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\nu^2 & \frac{\rho_{cov}}{D_{1,2}} & \dots & \frac{\rho_{cov}}{D_{1,K}} \\ \frac{\rho_{cov}}{D_{2,1}} & \sigma_\nu^2 & \dots & \frac{\rho_{cov}}{D_{2,K}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\rho_{cov}}{D_{K,1}} & \frac{\rho_{cov}}{D_{K,2}} & \dots & \sigma_\nu^2 \end{bmatrix} \right) \tag{6.3}$$

Here parameter σ_ν accounts for standard deviation in state shocks. Parameter ρ_{corr} accounts for the degree of autocorrelation and $D_{k,l}$ is the distance between state group centroid k and l in 1000 KMs. The results from the estimation for this exercise are given in columns (7) and (8) in Table 7.

Table 7: Robustness Tests

	Alt. State Groups: Isolate Florida		Swing States in Snyder and Yousaf (2020)		Aggregate Shocks σ_{agg}		Spatial Autocorrelation ρ_{cov}		Polling Error State-Wise	
Parameters	2012	2016	2012	2016	2012	2016	2012	2016	2012	2016
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
α_R	0.0271	0.0917	0.0428	0.0838	0.0932	0.0615	0.062	0.0416	0.0982	0.106
	0.00459	0.0166	0.00788	0.0155	0.0193	0.0272	0.00891	0.0102	0.016	0.0212
α_D	-0.0211	-0.0567	-0.0295	-0.0745	-0.0415	-0.0559	-0.042	-0.0807	-0.05	-0.1
	0.00428	0.0111	0.00671	0.0152	0.0127	0.0105	0.00674	0.015	0.0102	0.0171
ρ	0.991	0.99	0.987	0.991	0.99	0.991	0.988	0.991	0.99	0.994
	0.001	0.002	0.003	0.001	0.002	0.001	0.002	0.001	0.002	0.001
σ	0.0611	0.156	0.0899	0.16	0.129	0.121	0.144	0.154	0.147	0.161
	0.00729	0.0116	0.00623	0.0148	0.0115	0.00982	0.013	0.0124	0.014	0.0149
c_R	3.33	2.73	2.82	2.36	2.9	2.37	2.77	2.22	2.84	2.37
	0.27	0.233	0.292	0.208	0.29	0.211	0.275	0.2	0.286	0.191
c_D	3.44	3.44	2.88	3.26	2.86	3.24	2.87	3.39	2.83	3.23
	0.218	0.294	0.22	0.259	0.207	0.31	0.195	0.256	0.194	0.244
ρ_{corr}	-	-	-	-	-	-	0.006	0.01	-	-
	-	-	-	-	-	-	0.001	0.002	-	-
σ_{agg}	-	-	-	-	0.068	0.104	-	-	-	-
	-	-	-	-	0.0146	0.0133	-	-	-	-
LL	-285.56	-634.45	-465.89	-654.55	-645.68	-619.17	-642.47	-625.07	-658.38	-663.32
Observations	100	100	100	100	100	100	100	100	100	100

[a] Note: The table shows estimates for model parameters under 6 modifications. Columns (1) and (2) consider state groups where Florida constitutes South Eastern states and North Carolina along with Virginia are considered to be a part of North East states. Columns (3) and (4) consider states that are used by authors in Snyder and Yousaf (2020) for their event study. Columns (5) and (6) relax the assumption of uncorrelated popularity shocks and accommodates spatial autocorrelation in popularity. Columns (7) and (8) relaxes the assumption of uncorrelated popularity shocks and accommodates aggregate shocks in popularity. Columns (9) and (10) corrects for state-specific ex-post forecast error. Specifically, I correct for the difference between election day poll margin and the observed vote shares for each state separately. Here the standard errors have been computed by using observation wise gradient and likelihood hessian. I use HAC estimation for this purpose.

I find that for the 2012 election, the estimates are not significantly different from the baseline. For 2016, I do find significant divergence. However, the rally effectiveness of both candidates is still significant. Moreover, Vuong's test reveals that for 2016 the model with aggregate popularity shocks explains the data better. The correlation, I estimate, is likely an attempt to predict the variance in aggregate shocks. Therefore, this model is not valid for 2016.

6.2.2 Aggregate Popularity Shocks

The presence of aggregate shocks can also lead to a correlation in popularity shocks. For this purpose, I model the aggregate shocks and simplify the system of popularity equations to obtain the resulting variance and covariance matrix. Consider the baseline popularity equation with added popularity shocks.

$$p_{k,t+1} = \alpha_R a_{Rkt} + \alpha_D a_{Dkt} + \rho p_{kt} + \tilde{\nu}_{k,t+1} \tag{6.4}$$

Here $\tilde{\nu}_{k,t+1} = \sigma_\nu \nu_{k,t+1} + \sigma_{agg}\mu_{t+1}$. The parameter σ_{agg} represents the standard deviation in aggregate shocks, and μ_{t+1} represents the shock itself. I assume that state-specific shocks $\nu_{k,t+1}$ are orthogonal to each other and also to μ_{t+1}. Note that the net shock $\tilde{\nu}_{k,t+1}$ is correlated across states despite assuming orthogonality on $\nu_{k,t+1}$. This implies that the vector of popularity shocks $(\tilde{\nu}_{1,t+1}, \tilde{\nu}_{2,t+1}, \ldots, \nu_{K,t+1})$ satisfy the following:

$$\begin{bmatrix} \tilde{\nu}_{1,t+1} \\ \tilde{\nu}_{2,t+1} \\ \vdots \\ \tilde{\nu}_{K,t+1} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\nu^2 + \sigma_{agg}^2 & \sigma_{agg}^2 & \cdots & \sigma_{agg}^2 \\ \sigma_{agg}^2 & \sigma_\nu^2 + \sigma_{agg}^2 & \cdots & \sigma_{agg}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{agg}^2 & \sigma_{agg}^2 & \cdots & \sigma_\nu^2 + \sigma_{agg}^2 \end{bmatrix} \right) \tag{6.5}$$

The estimates from this case are given in columns (5) and (6) of Table 7. I find that the estimates for the 2012 and 2016 elections are not significantly different from the baseline. As mentioned, this model provides a better fit than the spatial autocorrelation model for 2016. This is evident by comparing the loglikelihood values, −625.07 for spatial autocorrelation and −619.17 for aggregate shocks.

6.3 Robustness to Polling Error

One of the key issues that were observed during the 2016 presidential elections was that polls failed to predict that Trump would win the election. This essentially points out to substantial measurement error in the polling data that I use for my empirical application. To test how robust my estimates are if we account for polling errors, I calculate ex-post state-specific polling errors. This is calculated by comparing election day R's poll margin obtained from FiveThirtyEight with the R's observed vote share margin (election result). I calculate the differences for each state and correct for this error for each daily poll. This corrected poll margin data is then used for estimation instead of the de-meaned polls margins in the baseline specification. The results from this exercise are provided in columns (9) and (10) in Table 7. Note that the estimates do not change significantly.

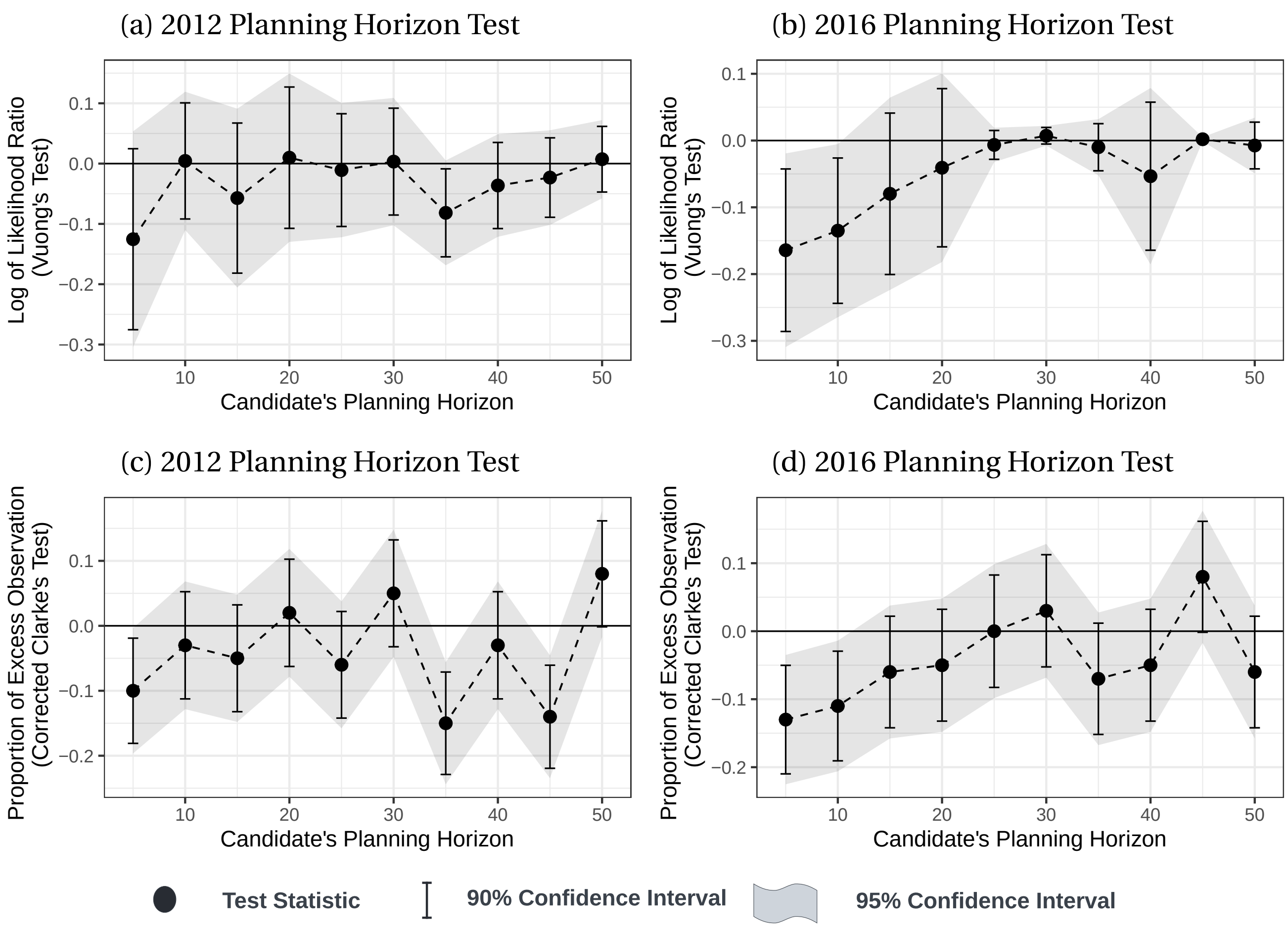


Figure 7: *This figure shows results from conducting Vuong's closeness test and also the Corrected Clarke's test (Brück et al., 2022). Here I consider the comparisons between the 'Full Planning Horizon' (baseline model) with models that have different possible length of the planning horizon. A negative statistic indicates that the 'Full Planning Horizon' model performs better than the competing model. Here x-axis shows the potential values of the planning horizon that I compare with the full planning horizon and the y-axis shows the corresponding Vuong's Closeness Test statistic for each comparison.*

7 Model Selection

This section tests the validity of behavioral assumptions made on candidates. The first assumption is that candidates can execute backward induction flawlessly. This might be untrue if candidates are myopic, if the voters are more attentive closer to the election, or if voters have limited memory spans (i.e., they do not recall earlier rallies by candidates). I assume that candidates, while optimizing, take into account these factors. Candidate behavior will be identical if any of the above is true. Therefore, for the statistical test, it does not matter whether politicians are myopic or their target population is. The second assumption is that candidates held rallies strategically. It could be the case that candidates do not consider opponents' rally decisions and only make their decisions based on a state's electoral payoff and their popularity in the state. I test whether this is true or not.

While researchers have considered strategic and myopic behavior in sequential voting settings, it has not been studied in a setting where these behavioral features are separable. For instance, in Spenkuch et al. (2018), suppressing strategic behavior is identical to suppressing forward-looking behavior. Therefore, whether candidates are myopic or non-strategic when they vote on bills is unclear. Here these features are separable and can be individually tested. I discuss the results and the procedures for these two tests in the following subsections.

7.1 Uncovering Planning Horizon

The model assumes that presidential candidates are rational and can indefinitely compute or forecast future outcomes. In this section, I analyze the extent to which this assumption holds. I assume that candidates may forecast potential outcomes up to $\tilde{D}$ days into the future. The payoffs candidates receive beyond $\tilde{D}$ are assumed to be zero. For each planning horizon limit $\tilde{D}$, I re-estimate the model and then carry out a 1) Vuong's closeness test and also 2) Corrected Clarke's Test (Brück et al., 2022). The results of this exercise are in Figure 7.

In this setting, myopic candidates will exhibit two key features. Consider the period $\tilde{t} = 4(D - \tilde{D}) + 1$. This is when election day enters the planning horizon of a candidate, who can backward induce up to $\tilde{D}$ days. Before this period, candidates do not consider their electoral payoffs, so their campaigning decisions do not relate to how popular they are in a state. However, once this period passes, the electoral payoffs enter their optimization problem. Equilibrium rally decisions of a candidate, at all periods from this point, exhibit a correlation with their relative popularity. This feature is absent if a candidate is not myopic. There is also an increase in the probability of holding a rally for myopic candidates (irrespective of their popularity) because there is a positive return to holding rallies, which is not the case in periods before $\tilde{t}$.

Note that these patterns can also arise if voters have a limited memory span or if they are more attentive towards politicians closer to the election. I assume that politicians consider these behavioral features of voters when they campaign. Therefore, it does not matter whether politicians act myopically because they are myopic or because their voters have the aforementioned behavioral features. Under either of these cases, as econometricians, we must see campaigning with myopic features.

From Figure 7, it is evident that we fail to reject the full horizon model against models that support planning horizon of less than 100 days. This reassures us that the data does not reject our assumption about candidates' forward-looking capacity or the fact that I ignore voters' behavioral constraints.

Moreover, these tests can also be used to find the minimum planning horizon needed to explain the data at the same level as the full horizon model. I find that for 2012, a 10 day planning horizon model can also explain the data as good as the baseline (using corrected Clarke's test), which indicates that candidates had at least a 10 day planning horizon. I also find that for 2016, a 15 day planning horizon model can also explain the data as good as the baseline, indicating that in 2016 candidates had at least a 15 day planning horizon.[46]

[46]These planning horizons are not interpreted as "booking a rally 15 days in advance," but as how far can candidate compute while campaigning. This is similar to what Spenkuch et al. (2018) studied for roll call voting.

Table 8: Strategic Behavior Test

	Election Year		
	2012	2016	Pool Both
Clarke's Test	5	9*	14**
90% CI	[-8,8]	[-8,8]	[-12,12]
95% CI	[-10,10]	[-10,10]	[-14,14]
Corrected Clarke	0.05	0.09*	0.07**
(Brück et al., 2022)	0.05	0.049	0.035
Vuong's Closeness Test	0.00133	0.00079	0.0083
	0.038	0.0453	0.0243
Observations	100	100	200

[a] Note: The table shows the results from model selection tests between strategic and non strategic model. For the non strategic model I assume that candidate do not observe nor anticipate opponent actions, they believe that the opponent does not rally. I estimate the model with this assumption for 2012 and 2016 elections. I calculate model selection test statistics where positive values indicate that strategic model performs better.

7.2 Strategic Campaigns

I test whether candidates chose rallies strategically or not. For this purpose, I construct a model of irrational candidates who decide based on their current popularity and the electoral size of a state. Here candidates do not anticipate nor observe opponent actions. They assume that the opponent does not rally while making decisions. The conditional choice probabilities under these assumptions still exhibit the pattern showed in Figure 2. However, these probabilities in equilibrium are now orthogonal across candidates.

I estimate the model and carry out three model selection tests. Here positive values indicate that the strategic model performs better than the non-strategic model. The first test I use is the Clarke's test, which shows that in 2016 the strategic model performed better at 90% level of significance. The corrected version of the Clarke Test (Brück et al., 2022) also produces similar conclusions. For the 2012 election, however we fail to reject the strategic model. Moreover, by pooling both elections together, I find strategic model performs better at 5% level of significance.

8 Counterfactual Experiments

I execute two main counterfactual experiments. The first counterfactual experiment estimates the cumulative effect of political rallies. The success of a rally today may not translate wholly into electoral outcomes because, in this dynamic game, the effects of rallies decay with time. However, on the other hand, candidates hold multiple rallies to cumulate these effects. The decay and cumulation formulate the opposing channels that determine whether rallies significantly affect electoral outcomes. If a candidate does not rally sufficiently, then decay dominates, making rallies ineffective.

I also consider counterfactual experiments that analyze how effectively campaign silences regulate elections. Campaign silence is also known as pre-election silence, where there is a partial to a complete ban on campaigning. Along with such bans, some countries also ban on release of polls to the public. Canen (2018) studied the complete ban feature and found that complete bans can reduce social welfare by reducing the amount of information available to voters. However, as argued before, little evidence shows rallies are truly informative (Snyder and Yousaf, 2020). Moreover, multiple anecdotes show that a substantial amount of misinformation (New York Times, 2020; Politico, 2020; New York Times, 2021) is released in rallies. Therefore, I see such intervention positively as it can limit the number of rallies (a potential source of misinformation).

8.1 Cumulative Effect of Rallies

In this exercise, I compare electoral outcomes under two cases. In the first case, call it "Only one candidate rallies," we allow only one candidate to hold political rallies while the opponent does not. This setup predicts the outcomes of one candidate's campaigning efforts alone and removes the effect of the opponent's counter-campaigning. The second case, call it "None rally," considers electoral outcomes when there are no rallies by any candidate. For this, we simulate the popularity regressions, defined in equation 2.1, for all states and set rally decisions to 0 for all states and both candidates.

The difference between the outcomes obtained under the two cases isolates the effect of *total rallying* by a candidate on election results, such as vote shares and winning probability. To calculate the cumulative effect of *total rallying* on vote shares and winning probability, I simulate the model S times, and for each simulation s, I calculate election day vote shares and whether the candidate won or lost the election for each case of "Only one candidate rallies" and "None rally." The average differences between these two cases give me the estimate of this cumulative effect. The following two equations define these quantities.

$$\text{Vote Margins} \quad \Delta V = \frac{1}{S}\sum_{s=1}^{S}\sum_{k=1}^{K}\left\{\left(p^{s}_{T+1,k,i\text{Rallies}} - p^{s}_{T+1,k,\text{None Rally}}\right)\times \text{Total Votes received by both in } k\right\} \tag{8.1}$$

$$\begin{aligned}\text{Winning Probability} \quad \Delta W = & \frac{1}{S}\sum_{s=1}^{S}\mathbb{1}\left\{\sum_{k=1}^{K}\mathbb{1}\left\{p^{s}_{T+1,k,i\text{ Rallies}} > 0\right\}E_k > 269 - \text{EC votes in resp. stronghold}\right\} \\ & - \frac{1}{S}\sum_{s=1}^{S}\mathbb{1}\left\{\sum_{k=1}^{K}\mathbb{1}\left\{p^{s}_{T+1,k,\text{None Rally}} > 0\right\}E_k > 269 - \text{EC votes in resp. stronghold}\right\}\end{aligned} \tag{8.2}$$

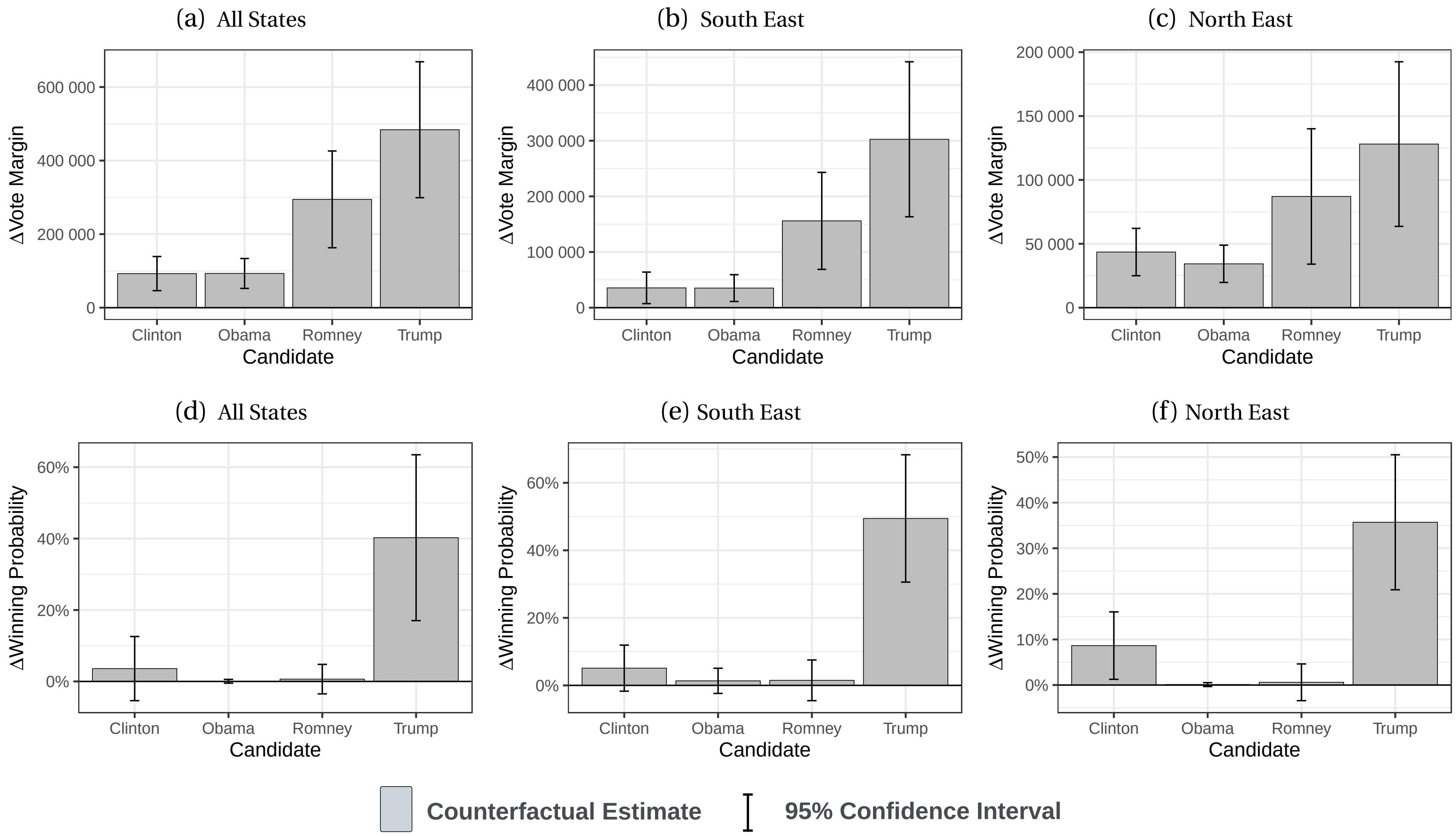


Figure 8: *This figure shows the cumulative effect a candidate's rallies had on their vote margin lead and winning probability. For each candidate, first I draw 400 draws parameter values from the asymptotic distribution of the model parameter estimates. Then for each draw I simulate the model outcomes for the cases of (i) only the candidate rallies and (ii) none rally. Then I take the differences of these outcomes across (i) and (ii). The variance of the distribution of these differences are used to formulate the confidence intervals.*

Note that for the case of vote margins, decay and cumulation are the sole channels determining the effect of *total rallying*. In the case of winning probability, a state's natural inclination is also a crucial factor. The parameter δ_k determines how pivotal a state is in a given year.

For estimating the standard error of the counterfactual estimates, I generate a sample (of size M) of parameter values from the asymptotic distribution of the estimated parameters.[47] For each draw, I calculate outcomes under the case of *"None rally"*, call it $y^m_{\text{None rally}}$, and *"i Rallies"*, call it $y^m_{i\text{ Rallies}}$, following the same procedure as before. The desired standard error are calculated by using the random variables $\left\{\Delta y^m : \Delta y^m = y^m_{i\text{ Rallies}} - y^m_{\text{None rally}}\right\}_{m=1}^{M}$.

The results from this exercise are given in Figure 8 and Table 17. It is important to note that all candidates significantly increased their vote shares. However, Trump's cumulative effect dwarfs the effect Clinton and Obama had. Further, the effect Trump had on winning probability is particularly large. It amounts to a 40% increase in winning probability. The rallies by other candidates did not affect the winning probability. This result shows that Trump's rallies were pivotal, while other candidates' rallies were not.

This finding contributes to the age old question of "Do campaigns matter?" (Lazarsfeld et al., 1968; Berelson et al., 1986; Jacobson, 2015). It highlights that in addition to frequently studied campaigning instrument, T.V. ads, presidential candidates can also effectively use political rallies to win elections. Moreover, political rallies in a competitive election can be electorally pivotal and secure a win. This finding disagrees with previous research that concludes electoral results have minimal dependence on presidential campaigns (Franz and Ridout, 2010; Huber and Arceneaux, 2007; Jacobson, 2015). It agrees with the findings obtained in Political Economy and Quantitative Marketing that presidential campaigns can have substantial effects on electoral results and voting decisions (Spenkuch and Toniatti, 2018; Gordon and Hartmann, 2013).

8.2 Campaign Silence

In this exercise, I introduce campaign silence for a fixed duration $\tilde{D}$ right before the election. When campaign silences are imposed, candidates can not hold rallies from day $D - \tilde{D} + 1$ to D. In terms of periods, the candidate can not hold rallies between periods $T - 4\tilde{D} + 1$ to period T. Imposing this restriction alters the continuation values for candidates in periods where they can rally and therefore alter their behavior.

In this setting, candidate responses vary from state to state. The total rallies held until the commencement of the campaign silence determines the accumulated popularity. If the campaign silence is short, the accumulated popularity will marginally decay. This will result in an ineffective campaign silence. If it is long, then the decay in accumulated popularity will be substantial and electoral outcomes will change significantly.

[47]The asymptotic distribution of the parameter values is the normal distribution with the mean given by the parameter estimates. The variance-covariance matrix of this normal distribution is given by the consistent estimator of the parameter estimate's variance-covariance matrix

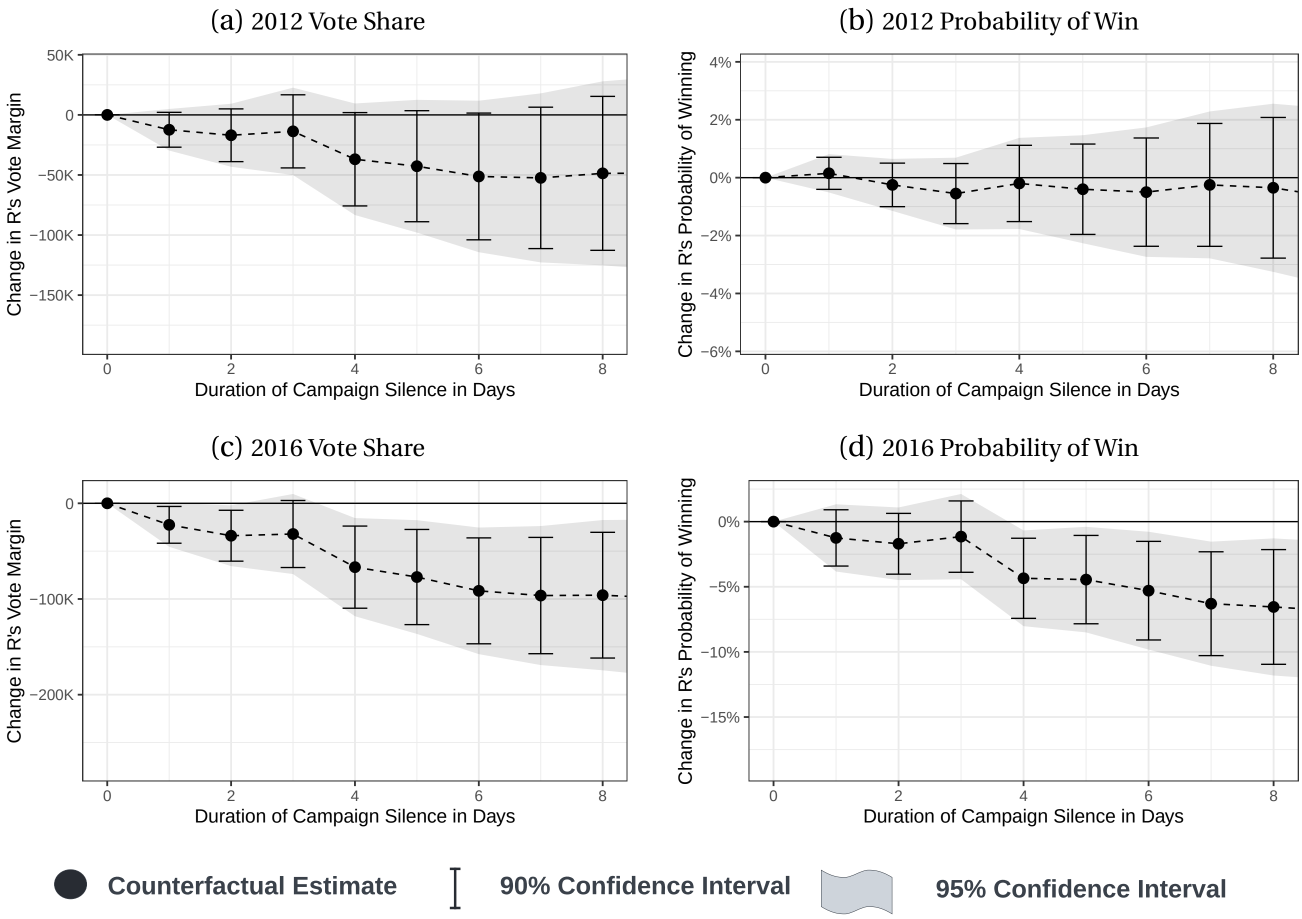


Figure 9: *Campaign Silence Duration and Election Day Poll Margin for 2012 Presidential Elections. This figure provides estimates for changes in electoral outcomes when campaign silence of varying duration lengths are imposed. For each campaign silence duration, I calculate the R's probability of winning along with the corresponding confidence intervals for these probabilities.*

I consider a grid of campaign silence length in days, $\tilde{D}$, ranging from 1 day up to 8.[48] For each campaign silence length, I compare electoral outcomes with the case without any campaign silence. I use a similar procedure to the previous subsection to estimate the expected change and the associated standard error.

I find campaign silences are ineffective if an election is completely lopsided. Campaign silences can have an effect when elections are highly competitive. In the data, 2012 was a lopsided election where Obama had massive support from voters, while 2016 was a competitive election where Trump won marginally over Clinton.

More importantly, shorter campaign silence— even in a competitive election— are ineffective. This can be seen by no effect on winning probability for campaign silences that last shorter than three days. If campaign silence is sufficiently long, the candidate that relies more on campaigning will see a decline in his/her chances of winning.

[48]This is for brevity, I initially considered up to 30 day campaign silence. The effect of campaign silence on change in election result weakly increases for 2016 and stays roughly the same for 2012.

9 Conclusion

This paper shows that political rallies can be persuasive, electorally pivotal, and hard to regulate, even in a consolidated democracy. I show this in two steps. In the first step, the paper constructs a dynamic game where politicians compete against each other to stay popular on election day. The game possesses a finite time horizon and a perfect information structure. The combination of these features is sufficient for applying backward induction to compute equilibrium conditional choice probabilities, which are unique. In this model, stage games satisfy the Markov Property, which is used to formulate a likelihood function and estimate model parameters. This model allows for estimation in settings where only one game is observed by using the stage games as a unit of observation.

From the analysis of electoral effects, I find that Trump's rallies were pivotal as they increased his chances of winning by 40%. However, rallies had no advantageous effects for Romney, Clinton, and Obama. I also analyze the persuasion rates (DellaVigna and Kaplan, 2007; Spenkuch and Toniatti, 2018) of rallies and find that a single rally is more persuasive than a Television ad. However, due to scalability constraints, cumulatively, rallies fall short as there are many more T.V. ads that candidates can use than rallies.

Policy-relevant counterfactual experiments reveal that campaign silences that last less than four days are ineffective. In many countries, campaign silence policies last only 1-2 days. However, these bans (as shown in the paper) lack regulatory power due to their short duration. Short durations give highly effective candidates sufficient time to hold multiple rallies and cumulate popularity among the voters. Ultimately, the induced decay by a short campaign silence law is insufficient to dissipate the accumulated popularity among voters.

Direct campaign communication is more prevalent in developing countries (Bidwell et al., 2020; Szwarcberg, 2012; De la Torre and Conaghan, 2009; Paget, 2019) such as India, Tanzania, and Siera-Leone. Some of these instruments, such as political rallies, can provide populist leaders with an uncontested platform where these leaders can make any claims. Less informative voters, especially those who lack the means to verify the claims made by politicians, can be easily persuaded using rallies. The proportion of such voters is higher in developing countries than in developed countries. Future research should explore if the persuasive effects of rallies in a developing country are much higher than in a developed country.

During the 2019 Indian General Elections, Narendra Modi held 115 rallies and majority of these rallies were held in states that once were INC's[49] stronghold. Massive rallies formed a significant part of Narendra Modi's electoral campaign in the 2014 Indian General Elections. Preliminary analysis shows a positive correlation between Modi rallies and the share of votes received by the BJP in 2019.[50] The model in this paper can be applied to this setting with suitable modifications. The model will address the selection bias Modi's rallies had and uncover the persuasive effects of his rallies.

[49] Indian National Congress party.

[50] Based on my research on rallies in India.

References

Acharya, A., E. Grillo, T. Sugaya, and E. Turkel (2022). Electoral campaigns as dynamic contests.

Aguirregabiria, V. and P. Mira (2007). Sequential estimation of dynamic discrete games. *Econometrica 75*, 1–53.

Ailliot, P. and F. Pene (2015). Consistency of the maximum likelihood estimate for non-homogeneous markov–switching models. *ESAIM: Probability and Statistics 19*, 268–292.

Al Jazeera (2019). Half a million attend opposition rally to remove india's modi.

Andonie, C. and D. Diermeier (2019). Impressionable voters. *American Economic Journal: Microeconomics 11*(1), 79–104.

Appleman, E. M. (2012). https://www.p2012.org/.

Appleman, E. M. (2016). https://www.p2016.org/.

Arcidiacono, P., P. Bayer, J. R. Blevins, and P. B. Ellickson (2016). Estimation of dynamic discrete choice models in continuous time with an application to retail competition. *The Review of Economic Studies 83*, 889–931.

Berelson, B. R., P. F. Lazarsfeld, and W. N. McPhee (1986). *Voting: A study of opinion formation in a presidential campaign*. University of Chicago Press.

Bidwell, K., K. Casey, and R. Glennerster (2020). Debates: Voting and expenditure responses to political communication. *Journal of Political Economy 128*, 2880–2924.

Bryan, W. J. (1909). *Speeches of William Jennings Bryan*, Volume 2. Funk & Wagnalls Company.

Brück, F., J.-D. Fermanian, and A. Min (2022). A corrected clarke test for model selection and beyond. *Journal of Econometrics*.

Buggle, J. C. and S. Vlachos (2022). Populist persuasion in electoral campaigns: Evidence from bryan's unique whistle-stop tour.

Canen, N. (2018). Information accumulation and the timing of voting decisions. *Available at SSRN 3514328*.

Chintagunta, P. K. and N. J. Vilcassim (1992). An empirical investigation of advertising strategies in a dynamic duopoly. *Management science 38*(9), 1230–1244.

De la Torre, C. and C. Conaghan (2009). The hybrid campaign: Tradition and modernity in ecuador's 2006 presidential election. *The International Journal of Press/Politics 14*(3), 335–352.

de Roos, N. and Y. Sarafidis (2018). Momentum in dynamic contests. *Economic Modelling 70*, 401–416.

DellaVigna, S. and E. Kaplan (2007, 08). The Fox News Effect: Media Bias and Voting*. *The Quarterly Journal of Economics 122*(3), 1187–1234.

Doganoglu, T. and D. Klapper (2006). Goodwill and dynamic advertising strategies. *Quantitative Marketing and Economics 4*(1), 5–29.

Donaldson, G. (1999). *Truman Defeats Dewey*. University Press of Kentucky.

Erikson, R. S. and T. R. Palfrey (2000). Equilibria in campaign spending games: Theory and data. *American Political Science Review 94*(3), 595–609.

Franz, M. M. and T. N. Ridout (2010). Political advertising and persuasion in the 2004 and 2008 presidential elections. *American Politics Research 38*(2), 303–329.

Garcia-Jimeno, C. and P. Yildirim (2017). Matching pennies on the campaign trail: An empirical study of senate elections and media coverage. Technical report, National Bureau of Economic Research.

Gerber, A. S., J. G. Gimpel, D. P. Green, and D. R. Shaw (2011). How large and long-lasting are the persuasive effects of televised campaign ads? results from a randomized field experiment. *American Political Science Review 105*, 135–150.

Globo (2020). Eleições 2020: propaganda eleitoral no rádio e na tv começa nesta sexta-feira; veja regras.

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

Gordon, B. R. and W. R. Hartmann (2016). Advertising competition in presidential elections. *Quantitative Marketing and Economics 14*, 1–40.

Grosset, L. and B. Viscolani (2004). Advertising for a new product introduction: A stochastic approach. *Top 12*, 149–167.

Gul, F. and W. Pesendorfer (2012). The war of information. *The Review of Economic Studies 79*(2), 707–734.

Heersink, B. and B. D. Peterson (2017). Truman defeats dewey: The effect of campaign visits on election outcomes. *Electoral Studies 49*, 49–64.

Heiss, F. and V. Winschel (2008, 5). Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics 144*, 62–80.

Hill, S. J., J. Lo, L. Vavreck, and J. Zaller (2013). How quickly we forget: The duration of persuasion effects from mass communication. *Political Communication 30*, 521–547.

Huber, G. A. and K. Arceneaux (2007). Identifying the persuasive effects of presidential advertising. *American Journal of Political Science 51*(4), 957–977.

IFES (2012). Ifes-indonesia unofficial translation of law no. 8 /2012 on legislative elections.

Jacobson, G. C. (2015). How do campaigns matter? *Annual Review of Political Science 18*, 31–47.

Johnstone, C. L. and R. J. Graff (2018). Situating deliberative rhetoric in ancient greece: The bouleutêrion as a venue for oratorical performance. *Advances in the History of Rhetoric 21*(1), 2–88.

Judd, K. L. (1992, 12). Projection methods for solving aggregate growth models. *Journal of Economic Theory 58*, 410–452.

Judd, K. L., L. Maliar, S. Maliar, and I. Tsener (2017). How to solve dynamic stochastic models computing expectations just once. *Quantitative Economics 8*, 851–893.

Judd, K. L., L. Maliar, S. Maliar, and R. Valero (2014). Smolyak method for solving dynamic economic models: Lagrange interpolation, anisotropic grid and adaptive domain. *Journal of Economic Dynamics and Control 44*, 92–123.

Kawai, K. and T. Sunada (2022). Estimating candidate valence. Technical report, National Bureau of Economic Research.

Knews (2022). Cyprus enters dome of silence ahead of election.

Kwon, H. D. and H. Zhang (2015). Game of singular stochastic control and strategic exit. *Mathematics of Operations Research 40*(4), 869–887.

Lazarsfeld, P. F., B. Berelson, and H. Gaudet (1968). The people's choice. In *The people's choice*. Columbia University Press.

Marinelli, C. (2007). The stochastic goodwill problem. *European Journal of Operational Research 176*, 389–404.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*.

Meirowitz, A. (2008). Electoral contests, incumbency advantages, and campaign finance. *The Journal of Politics 70*(3), 681–699.

Milgrom, P. R. and R. J. Weber (1985). Distributional strategies for games with incomplete information. *Mathematics of operations research 10*(4), 619–632.

New York Times (2020). Rallies are the core of trump's campaign, and a font of lies and misinformation.

New York Times (2021). How a presidential rally turned into a capitol rampage.

Paget, D. (2019). The rally-intensive campaign: A distinct form of electioneering in sub-saharan africa and beyond. *The International Journal of Press/Politics 24*(4), 444–464.

Pickles, W. (1960). *The French constitution of October 4th, 1958.* Stevens.

Polborn, M. K. and D. T. Yi (2006). Informative positive and negative campaigning. *Quarterly Journal of Political Science 1*(4), 351–371.

Politico (2020). Trump rallies his base to treat coronavirus as a 'hoax'.

Pouzo, D., Z. Psaradakis, and M. Sola (2022). Maximum likelihood estimation in markov regime-switching models with covariate-dependent transition probabilities. *Econometrica 90*(4), 1681–1710.

Shaw, D. R. (1999). A study of presidential campaign event effects from 1952 to 1992. *The Journal of Politics 61*, 387–422.

Shaw, D. R. and J. G. Gimpel (2012). What if we randomize the governor's schedule? evidence on campaign appearance effects from a texas field experiment. *Political Communication 29*, 137–159.

Shaw, D. R. and B. E. Roberts (2000). Campaign events, the media and the prospects of victory: The 1992 and 1996 us presidential elections. *British Journal of Political Science 30*, 259–289.

Snyder, J. M. (1989). Election goals and the allocation of campaign resources. *Econometrica 57*, 637.

Snyder, J. M. and H. Yousaf (2020). Making rallies great again: The effects of presidential campaign rallies on voter behavior, 2008-2016. Technical report, National Bureau of Economic Research.

Spenkuch, J. L., B. P. Montagnes, and D. B. Magleby (2018). Backward induction in the wild? evidence from sequential voting in the us senate. *American Economic Review 108*, 1971–2013.

Spenkuch, J. L. and D. Toniatti (2018, 05). Political Advertising and Election Results*. *The Quarterly Journal of Economics 133*(4), 1981–2036.

Strömberg, D. (2008). How the electoral college influences campaigns and policy: The probability of being florida: American economic review, 98(3), 769-807. *American Economic Review 98*, 769–807.

Szwarcberg, M. (2012). Uncertainty, political clientelism, and voter turnout in latin america: Why parties conduct rallies in argentina. *Comparative Politics 45*(1), 88–106.

van der Blom, H. (2016). *Oratory and Political Career in the Late Roman Republic.* Cambridge University Press.

Watanabe, T. and H. Yamashita (2017). Existence of a pure strategy equilibrium in markov games with strategic complementarities for finite actions and finite states. *Journal of the Operations Research Society of Japan 60*, 201–214.

Wood, T. (2016). What the heck are we doing in ottumwa, anyway? presidential candidate visits and their political consequence. *The ANNALS of the American Academy of Political and Social Science 667*, 110–125.

A Proofs

A.1 Generalization of Proposition 2.1

A.1.1 Generalization of the Model

Here I will describe a more general model that can support intertemporal correlations in costs. The timing of information revelation and decision making remains the same. The game will remain a game of perfect information, as in the main text, as we assume every candidate observes every nature realization in the prior periods. However, to allow for general correlation structures the bellman equations need to be redefined.

History: First, I introduce another set of costs, ξ_{imtk}, that can be deterministic or random variables. These will allow for various different types of correlations in costs. I will provide more details on these costs later in the section. Second I define history for the game as followed. Start with $h^1 = p_1$ and h^t for $t = 2, 3, \ldots, T, T+1$ is defined as:

$$h^t = \left(h^{t-1}, f_{t-1}, \xi_{f_{t-1},f,t-1}, \epsilon_{f_{t-1},f,t-1}, a_{f_{t-1},f,t-1}, \xi_{s_{t-1},s,t-1}, \epsilon_{f_{t-1},f,t-1}, a_{s_{t-1},s,t-1}, \nu_t, p_t\right) \tag{A.1}$$

Where f_t denotes the first mover picked by nature in period t and $s_t \in \{R, D\}/f_t$ denotes the second mover. Moreover, recall that ν_t denotes popularity shocks. Here $a_{f_{t-1},f,t-1}$ denotes the action chosen by the first mover, similarly $a_{s_{t-1},s,t-1}$ is defined. The idiosyncratic costs shocks $\epsilon_{f_{t-1},f,t-1}$ and $\epsilon_{s_{t-1},s,t-1}$, and the costs $\xi_{f_{t-1},f,t-1}$ and $\xi_{s_{t-1},s,t-1}$ follow the same nomenclature.

Popularity: Popularity follows the same definition as in the main text. It follows the following AR(1) process.

$$p_{k,t+1} = \alpha_R a_{Rk,t} + \alpha_D a_{Dk,t} + \rho p_{kt} + \delta_k + \nu_{k,t+1} \tag{A.2}$$

Here the order of play doesn't matter. What matters is who chose what. That is why the subscripts f and s have been redacted from this equation. All terms have the same definition as in the main text. $\nu_{k,t}$ is a random variable indicating a generic popularity shock. Assumption A.1 states the assumption that allows for arbitrary correlation in $\nu_{k,t}$ across states.

Assumption A.1 (General Popularity Shocks) *The popularity shocks $(\nu_1, \nu_2, \ldots, \nu_K)$ are distributed according to a multivariate normal distribution.*

$$(\nu_{1,t}, \nu_{2,t}, \ldots, \nu_{K,t}) \sim N(0, \Omega) \tag{A.3}$$

where Ω is a positive definite matrix.

The popularity shock in period t for state k is governed by $\nu_{k,t}$. The shock vector is distributed according to a multivariate normal distribution with mean vector as the null vector and variance covariance matrix as the positive definite matrix Ω, i.e. $(\nu_{1,t}, \nu_{2,t}, \ldots, \nu_{K,t}) \sim N(0, \Omega)$

Let the density of popularity in period $t+1$ given period t primitives be denoted by $f(p_{t+1}|a_{R,t}, a_{D,t}, p_t)$. Here $p_s = (p_{1,s}, p_{2,s}, \ldots, p_{K,s})$ and $a_{it} = (a_{i,1t}, a_{i,2t}, \ldots, a_{i,Kt})$ for $i \in \{R, D\}$. Let $\delta = (\delta_1, \delta_2, \ldots, \delta_K)$ then by assumption A.1 this density is given by:

$$\begin{aligned} &f(p_{t+1}|a_{R,t}, a_{D,t}, p_t) = \\ &\quad \frac{1}{\sqrt{2\pi\,|\Omega|}} \exp\left[-\frac{1}{2}\left(p_{t+1} - \alpha_R a_{Rt} - \alpha_D a_{Dt} - \rho p_t - \delta\right)' \Omega^{-1} \left(p_{t+1} - \alpha_R a_{Rt} - \alpha_D a_{Dt} - \rho p_t - \delta\right)\right] \end{aligned} \tag{A.4}$$

Here $|.|$ denotes the determinant operator and x' denotes the transpose of x.

Electoral Pay-offs: It follows the same definition as in equation 2.4.

Second Mover Problem: The bell-man equation is quite different here. The option-specific value function is defined as followed:

$$u_{is,t}(k;l,h^t,\xi_{jft},\xi_{ist}) = -c_i \times \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[V_{i,t+1}(h^{t+1})\Big|\, h^t, f_t = j, \xi_{jf,t}, a_{j,f,t} = l, \xi_{is,t}, a_{i,s,t} = k\right]$$
$$\text{Where} \quad h^{t+1} = (h^t, f_t = j, \xi_{jf,t}, a_{j,f,t} = l, \xi_{is,t}, a_{i,s,t} = k, \nu_{t+1}, p_{t+1}) \tag{A.5}$$

Here, ξ_{jft}, ξ_{ist}, ν_{t+1} can determine the continuation value. The above function needs to be well defined, that is it takes finite values in $\mathbb{R}$ for all possible values of $(k,l,h^t,\xi_{jft},\xi_{ist})$. The bellman equation for the second mover is defined as followed:

$$V_{is,t}(l,h^t,\xi_{jft},\xi_{ist},\epsilon_{ist}) = \max_{k\in\{0,1,\ldots,K\}}\left\{u_{is,t}(k;l,h^t,\xi_{jft},\xi_{ist}) - \epsilon_{istk} - \xi_{istk}\right\} \tag{A.6}$$

Note that the term $u_{is,t}(k;l,h^t,\xi_{jft},\xi_{ist}) - \epsilon_{istk} - \xi_{istk}$ is not additively separable in ξ_{ist}. It is only separable in the idiosyncratic shocks ϵ_{istk}.

First Mover Problem: The bell-man equation is quite different here. The option-specific value function is defined as followed:

$$\begin{aligned} u_{if,t}(k;h^t,\xi_{ift}) = &-c_i \times \mathbb{1}\{k \neq 0\} \\ &+ \beta\mathbb{E}\left[\sum_{l=0}^{K}\left\{V_{i,t+1}\left(h^t, f_t = i, \xi_{if,t}, a_{i,f,t} = k, \xi_{js,t}, a_{j,s,t} = l, \nu_{t+1}, p_{t+1}\right) \times \mathbb{1}\left\{a_{jst} = l\right\}\right\}\Big|\, h^t, f_t = i, \xi_{ift}\right] \end{aligned} \tag{A.7}$$

Here, ξ_{ift} can determine the continuation value. The above function needs to be well defined, that it takes finite values in $\mathbb{R}$ for all possible values of (k,h^t,ξ_{ift}). The bellman equation for the second mover is defined as followed:

$$V_{if,t}(h^t,\xi_{ift},\epsilon_{ift}) = \max_{k\in\{0,1,\ldots,K\}}\left\{u_{is,t}(k;h^t,\xi_{ift}) - \epsilon_{iftk} - \xi_{iftk}\right\} \tag{A.8}$$

Note that the term $u_{is,t}(k;h^t,\xi_{ift}) - \epsilon_{iftk} - \xi_{iftk}$ is not additively separable in ξ_{ift}. It is only separable in the idiosyncratic shocks ϵ_{iftk}.

Examples of ξ_{imt}: Here I provide two examples of ξ_{imt} that are also relevant to the setting. First allows dependence of cost of a rally on previous period location in a spatial manner.

$$\xi_{imtk} = \begin{cases} \frac{\bar{c}}{D_{a_{it-1},k}} \text{ with probability 1} & a_{it-1} \neq k\ ,\ k \neq 0\ \&\ a_{it-1} \neq 0 \\ 0 \text{ with probability 1} & \text{Otherwise} \end{cases} \tag{A.9}$$

Here $D_{a_{it-1},k}$ is the distance between last rally location of i and the current potential rally location k. Here, the option of not rallying is costless and if there was no rally in the previous period then that has no carryover cost in the current period. The idea here is to test whether there is some traveling inertia in candidate's movements from one location to another.

A second case is to allow costs to have increasing marginal costs within a specified time limit. The time limit could be a group of periods. Let this time limit be $\bar{t}$, and then define d and l such that $t = \bar{t}(d-1) + l$ for $l =\in \{1,2,\ldots,\bar{t}\}$. Change the indices of all variables using this notation and consider the following cost structure:

$$\xi_{im,\bar{t}(t-1)+l,k} = \begin{cases} \bar{c}_1\left(\sum_{j=1}^{l-1}\mathbb{1}\left\{a_{if,\bar{t}(t-1)+j} \neq 0\right\}\right) + \bar{c}_2\left(\sum_{j=1}^{l-1}\mathbb{1}\{a_{if,\bar{t}(t-1)+j} \neq 0\}\right)^2 \text{ with probability 1} & l \in \{2,3,\ldots,\bar{t}\} \\ 0 \text{ with probability 1} & l = 1 \end{cases} \tag{A.10}$$

Here $\bar{c}_1$ and $\bar{c}_2$ back out the cost function of rallies if marginal costs for rallies are increasing.

Note in both these case I have treated ξ_{imt} as deterministic functions. They can also be random. One can allow $\xi_{imt} \sim N(0,\Omega_\xi)$ as well. As long as these costs, deterministic or random, have a finite second moment a unique equilibrium can be supported. I will provide these assumptions, proposition and the proof in the next subsection.

A.1.2 Assumptions and Equilibrium

I assume nature's draw of first and second mover is independent across all histories. Moreover these draws are orthogonal to any other random variable in the model.

Assumption A.2 (Independent First Mover Draws) *First mover draw in any period t is independent of cost and popularity shocks in any period. Moreover it is also independent of first mover draws in other periods.*

For a unique equilibrium it is absolutely necessary that the idiosyncratic costs shock produce convolutions that are continuous random variables. If the convolutions are not continuous then there can be mass-points that can lead to indifference with positive probabilities and therefore multiplicity will arise.

If these idiosyncratic costs shocks are correlated with future idiosyncratic costs shocks then there can be certain types of correlations that can also produce indifference. Here— conditional on ξ_{imt}, h^t, $\xi_{jm't}$ and a_{jft}[51]— the continuation values are also random variables. The conditional convolutions of these continuation values with the idiosyncratic costs shocks can also produce mass points.

In order to avoid these possibilities I assume that idiosyncratic costs shocks are independent across all histories and actions and they are absolutely continuous random variables.

Assumption A.3 (Distribution of Idiosyncratic Cost Shocks) *The random vectors ϵ_{ift} and ϵ_{ist} are absolutely continuous random vectors with respect to Lebesgue measure. In addition to this, they are independently across all histories and actions.*

I need second moments of the cost shocks to be bounded. This is required to ensure that the second mover and the first mover's problems are well-defined and the E max operators produce real numbers. This is a regularity condition which is satisfied by many distributions used in practice.

Assumption A.4 (Cost Shocks Regularity) *The cost shocks, $\epsilon_{i,s,t,k}$ have finite conditional 2nd order moments. That is there exists $\bar{C}$ such that*

$$\begin{aligned} \mathbb{E}\left[\epsilon^2_{i,f,t,k}\right], \mathbb{E}\left[\xi^2_{i,f,t,k}\middle| h^t\right] < \bar{C} < \infty \\ \mathbb{E}\left[\epsilon^2_{i,s,t,k}\right], \mathbb{E}\left[\xi^2_{i,s,t,k}\middle| h^t, a_{jft}, \xi_{jft}\right] < \bar{C} < \infty \end{aligned} \tag{A.11}$$

Below I state and prove the proposition for the general model.

Proposition A.1 (Equilibrium Under Weakest Assumptions) *Given Assumptions A.2, A.3 and A.4, and V_{T+1} defined by equation 2.4. Then an equilibrium exists, it is essentially unique and is characterized as:*

1. *First mover, i's, equilibrium chosen action given $h^t t$ and the cost shocks $\epsilon_{ift0}, \ldots, \epsilon_{iftK}$ is unique with probability 1 and is given by:*
$$a^*_{ift}\left(h^t, \xi_{ift}, \epsilon_{ift}\right) = \arg\max_{k=0,1,\ldots,K} \left\{u_{if,t}(k;h^t) - \epsilon_{i,f,t,k} - \xi_{i,f,t,k}\right\} \tag{A.12}$$

2. *Second mover, i's, equilibrium chosen action given h^t, a_{jft}, and the cost shocks $\epsilon_{ist0}, \ldots, \epsilon_{istK}$ is unique with probability 1 and is given by:*
$$a^*_{ist}\left(h^t, a_{jft}, \xi_{ift}, \xi_{ist}, \epsilon_{ist}\right) = \arg\max_{k=0,1,\ldots,K} \left\{u_{is,t}(k;h^t, a_{jft}, \xi_{ift}, \xi_{ist}) - \epsilon_{i,s,t,k} - \xi_{i,s,t,k}\right\} \tag{A.13}$$

The option specific value function, $u_{if,t}(k;h^t)$, for i when he is the first mover at period t at popularity level p_t satisfies the following:

$$\begin{aligned} u_{if,t}(k;h^t,\xi_{ift}) =& -c_i \times \mathbb{1}\{k \neq 0\} \\ &+ \beta \mathbb{E}\left[\sum_{l=0}^{K} \left\{V_{i,t+1}\left(h^t, f_t = i, \xi_{if,t}, a_{i,f,t} = k, \xi_{js,t}, a_{j,s,t} = l, \nu_{t+1}, p_{t+1}\right) \times \mathbb{1}\left\{a_{jst} = l\right\}\right\} \middle| h^t, f_t = i, \xi_{ift}\right] \end{aligned} \tag{A.14}$$

[51] If second mover is considered

The option specific value function, $u_{is,t}(k;h^t)$, for i when he is the second mover at period t at popularity level p_t and the first mover chose l satisfies the following eq. 2.14:

$$u_{is,t}(k;l,h^t,\xi_{jft},\xi_{ist}) = -c_i \times \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[V_{i,t+1}(h^{t+1})\middle| h^t, f_t = j, \xi_{jf,t}, a_{j,f,t} = l, \xi_{is,t}, a_{i,s,t} = k\right]$$
$$\textit{Where} \quad h^{t+1} = (h^t, f_t = j, \xi_{jf,t}, a_{j,f,t} = l, \xi_{is,t}, a_{i,s,t} = k, \nu_{t+1}, p_{t+1}) \tag{A.15}$$

Proof: This proposition is proven by using the Lemmas A.1, A.2, A.3, A.4, A.5, A.6 and A.7. First A.1 and A.3 show first and second mover option specific value functions are bounded if the next period value function is bounded. Then Lemmas A.2 and A.4 show that the maximization problems of the first and the second mover are well defined and have a unique maximizer with probability 1 if the first and second mover option specific value functions are bounded themselves.

Then Lemmas A.5 and A.6 show that the expected value function of a first mover and the second mover is bounded. Finally A.7 will show that if the next period value function is bounded then the current value function is also bounded by invoking lemmas A.1, A.3, A.5, and A.6. This final Lemma A.7 bridges connection between two periods allows for induction to prove the whole proposition.

Then finally these Lemmas together can prove the Proposition A.1 if V_{iT+1} is bounded. This holds true by definition since consider V_{it} which is given by:

$$V_{i,t}(h^t) = V_{iT+1}(p_{T+1}) = \sum_{k=0} e_k E\mathbb{1}\{p_{kT+1} > 0\}$$
$$\Rightarrow \infty < -E < V_{i,t}(h^t) = V_{iT+1}(p_{T+1}) < E < \infty \tag{A.16}$$

Therefore choosing $M_{T+1} = E$ and then repeated applications of these Lemmas for $t = T, T-1, T-2, \ldots, 1$ proves the Proposition.

Q.E.D.

Lemma A.1 *Given Assumption A.2, A.3, and A.4, and suppose the following condition holds for all h^{t+1}:*

$$-\infty < -M_{t+1} \leq V_{it+1}(h^{t+1}) \leq M_{t+1} < \infty \tag{A.17}$$

Then the following holds

$$|u_{is,t}(k; a_{jft}, h^t, \xi_{jft}, \xi_{ist})| \leq c_i \times \mathbb{1}\{k \neq 0\} + \beta M_{t+1} < \infty \tag{A.18}$$

For all h^t.

Proof: Pick any arbitrary $(h^t, f_t, \epsilon_{if,t}, a_{f_t,f,t}, \epsilon_{is,t}, a_{s_t,s,t})$, for all i, s, k, l we have the following holding

$$\begin{aligned} u_{is,t}(k; a_{jft}, h^t, \xi_{jft}, \xi_{ist}) &= -c_i \times \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[V_{i,t+1}(h^{t+1})\middle| h^t, f_t = j, \xi_{jf,t}, a_{j,f,t}, \xi_{is,t}, a_{i,s,t}\right] \\ &\leq -c_i \times \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[M_{t+1}\middle| h^t, f_t = j, \xi_{jf,t}, a_{j,f,t}, \xi_{is,t}, a_{i,s,t}\right] \\ &\leq -c_i \times \mathbb{1}\{k \neq 0\} + \beta M_{t+1} < c_i \times \mathbb{1}\{k \neq 0\} + \beta M_{t+1} < \infty \end{aligned} \tag{A.19}$$

The first inequality holds because $V_{i,t+1}$ is pointwise lower than M_{t+1}. Pick any arbitrary $(h^t, f_t, \xi_{if,t}, a_{f_t,f,t}, \xi_{is,t}, a_{s_t,s,t})$, for all i, s, k, l we have the following holding

$$\begin{aligned} u_{is,t}(k; a_{jft}, h^t, \xi_{jft}, \xi_{ist}) &= -c_i \times \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[V_{i,t+1}(h^{t+1})\middle| h^t, f_t = j, \xi_{jf,t}, a_{j,f,t}, \xi_{is,t}, a_{i,s,t}\right] \\ &\geq -c_i \times \mathbb{1}\{k \neq 0\} - \beta\mathbb{E}\left[M_{t+1}\middle| h^t, f_t = j, \xi_{jf,t}, a_{j,f,t}, \xi_{is,t}, a_{i,s,t}\right] \\ &\geq -c_i \times \mathbb{1}\{k \neq 0\} - \beta M_{t+1} > -\infty \end{aligned} \tag{A.20}$$

The first inequality holds because $V_{i,t+1}$ is pointwise higher than M_{t+1}. Note given this for all $|\xi_{i,s,t}| \in \mathbb{R}^{K+1}$ we have $\xi_{istk} < \infty$ and $|u_{ist}| < \infty$.

Q.E.D.

Lemma A.2 *Given Assumption A.2, A.3, and A.4, and suppose the following condition holds for all* $(k, a_{jft}, h^t, \xi_{jft}, \xi_{ist})$*:*

$$|u_{is,t}(k; a_{jft}, h^t, \xi_{jft}, \xi_{ist})| \leq c_i \times \mathbb{1}\{k \neq 0\} + \beta M_{t+1} < \infty \tag{A.21}$$

Then for all $(a_{jft}, h^t, \xi_{jft}, \xi_{ist})$, $\max_{k\in\{0,1,\ldots,K\}} \left\{u_{i,s,t}(k; l, h^t, \epsilon_{jft}, \epsilon_{ist}) - \epsilon_{ist,k}\right\}$ *is well-defined maximization problem. Moreover, a unique maximizer exists with probability 1.*

Proof: Pick an arbitrary $(a_{jft}, h^t, \xi_{jft}, \xi_{ist})$. Note the following holds for all $k \in \{0, 1, \ldots, K\}$, $|u_{is,t}(k; a_{jft}, h^t, \xi_{jft}, \xi_{ist})| \leq c_i \times \mathbb{1}\{k \neq 0\} + \beta M_{t+1} < \infty$ then we must also have

$$-\infty < u_{i,s,t}(k; l, h^t, \xi_{jft}, \xi_{ist}) - \xi_{ist,k} < \infty \tag{A.22}$$

Since $-\infty < \xi_{ist,k} < \infty$. Therefore, $\max_{k\in\{0,1,\ldots,K\}} \left\{u_{i,s,t}(k; l, h^t, \epsilon_{jft}, \epsilon_{ist}) - \epsilon_{ist,k}\right\}$ is well defined and exists. It is a maximum over finite set of real numbers.

Moreover, note $a^*_{ist}(h^t, a_{jft}, \xi_{ist}, \xi_{jft}, \epsilon_{ist})$ is unique with probability 1 for any $(h^t, a_{jft}, \xi_{ist}, \xi_{jft})$.

Suppose not then there exists $(h^t, a_{jft}, \xi_{ist}, \xi_{jft})$ such that candidate i is indifferent between two actions k and l with positive probability. That is the following is true

$$\begin{aligned}
&\mathbb{P}\Big(u_{ist,k} - \xi_{ist,k} - \epsilon_{ist,k} = u_{ist,l} - \xi_{ist,l} - \epsilon_{ist,l} | h^t, a_{jft}, \xi_{ist}, \xi_{jft}\Big) \\
&\qquad = \mathbb{P}\Big(\epsilon_{ist,k} - \epsilon_{ist,l} = u_{ist,l} - u_{ist,k} + \xi_{ist,k} - \xi_{ist,k} | h^t, a_{jft}, \xi_{ist}, \xi_{jft}\Big) > 0
\end{aligned} \tag{A.23}$$

Note that conditional on $(h^t, a_{jft}, \xi_{ist}, \xi_{jft})$ the term $u_{istl} - u_{istk} + \xi_{istk} - \xi_{istl}$ is deterministic and is equal to a real number. Therefore, the above probability is zero by assumption A.3.

Q.E.D.

Lemma A.3 *Given Assumption A.2, A.3, and A.4, and suppose the following condition holds for all* h^{t+1}*:*

$$-\infty < -M_{t+1} \leq V_{it+1}(h^{t+1}) \leq M_{t+1} < \infty \tag{A.24}$$

Then the following holds

$$|u_{if,t}(k; h^t, \xi_{ift})| \leq c_i \times \mathbb{1}\{k \neq 0\} + \beta M_{t+1} < \infty \tag{A.25}$$

For all h^t.

Proof: Pick an arbitrary (k, h^t, ξ_{ift}), then

$$\begin{aligned}
u_{if,t}(k; h^t, \xi_{ift}) &= -c_i \times \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[\sum_{l=0}^{K} \left\{V_{i,t+1}\left(h^t, f_t = i, \xi_{if,t}, a_{i,f,t} = k, \xi_{js,t}, a_{j,s,t} = l, \nu_{t+1}, p_{t+1}\right) \times \mathbb{1}\left\{a_{jst} = l\right\}\right\} \Big| \, h^t, f_t = i, \xi_{ift}\right] \\
&\leq -c_i \times \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[\sum_{l=0}^{K} M_{t+1} \times \mathbb{1}\left\{a_{jst} = l\right\} \Big| \, h^t, f_t = i, \xi_{ift}\right] \\
&= -c_i \times \mathbb{1}\{k \neq 0\} + \beta M_{t+1} < c_i \times \mathbb{1}\{k \neq 0\} + \beta M_{t+1} < \infty
\end{aligned} \tag{A.26}$$

The first inequality holds because we showed u_{ist} is bounded above by $-c_i \times \mathbb{1}\{k \neq 0\} + \beta M_{t+1}$. Similarly we can show

$$\begin{aligned}
u_{if,t}(k; h^t, \xi_{ift}) &= -c_i \times \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[\sum_{l=0}^{K} \left\{V_{i,t+1}\left(h^t, f_t = i, \xi_{if,t}, a_{i,f,t} = k, \xi_{js,t}, a_{j,s,t} = l, \nu_{t+1}, p_{t+1}\right) \times \mathbb{1}\left\{a_{jst} = l\right\}\right\} \Big| \, h^t, f_t = i, \xi_{ift}\right] \\
&\geq -c_i \times \mathbb{1}\{k \neq 0\} - \beta\mathbb{E}\left[\sum_{l=0}^{K} M_{t+1} \times \mathbb{1}\left\{a_{jst} = l\right\} \Big| \, h^t, f_t = i, \xi_{ift}\right] \\
&= -c_i \times \mathbb{1}\{k \neq 0\} - \beta M_{t+1} > -\infty
\end{aligned} \tag{A.27}$$

Q.E.D.

Lemma A.4 *Given Assumption A.2, A.3, and A.4, and suppose the following condition holds for all* (h^t, ξ_{ift})*:*

$$|u_{if,t}(k; h^t, \xi_{ift})| \le c_i \times \mathbb{1}\{k \ne 0\} + \beta M_{t+1} < \infty \tag{A.28}$$

Then for all $(a_{jft}, h^t, \xi_{jft}, \xi_{ist})$, $\max_{k\in\{0,1,\ldots,K\}} \left\{u_{i,s,t}(k; l, h^t, \epsilon_{jft}, \epsilon_{ist}) - \epsilon_{ist,k}\right\}$ *is well-defined maximization problem. Moreover, a unique maximizer exists with probability 1.*

Proof: Arguments similar to the case of Lemma A.2 proves the statement.
Q.E.D.

Lemma A.5 *Given Assumption A.2, A.3, and A.4, and suppose the following condition holds for all* (h^t, ξ_{ift})*:*

$$|u_{if,t}(k; h^t, \xi_{ift})| \le c_i \times \mathbb{1}\{k \ne 0\} + \beta M_{t+1} < \infty \tag{A.29}$$

Then for all h^t *the following inequality holds*

$$-\beta M_{t+1} - 2\sqrt{\bar{C}} \le \mathbb{E}\left[\max_{k\in\{0,1,\ldots,K\}} \left\{u_{if,t}(k; h^t, \xi_{ift}) - \epsilon_{i,f,t,k} - \xi_{i,f,t,k}\right\} \;\middle|\; h^t\right] \le Kc_i + (K+1)\left(\beta M + 2\sqrt{\bar{C}}\right) \tag{A.30}$$

Proof: Pick an arbitrary h^t and consider the expected pay-off when i is the first mover

$$\begin{aligned}
&\mathbb{E}\left[\max_{k\in\{0,1,\ldots,K\}} \left\{u_{if,t}(k; h^t, \xi_{ift}) - \epsilon_{i,f,t,k} - \xi_{i,f,t,k}\right\} \;\middle|\; h^t\right] \\
&\le \mathbb{E}\left[\max_{k\in\{0,1,\ldots,K\}} \left\{\left|u_{if,t}(k; h^t, \xi_{ift}) - \epsilon_{i,f,t,k} - \xi_{ift}\right|\right\} \;\middle|\; h^t\right] \\
&\because \max\{x_0, x_1, \ldots, x_K\} \le \max\{|x_0|, |x_1|, \ldots, |x_K|\} \\
&\le \mathbb{E}\left[\sum_{k\in\{0,1,\ldots,K\}} \left|u_{if,t}(k; h^t, \xi_{ift}) - \epsilon_{i,f,t,k} - \xi_{ift}\right| \;\middle|\; h^t\right] \\
&\because \max\{|x_0|, |x_1|, \ldots, |x_K|\} \le |x_0| + |x_1| + \cdots + |x_K| \\
&\le \mathbb{E}\left[\sum_{k\in\{0,1,\ldots,K\}} \left|u_{if,t}(k; h^t, \xi_{ift})\right| + \left|\epsilon_{i,f,t,k}\right| + \left|\xi_{ift}\right| \;\middle|\; h^t\right] \\
&\because \text{ Triangle Inequality} \\
&\le \sum_{k\in\{0,1,\ldots,K\}} \left(\mathbb{E}\left[\left|u_{if,t}(k; h^t, \xi_{ift})\right| \;\middle|\; h^t\right] + \mathbb{E}\left[\left|\epsilon_{i,f,t,k}\right| \;\middle|\; h^t\right] + \mathbb{E}\left[\left|\xi_{i,f,t,k}\right| \;\middle|\; h^t\right]\right) \\
&\because \text{ Since expectation is a linear operator} \\
&\le \sum_{k\in\{0,1,\ldots,K\}} \left(\mathbb{E}\left[\left|u_{if,t}(k; h^t, \xi_{ift})\right| \;\middle|\; h^t\right] + \sqrt{\mathbb{E}\left[\epsilon^2_{i,f,t,k} \;\middle|\; h^t\right]} + \sqrt{\mathbb{E}\left[\xi^2_{i,f,t,k} \;\middle|\; h^t\right]}\right) \\
&\because \mathbb{E}\left[\epsilon^2_{i,f,t,k} \;\middle|\; h^t\right] = \mathbb{E}\left[\epsilon^2_{i,f,t,k}\right] \text{ and Hölder Inequality} \\
&\le Kc_i + (K+1)\left(\beta M_{t+1} + 2\sqrt{\bar{C}}\right) \\
&\because \text{ Substituting the bounds for } u_{ift} \text{ and } \mathbb{E}\left[\xi^2_{i,f,t,k} \;\middle|\; h^t\right], \mathbb{E}\left[\xi^2_{i,f,t,k} \;\middle|\; h^t\right]
\end{aligned}$$

For the lower bound of $\mathbb{E}\Big[\max_{k\in\{0,1,\ldots,K\}}\Big\{u_{if,t}(k;h^t)-\epsilon_{i,f,t,k}-\xi_{i,f,t,k}\Big\}\Big|\,h^t\Big]$ consider the following set of arguments:

$$
\begin{aligned}
&\mathbb{E}\Big[\max_{k\in\{0,1,\ldots,K\}}\Big\{u_{if,t}(k;h^t,\xi_{ift})-\epsilon_{i,f,t,k}-\xi_{i,f,t,k}\Big\}\Big|\,h^t\Big]\\
&\qquad\geq \mathbb{E}\Big[u_{if,t}(0;h^t,\xi_{ift})-\epsilon_{i,f,t,0}-\xi_{i,f,t,0}\Big|\,h^t\Big]\\
&\qquad\because \max\{x_0,x_1,\ldots,x_K\}\geq x_0\\
&\qquad\geq \mathbb{E}\Big[u_{if,t}(0;h^t,\xi_{ift})\Big]-\mathbb{E}\Big[\epsilon_{i,f,t,k}\Big|\,h^t\Big]-\mathbb{E}\Big[\xi_{i,f,t,k}\Big|\,h^t\Big]\\
&\qquad\because \mathbb{E}\text{ is a linear operator}\\
&\qquad\geq -\beta M_{t+1}-\mathbb{E}\Big[\big|\epsilon_{i,f,t,k}\big|\Big|\,h^t\Big]-\mathbb{E}\Big[\big|\xi_{i,f,t,k}\big|\Big|\,h^t\Big]\\
&\qquad\because \mathbb{E}\left[x\right]\leq \mathbb{E}\left[|x|\right]\\
&\qquad\geq -\beta M_{t+1}-\sqrt{\mathbb{E}\Big[\big|\epsilon_{i,f,t,k}\big|^2\Big|\,h^t\Big]}-\sqrt{\mathbb{E}\Big[\big|\xi_{i,f,t,k}\big|^2\Big|\,h^t\Big]}\\
&\qquad\because \text{Hölder Inequality}\\
&\qquad\geq -\beta M_{t+1}-2\sqrt{\bar{C}}\\
&\qquad\because \text{Substituting the bounds for }\mathbb{E}_{\epsilon_{i,f,t}}\Big[\epsilon^2_{i,f,t,k}\Big|\,h^t\Big]
\end{aligned}
$$

Therefore this gives us that

$$
-\beta M_{t+1}-2\sqrt{\bar{C}}\leq \mathbb{E}\Big[\max_{k\in\{0,1,\ldots,K\}}\Big\{u_{if,t}(k;h^t,\xi_{ift})-\epsilon_{i,f,t,k}-\xi_{i,f,t,k}\Big\}\Big|\,h^t\Big]\leq Kc_i+(K+1)\Big(\beta M+2\sqrt{\bar{C}}\Big) \tag{A.31}
$$

Q.E.D.

Lemma A.6 *Given Assumption A.2, A.3, and A.4, and suppose the following condition holds for all* $(k,a_{jft},h^t,\xi_{jft},\xi_{ist})$*:*

$$
|u_{is,t}(k;a_{jft},h^t,\xi_{jft},\xi_{ist})|\leq c_i\times \mathbb{1}\{k\neq 0\}+\beta M_{t+1}<\infty \tag{A.32}
$$

Then for all (h^t,a_{jft},ξ_{jft}), $\max_{k\in\{0,1,\ldots,K\}}\Big\{u_{i,s,t}(k;l,h^t,\epsilon_{jft},\epsilon_{ist})-\epsilon_{ist,k}\Big\}$ *is well-defined maximization problem. Moreover, a unique maximizer exists with probability 1.*

$$
-\beta M_{t+1}-2\sqrt{\bar{C}}\leq \mathbb{E}\Big[\max_{l\in\{0,1,\ldots,K\}}\Big\{u_{is,t}(l;a_{jft},h^t,\xi_{jft},\xi_{ist})-\epsilon_{i,s,t,l}-\xi_{i,s,t,l}\Big\}\Big|\,h^t,a_{jft},\xi_{jft}\Big]\leq Kc_i+(K+1)\Big(\beta M+2\sqrt{\bar{C}}\Big) \tag{A.33}
$$

Proof: Similar set of arguments used in the proof of Lemma A.5 prove this lemma.

Q.E.D.

Lemma A.7 *Given Assumption A.2, A.3, and A.4, and suppose the following condition holds for all* h^{t+1}*:*

$$
-\infty<-M_{t+1}\leq V_{it+1}(h^{t+1})\leq M_{t+1}<\infty \tag{A.34}
$$

Then there exists $M_t\in\mathbb{R}_+$ *such that*

$$
-\infty<-M_t\leq V_{it}(h^t)\leq c_i\times \mathbb{1}\{k\neq 0\}+\beta M_{t+1}<\infty \tag{A.35}
$$

Where V_{it} *is period* t *value function and is given by:*

$$
\begin{aligned}
V_{i,t}(h^t)| =& f_i\times\mathbb{E}\Big[\max_{k\in\{0,1,\ldots,K\}}\Big\{u_{if,t}(k;h^t,\xi_{ift})-\epsilon_{i,f,t,k}-\xi_{i,f,t,k}\Big\}\Big|\,h^t\Big]\\
&+(1-f_i)\times\sum_{k=0}^{K}\mathbb{E}\Big[\max_{l\in\{0,1,\ldots,K\}}\Big\{u_{is,t}(l;k,h^t,\xi_{jft},\xi_{ist})-\epsilon_{i,s,t,l}-\xi_{i,s,t,k}\Big\}\times\mathbb{1}\{a_{jft}=k\}\Big|\,h^t\Big]
\end{aligned}\tag{A.36}
$$

For all h^t.

Proof: Pick an arbitrary h^t and note that $|u_{ift}(k,h^t,\xi_{ift})|$ is bounded by $c_i \times \mathbb{1}\{k \neq 0\} + \beta M_{t+1}$ by Lemma A.3 for all (k,ξ_{ift}). Therefore by Lemma A.5 we have the following:

$$-\beta M_{t+1} - 2\sqrt{\bar{C}} \le \mathbb{E}\left[\max_{k\in\{0,1,\dots,K\}}\left\{u_{if,t}(k;h^t,\xi_{ift}) - \epsilon_{i,f,t,k} - \xi_{i,f,t,k}\right\}\ \middle|\ h^t\right] \le Kc_i + (K+1)\left(\beta M + 2\sqrt{\bar{C}}\right) \tag{A.37}$$

Also note for any $(k, a_{jft}, \xi_{jft}, \xi_{ist})$[52] the following holds $|u_{is,t}(k;a_{jft},h^t,\xi_{jft},\xi_{ist})| \le c_i \times \mathbb{1}\{k \neq 0\} + \beta M_{t+1} < \infty$ by Lemma A.2. Therefore by Lemma A.6 we have the following holding true:

$$-\beta M_{t+1} - 2\sqrt{\bar{C}} \le \mathbb{E}\left[\max_{l\in\{0,1,\dots,K\}}\left\{u_{is,t}(l;a_{jft},h^t,\xi_{jft},\xi_{ist}) - \epsilon_{i,s,t,l} - \xi_{i,s,t,l}\right\}\ \middle|\ h^t, a_{jft}, \xi_{jft}\right] \le Kc_i + (K+1)\left(\beta M + 2\sqrt{\bar{C}}\right) \tag{A.38}$$

Note that $Kc_i + (K+1)\left(\beta M + 2\sqrt{\bar{C}}\right) > \beta M_{t+1} + 2\sqrt{\bar{C}}$. Let $M_t = \max_{i\in\{R,D\}}\left\{Kc_i + (K+1)\left(\beta M + \sqrt{\bar{C}}\right)\right\}$. This bounds both conditional expectations. Now consider $V_{it}(h^t)$ given by:

$$\begin{aligned}
|V_{i,t}(h^t)| =& \Bigg| f_i \times \mathbb{E}\left[\max_{k\in\{0,1,\dots,K\}}\left\{u_{if,t}(k;h^t,\xi_{ift}) - \epsilon_{i,f,t,k} - \xi_{i,f,t,k}\right\}\ \middle|\ h^t\right] \\
&+ (1-f_i) \times \sum_{k=0}^{K} \mathbb{E}\left[\max_{l\in\{0,1,\dots,K\}}\left\{u_{is,t}(l;k,h^t,\xi_{jft},\xi_{ist}) - \epsilon_{i,s,t,l} - \xi_{i,s,t,k}\right\} \times \mathbb{1}\{a_{jft}=k\}\ \middle|\ h^t\right]\Bigg| \\
=& \Bigg| f_i \times \mathbb{E}\left[\max_{k\in\{0,1,\dots,K\}}\left\{u_{if,t}(k;h^t,\xi_{ift}) - \epsilon_{i,f,t,k} - \xi_{i,f,t,k}\right\}\ \middle|\ h^t\right] \\
&+ (1-f_i) \times \sum_{k=0}^{K} \mathbb{E}\left[\mathbb{E}\left[\max_{l\in\{0,1,\dots,K\}}\left\{u_{is,t}(l;k,h^t,\xi_{jft},\xi_{ist}) - \epsilon_{i,s,t,l} - \xi_{i,s,t,k}\right\}\ \middle|\ h^t, a_{jft}=k, \epsilon_{jft}\right] \times \mathbb{1}\{a_{jft}=k\}\ \middle|\ h^t\right]\Bigg| \\
&\because \text{Iterative Law of Expectations} \\
\le& f_i \times \left|\mathbb{E}\left[\max_{k\in\{0,1,\dots,K\}}\left\{u_{if,t}(k;h^t,\xi_{ift}) - \epsilon_{i,f,t,k} - \xi_{i,f,t,k}\right\}\ \middle|\ h^t\right]\right| \\
&+ (1-f_i) \times \sum_{k=0}^{K} \mathbb{E}\left[\left|\mathbb{E}\left[\max_{l\in\{0,1,\dots,K\}}\left\{u_{is,t}(l;k,h^t,\xi_{jft},\xi_{ist}) - \epsilon_{i,s,t,l} - \xi_{i,s,t,k}\right\}\ \middle|\ h^t, a_{jft}=k, \xi_{jft}\right]\right| \times \mathbb{1}\{a_{jft}=k\}\ \middle|\ h^t\right] \\
&\because \text{Triangle Inequality followed by Cauchy-Schwarz Inequality for the second term.} \\
\le& f_i \max_{i\in\{R,D\}}\left\{Kc_i + (K+1)\left(\beta M_{t+1} + 2\sqrt{\bar{C}}\right)\right\} \\
&+ (1-f_i) \times \sum_{k=0}^{K} \mathbb{E}\left[\max_{i\in\{R,D\}}\left\{Kc_i + (K+1)\left(\beta M + 2\sqrt{\bar{C}}\right)\right\} \times \mathbb{1}\{a_{jft}=k\}\ \middle|\ h^t\right] \\
&\because \text{ Due to bounds obtained in inequalities A.38 and A.31} \\
=& f_i \max_{i\in\{R,D\}}\left\{Kc_i + (K+1)\left(\beta M_{t+1} + 2\sqrt{\bar{C}}\right)\right\} + (1-f_i) \times \sum_{k=0}^{K}\left(\max_{i\in\{R,D\}}\left\{Kc_i + (K+1)\left(\beta M_{t+1} + 2\sqrt{\bar{C}}\right)\right\} \times \sigma_{jf,t}(k;h^t)\right) \\
&\because \max_{i\in\{R,D\}}\left\{Kc_i + (K+1)\left(\beta M_{t+1} + 2\sqrt{\bar{C}}\right)\right\} \text{ is a deterministic constant and } \mathbb{E}[\mathbb{1}\{a_{jft}=k,h^t\}|h^t] = \sigma_{jst}(k;h^t) \\
=& \max_{i\in\{R,D\}}\left\{Kc_i + (K+1)\left(\beta M_{t+1} + 2\sqrt{\bar{C}}\right)\right\} = M_t < \infty
\end{aligned}$$

Q.E.D.

[52] h^t has been picked here.

A.2 Proof for Proposition 2.1

In this section I will prove proposition 2.1, however first I will restate the proposition with more details:

Proposition A.2 (Value Functions and CCPs) *Given assumptions A.2, 2.1 (or A.1) ,2.2, 2.3, and eq. 2.4, which defines electoral pay-off $V_{i,T+1}$, the following holds for all $t = 1, 2, ..., T$*

1. *The value function $V_{i,t}$ takes the following functional form:*

$$V_{i,t}(p_t) = f_i \times \ln\left(\sum_{k=0}^{K} \exp\left\{u_{if,t}(k;p_t)\right\}\right) + (1-f_i) \times \sum_{k=0}^{K}\left[\sigma_{jf,t}(k;p_t)\ln\left(\sum_{l=0}^{K}\exp\left\{u_{is,t}(l;k,p_t))\right\}\right)\right] + \gamma \tag{A.39}$$

2. *First mover, i's, equilibrium chosen action given p_t and the cost shocks $\epsilon_{ift0}, \ldots, \epsilon_{iftK}$ is unique with probability 1 and is given by:*

$$a^*_{ift}\left(p_t, \epsilon_{ist0}, \ldots, \epsilon_{istK}\right) = \arg\max_{k=0,1,\ldots,K}\left\{u_{ift}(k;p_t) + \epsilon_{iftk}\right\} \tag{A.40}$$

3. *The expected probability of i choosing action k as the first mover is given by:*

$$\sigma_{if,t}(k;p_t) = P\Big(k = a^*_{if,t}(p_t, \epsilon_{if,t})\Big) = \frac{\exp\left(u_{if,t}(k;p_t) - u_{if,t}(0;p_t)\right)}{1 + \sum_{l=1}^{K}\exp\left(u_{if,t}(l;p_t) - u_{if,t}(0;p_t)\right)} \tag{A.41}$$

4. *Second mover, i's, equilibrium chosen action given a_{jft}, p_t and the cost shocks $\epsilon_{ist0}, \ldots, \epsilon_{istK}$ is unique with probability 1 and is given by:*

$$a^*_{ist}\left(a_{jft}, p_t, \epsilon_{ist0}, \ldots, \epsilon_{istK}\right) = \arg\max_{k=0,1,\ldots,K}\left\{u_{ist}(k;a_{jft},p_t) + \epsilon_{istk}\right\} \tag{A.42}$$

5. *The probability of i choosing action k as the second mover is given by:*

$$\sigma_{is,t}(k;l,p_t) = P\Big(k = a^*_{is,t}(a_{jft} = l, p_t, \epsilon_{is,t})\Big) = \frac{\exp\left(u_{is,t}(k;l,p_t) - u_{is,t}(0;l,p_t)\right)}{1 + \sum_{q=1}^{K}\exp\left(u_{is,t}(q;l,p_t) - u_{is,t}(0;l,p_t)\right)} \tag{A.43}$$

   Where, the option specific value function, $u_{if,t}(k;p_t)$, for i when he is the first mover at period t at popularity level p_t satisfies the following:

$$u_{if,t}(k;p_t) = \sum_{l=0}^{K} u_{is,t}(k;l,p_t) \times \sigma_{js,t}(l;k,p_t) \tag{A.44}$$

   The option specific value function, $u_{is,t}(k;p_t)$, for i when he is the second mover at period t at popularity level p_t and the first mover chose l satisfies the following eq. 2.14:

$$u_{is,t}(k;l,p_t) = -c_i \times \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[V_{i,t+1}(p)\Big|\, a_{it} = k, a_{jt} = l, p' = p_t\right] \tag{A.45}$$

Proof: I use Proposition A.1 to prove the extended version of Proposition 2.1. First note under the Assumption 2.2 and the Assumption 2.3, cost shocks satisfy the Assumption A.3. This is because ϵ_{imtk} and ϵ_{imtl} are two independently drawn T1EV random variables and therefore $\epsilon_{imtk} - \epsilon_{imtl}$ is a Logistic random variable which is also a continuously distributed random variable. In the baseline set up $P\left(\xi_{iftk} = 0|h^t\right) = 1$ and $P\left(\xi_{istk} = 0|h^t, \xi_{jft}, a_{jft}\right) = 1$. Which satisfies Assumption A.4.

Moreover, since $\epsilon_{imtk} \sim T1EV$ and due to Assumption 2.2 we have the following holding:

$$\begin{aligned}\mathbb{E}\left[\epsilon^2_{i,f,t,k}\middle|h^t\right] &= \mathbb{E}\left[\epsilon^2_{i,f,t,k}\right] = \gamma^2 + \frac{\pi^2}{6} < \bar{C} = \gamma^2 + \frac{\pi^2}{6} + 1 < \infty \\ \mathbb{E}\left[\epsilon^2_{i,s,t,k}\middle|h^t, a_{jft}, \epsilon_{jft}\right] &= \mathbb{E}\left[\epsilon^2_{i,s,t,k}\right] = \gamma^2 + \frac{\pi^2}{6} < \bar{C} = \gamma^2 + \frac{\pi^2}{6} + 1 < \infty\end{aligned} \tag{A.46}$$

Here $\gamma = -\ln(\ln(2))$ is the Euler–Mascheroni constant. Therefore Assumptions A.4 and A.3 hold here. Proposition A.1 can be applied here. Therefore, equilibrium exists, it is *essentially unique* (multiplicity takes place with probability zero) and equilibrium is characterized in the same fashion as in Proposition A.1.

Also note that under Assumptions 2.1 or A.1 there is no intertemporal dependence between ν_s and ν_t for any $t \neq s$. Therefore past shocks $\nu_{t-1}, \nu_{t-2}, \ldots \nu_1$ are not relevant state variables in period t. Also note under Assumption 2.2 past cost shocks are not relevant either for any player or any mover. Under Assumption A.2 past mover order is also not a relevant state variable. Therefore for the first mover the relevant state variables are p_t and ϵ_{ift}. Since $\xi_{imtk} = 0$ a.s. these can be dropped from set of relevant state variables. For the second mover the relevant state variables are p_t, a_{jft} and ϵ_{ist}.

Now I will further simplify the proposition as followed. I start with second mover option specific value function, u_{ist}.

$$u_{is,t}(k;l,h^t,\xi_{jft}=0,\xi_{ist}=0) = -c_i \times \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[V_{i,t+1}(h^{t+1})\middle|\, h^t, f_t = j, \xi_{jf,t}=0, a_{j,f,t}=l, \xi_{is,t}=0, a_{i,s,t}=k\right]$$
$$\Rightarrow u_{is,t}(k;l,h^t,\xi_{jft}=0,\xi_{ist}=0) = u_{is,t}(k;l,p_t) = -c_i \times \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[V_{i,t+1}(p_{t+1})\middle|\, p_t, a_{j,f,t}=l, a_{i,s,t}=k\right]$$
$$\text{where } p_{t+1} \text{ is defined in Equation 2.1 component wise} \tag{A.47}$$

The first equality holds because only p_{t+1} is a relevant state variable for $V_{i,t+1}$. Given this note a^*_{ist} is defined as

$$a^*_{ist}\left(h^t, a_{jft}, \xi_{jft}=0, \xi_{ist}=0, \epsilon_{ist}\right) = \arg\max_{k=0,1,\ldots,K}\left\{u_{if,t}(k;h^t,\xi_{jft},a_{jft}) - \epsilon_{i,s,t,k} - 0\right\}$$
$$\Rightarrow a^*_{ist}\left(h^t, a_{jft}, \xi_{jft}=0, \xi_{ist}=0, \epsilon_{ist}\right) = a^*_{ist}\left(p_t, a_{jft}, \epsilon_{ist}\right) = \arg\max_{k=0,1,\ldots,K}\left\{u_{if,t}(k;a_{jft},p_t) - \epsilon_{i,s,t,k}\right\} \tag{A.48}$$

Given this, by Lemma 1 from McFadden (1973) we can see that the following holds:

$$\sigma_{ist}(k;l,p_t) = P\left(a^*_{ist}(a_{jft},p_t,\epsilon_{ist}) = k \middle| a_{jft}, p_t\right) = \frac{\exp\left(u_{is,t}(k;a_{jft},p_t) - u_{is,t}(0;a_{jft},p_t)\right)}{1 + \sum_{q=1}^{K}\exp\left(u_{is,T}(q;a_{jft},p_t) - u_{is,t}(0;a_{jft},p_t)\right)} \tag{A.49}$$

Moreover, note if $-\epsilon_k \sim T1EV$ and independent then $\max_{k\in\{0,1,\ldots,K\}}\{\tilde{\delta}_k - \epsilon_{iftk}\} \sim \text{Gumbel}(\mu = \ln\sum_k \tilde{\delta}_k,\ \beta = 1)$. Where μ denotes the location parameter of a Gumbel Distribution and β denotes the scale parameter. Therefore we must have that the following holds:

$$\mathbb{E}\left[\max_{l\in\{0,1,\ldots,K\}}\left\{u_{is,t}(l;a_{jft},p_t) - \epsilon_{i,s,t,l}\right\}\,\middle|\, p_t, a_{jft}=k\right] = \log\left(\sum_{l=0}^{K}\exp\left\{u_{ist}(l;k,p_t)\right\}\right) + \gamma \tag{A.50}$$

Here note p_t is the relevant state variable replacing h^t.

Consider the following set of arguments to attain a simplification of u_{ift}.

$$u_{if,t}(k;h^t,\xi_{ift}=0) = -c_i \times \mathbb{1}\{k \neq 0\}$$
$$+\beta\mathbb{E}\left[\sum_{l=0}^{K}\left\{V_{i,t+1}\left(h^t, f_t=i, \xi_{if,t}, a_{i,f,t}=k, \xi_{js,t}, a_{j,s,t}=l, \nu_{t+1}, p_{t+1}\right)\times \mathbb{1}\left\{a_{jst}=l\right\}\right\}\Big|\, h^t, f_t=i, \xi_{ift}=0, a_{ift}=k\right]$$

By substituting for relevant state variables we obtain

$$\Rightarrow u_{if,t}(k;p) = -c_i \times \mathbb{1}\{k\neq 0\} + \beta\mathbb{E}\left[\sum_{l=0}^{K}\left\{V_{i,t+1}\left(p_{t+1}\right)\times \mathbb{1}\left\{a_{jst}=l\right\}\right\}\Big|\, p_t=p, a_{ift}=k\right]$$

Note that $c_i \times \mathbb{1}\{k\neq 0\}$ is deterministic term, therefore

$$\Rightarrow u_{if,t}(k;p) = \mathbb{E}\left[\sum_{l=0}^{K}\left\{-c_i \times \mathbb{1}\{k\neq 0\} + \beta V_{i,t+1}\left(p_{t+1}\right)\times \mathbb{1}\left\{a_{jst}=l\right\}\right\}\Big|\, p_t=p, a_{ift}=k\right]$$

Using the iterative law and linearity of expectations

$$\Rightarrow u_{if,t}(k;p) = \mathbb{E}\left[\sum_{l=0}^{K}\left\{-c_i \times \mathbb{1}\{k\neq 0\} + \beta\mathbb{E}\left[V_{i,t+1}\left(p_{t+1}\right)|p_t=p, a_{ift}=k, a_{jst}=l\right]\times \mathbb{1}\left\{a_{jst}=l\right\}\right\}\Big|\, p_t=p, a_{ift}=k\right]$$

Substituting for u_{ist} from equation A.47

$$\Rightarrow u_{if,t}(k;p) = \mathbb{E}\left[\sum_{l=0}^{K}\left\{u_{ist}\left(k;l,p\right)\times \mathbb{1}\left\{a_{jst}=l\right\}\right\}\Big|\, p_t=p, a_{ift}=k\right]$$

Note that $u_{ist}\left(k;l,p\right)$ is a deterministic constant given the values k,l,p

$$\Rightarrow u_{if,t}(k;p) = \sum_{l=0}^{K}\left\{u_{ist}\left(k;l,p\right)\times \mathbb{E}\left[\mathbb{1}\left\{a_{jst}=l\right\}\Big|\, p_t=p, a_{ift}=k\right]\right\}$$
$$\Rightarrow u_{if,t}(k;p) = \sum_{l=0}^{K} u_{ist}\left(k;l,p\right)\times \sigma_{jst}\left(l;k,p\right) \tag{A.51}$$

Given all this one can use (1) Lemma 1 of McFadden (1973) and (2) the fact that max of T1EV random variables is another Gumbel random variable with a specific location parameter to show that the following two relations hold:

$$\sigma_{if,t}(k;p_t) = P\Big(k = a^*_{if,t}(p_t,\epsilon_{if,t})\Big) = \frac{\exp\Big(u_{if,t}(k;p_t)-u_{if,t}(0;p_t)\Big)}{1+\sum_{l=1}^{K}\exp\Big(u_{if,t}(l;p_t)-u_{if,t}(0;p_t)\Big)}$$
$$\mathbb{E}\left[\max_{k\in\{0,1,\ldots,K\}}\left\{u_{if,t}(k;p_t)-\epsilon_{i,f,t,k}\right\}\,\Big|\, h^t\right] = \ln\left(\sum_{k=0}^{K}\exp\left\{u_{if,t}(k;p_t)\right\}\right)+\gamma \tag{A.52}$$

Given that in the beginning of the period only p_t is the relevant state variable, the value function for player i will be a function of p_t alone. Dropping all irrelevant state variables can also show that the following will hold.

$$V_{i,t}(p_t) = f_i \times \mathbb{E}\left[\max_{k\in\{0,1,\ldots,K\}}\left\{u_{if,t}(k;p_t)-\epsilon_{i,f,t,k}\right\}\,\Big|\, h^t\right] + (1-f_i)\times\mathbb{E}\left[\max_{l\in\{0,1,\ldots,K\}}\left\{u_{is,t}(l;a_{jft},p_t)-\epsilon_{i,s,t,l}\right\}\,\Big|\, h^t, a_{jft}=k\right]$$
$$\Rightarrow V_{i,t}(p_t) = f_i \times \ln\left(\sum_{k=0}^{K}\exp\left\{u_{if,t}(k;p_t)\right\}\right) + (1-f_i)\times\sum_{k=0}^{K}\left[\sigma_{jf,t}(k;p_t)\ln\left(\sum_{l=0}^{K}\exp\left\{u_{is,t}(l;k,p_t))\right\}\right)\right]+\gamma \tag{A.53}$$

Q.E.D.

A.3 Estimation

A.3.1 Proof for Lemma 4.1

Recall $\tilde{X}_t = (A_t, P_{t+1})$. Moreover, random vector $A_t \in \{0, 1, \ldots, K\}^2$ and $P_{t+1} \in \mathbb{R}^K$. In order to derive this density we consider the following probability:

$$\mathbb{P}\left[\tilde{X}_t \in B|\tilde{X}_{t-1}\right] = \sum_{a\in\{0,1,\ldots,K\}^2} \mathbb{P}\left[A_t = a, P_{t+1} \in B_a|(A_{t-1}, P_t)\right] \tag{A.54}$$

The above decomposition is well defined because A_t is a discrete random variable. Note that P_{t+1} is contained in a set and not equal to a point here. Therefore the above probability is not always zero. Moreover, B is a measurable subset of $\left\{\{0, 1, \ldots, K\}^2 \times \mathbb{R}^K\right\}$ and $B_a = \left\{p \in \mathbb{R}^K : (a, p) \in B\right\}$. Also, in case $\nexists p \in \mathbb{R}^K$ s.t. $(a, p) \in B$ then $B_a = \emptyset$, i.e. B_a is empty. The corresponding probability will be 0. The sum appears because $\{0, 1, \ldots, K\}^2$ is finite. Note by model assumption on the popularity process the following holds:

$$\mathbb{P}\left[P_{t+1} \in B_a|(A_t = a, P_t)\right] = \int_{p\in B_a} f\left(p|A_t = a, P_t\right) dp \tag{A.55}$$

Where $f(.|.)$ is defined in equation 2.3. Also note that by equation 2.15 the following is also given to us:

$$\mathbb{P}\left[A_t = a|P_t\right] = \sigma_t\left(A_t = a; P_t\right) \tag{A.56}$$

Therefore, we can express $\mathbb{P}\left[\tilde{X}_t \in B|(A_{t-1}, P_t)\right]$ as followed:

$$\begin{aligned}
\mathbb{P}\left[\tilde{X}_t \in B|\tilde{X}_{t-1}\right] &= \sum_{a\in\{0,1,\ldots,K\}^2} \mathbb{P}\left[A_t = a, P_{t+1} \in B_a|(A_{t-1}, P_t)\right] \\
&= \sum_{a\in\{0,1,\ldots,K\}^2} \mathbb{P}\left[P_{t+1} \in B_a|(A_t = a, P_t, A_{t-1})\right] \mathbb{P}\left[A_t = a|P_t, A_{t-1}\right] \\
&= \sum_{a\in\{0,1,\ldots,K\}^2} \mathbb{P}\left[P_{t+1} \in B_a|(A_t = a, P_t)\right] \mathbb{P}\left[A_t = a|P_t\right] \\
&= \sum_{a\in\{0,1,\ldots,K\}^2} \left(\int_{p\in B_a} f\left(p|A_t = a, P_t\right) dp\right) \sigma_t\left(A_t = a; P_t\right) \\
&= \sum_{a\in\{0,1,\ldots,K\}^2} \int_{p\in B_a} f\left(p|A_t = a, P_t\right) \sigma_t\left(A_t = a; P_t\right) dp \\
\Rightarrow \int_{x\in B} \psi_t(x|X_{t-1}) dx &= \int_{(a,p)\in B} f\left(p|A_t = a, P_t\right) \sigma_t\left(A_t = a; P_t\right) d(a, p)
\end{aligned} \tag{A.57}$$

The second equality holds by law of total probability. This is well defined as A_t is a discrete random variable therefore $\mathbb{P}[A_t = a|.]$ is not always zero. Since $\mathbb{P}\left[\tilde{X}_t \in B|\tilde{X}_{t-1}\right]$ is well-defined then $\mathbb{P}\left[\tilde{X}_t \in B|\tilde{X}_{t-1}, \ldots, \tilde{X}_0\right] = \mathbb{P}\left[\tilde{X}_t \in B|\tilde{X}_{t-1}\right]$. Hence the Markov property is satisfied where ψ_t is the transition density.

Q.E.D.

A.3.2 Proof for Proposition 4.1

Consider the following probability for a measurable set $B \subset \left\{\{0,1,\ldots,K\}^8 \times \mathbb{R}^{4K}\right\}$:

$$\begin{aligned}
\mathbb{P}\left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3}\right) \in B \middle| \tilde{X}_{4d-4}\right] &= \int_{x_1} \mathbb{P}\left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}\right) \in B_{x_1} | X_{4d-3} = x_1\right] \psi_{4d-3}(x_1|\tilde{X}_{4d-4}) dx_1 \\
&= \int_{x_1,x_2} \mathbb{P}\left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}\right) \in B_{x_1,x_2}, \tilde{X}_{4d-2} = x_2 | \tilde{X}_{4d-3} = x_1\right] \psi_{4d-3}(x_1|\tilde{X}_{4d-4}) d(x_1,x_2) \\
&= \int_{x_1,x_2} \mathbb{P}\left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}\right) \in B_{x_1,x_2} | \tilde{X}_{4d-2} = x_2\right] \psi_{4d-2}(x_2|x_1) \psi_{4d-3}(x_1|\tilde{X}_{4d-4}) d(x_1,x_2) \\
&\;\vdots \\
&= \int_{x_1,x_2,x_3,x_4 \in B} \prod_{l=1}^{4} \psi_{4(d-1)+l}\left(x_l|x_{l-1}\right) \Bigg|_{x_0 = \tilde{X}_{4d-4}} d(x_1,x_2,x_3,x_4)
\end{aligned}$$

$$\Rightarrow \mathbb{P}\left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3}\right) \in B \middle| \tilde{X}_{4d-4}\right] = \int_{(a_l,p_{l+1})_{l=1,\ldots,4} \in B} \left(\prod_{l=1}^{4} \sigma_{4(d-1)+l}(a_l;p_l)\right) \times \left(\prod_{l=1}^{4} f\left(p_{l+1}|a_l,p_l\right)\right) d\left((a_l,p_{l+1})_{l=1,\ldots,4}\right) \tag{A.58}$$

Where $p_1 = P_{3d-4}$. We wish to evaluate the following the probability, for an arbitrary measurable set $C \subset \{\{0,1,\ldots,K\}^8 \times \mathbb{R}^K\}$. Note that $B^C = C \times \mathbb{R}^{3K}$ is a measurable subset of $\left\{\{0,1,\ldots,K\}^8 \times \mathbb{R}^{4K}\right\}$ and the following holds

$$\begin{aligned}
\mathbb{P}\left[X_d \in C | X_{d-1}\right] &= \mathbb{P}\left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3}\right) \in B^C \middle| X_{d-1}\right] \\
&= \mathbb{P}\left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3}\right) \in B^C | (P_{4d-3}, A_{4d-4})\right] \\
&= \mathbb{P}\left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3}\right) \in B^C | \tilde{X}_{4d-4}\right] \\
&= \int_{(a_l,p_{l+1})_{l=1,\ldots,4} \in B^C} \left(\prod_{l=1}^{4} \sigma_{4(d-1)+l}(a_l;p_l)\right) \times \left(\prod_{l=1}^{4} f\left(p_{l+1}|a_l,p_l\right)\right) d\left((a_l,p_{l+1})_{l=1,\ldots,4}\right) \\
&= \int_{x \in C} \left(\int_{p_2,p_3,p_4 \in \mathbb{R}^{3K}} \left(\prod_{l=1}^{4} \sigma_{4(d-1)+l}(a_l;p_l)\right) \times \left(\prod_{l=1}^{4} f\left(p_{l+1}|a_l,p_l\right)\right) d(p_1,p_2,p_3)\right) d(x)
\end{aligned} \tag{A.59}$$

Where $x = (a_1, a_2, a_3, a_4, p_5)$ and $p_1 = P_{4d-3}$. The first equality holds by Chapman-Kolmogorov equation for this setting. The second equality holds because X_{d-1} has $X_{\tilde{4d-4}}$ as its component and the corresponding probability is well defined by equation A.58. The following equality is substitution of the expression found in A.58. The last equality is a mere re-writing of the preceding integral. The probability distribution of $X_d \in C$ is nothing but the marginalization of $\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3}\right)$ along the dimensions of P_{4d-2}, P_{4d-1} and P_{4d}.

Q.E.D.

A.4 Bounds

Lemma A.8 *If* $E\left[V_{it+1}(p_t) \middle| a_{Rt}, a_{Dt}, p_t\right] \leq \overline{V}_{t+1}$ *for all* $a_{Rt} = 0,1,\ldots,K$, $a_{Dt} = 0,1,\ldots,K$ *and* $p_t \in \mathbb{R}^K$ *then*

$$V_{it}(p_t) \leq \ln(1 + Ke^{-c_i}) + \gamma + \beta \overline{V}_{t+1} \tag{A.60}$$

We have that $E\left[V_{it+1}(p_t) \middle| a_{Rt}, a_{Dt}, p_t\right] \leq \overline{V}_{t+1}$, then consider u_{ist} as given by proposition 2.1 we have:

$$u_{ist}(k;l,p_t) = -c_i \mathbb{1}\{k \neq 0\} + \beta \mathbb{E}\left[V_{it+1}(p_{t+1}) \middle| k,l,p_t\right] \leq -c_i \mathbb{1}\{k \neq 0\} + \beta \overline{V}_{t+1} \tag{A.61}$$

Similarly, u_{ift}, by proposition A.2 it is given by:

$$u_{ift}(k;l,p_t) = \sum_{l=1}^{K} u_{ist}(k;l,p_T)\sigma_{jst}(l;k,p_T) \leq \sum_{l=1}^{K}\left(\beta\overline{V}_{t+1}\right)\sigma_{ift}(l;k,p_t) = -c_i\mathbb{1}\{k\neq 0\} + \beta\sum_{k=1}^{K} E_k \tag{A.62}$$

Now, we can bound $V_{it}(p_t)$, note by proposition A.2 we have the following:

$$\begin{aligned}
V_{i,t}(p_t) =& f_i \times \ln\left(\sum_{k=0}^{K}\exp\left\{u_{if,t}(k;p_t)\right\}\right) + (1-f_i)\times\sum_{k=0}^{K}\left[\sigma_{jf,t}(k;p_t)\ln\left(\sum_{l=0}^{K}\exp\left\{u_{is,t}(l;k,p_t))\right\}\right)\right] + \gamma \\
\leq& f_i\times\ln\left(\exp\left\{\beta\overline{V}_{t+1}\right\} + K\times\exp\left\{-c_i+\beta\overline{V}_{t+1}\right\}\right) \\
&+ (1-f_i)\times\sum_{k=0}^{K}\left[\sigma_{jf,t}(k;p_t)\ln\left(\exp\left\{\beta\overline{V}_{t+1}\right\} + K\times\exp\left\{-c_i+\beta\overline{V}_{t+1}\right\}\right)\right] + \gamma \\
=& f_i\times\ln\left(\exp\left\{\beta\overline{V}_{t+1}\right\} + K\times\exp\left\{-c_i+\beta\overline{V}_{t+1}\right\}\right) \\
&+ (1-f_i)\ln\left(\exp\left\{\beta\overline{V}_{t+1}\right\} + K\times\exp\left\{-c_i+\beta\overline{V}_{t+1}\right\}\right) + \gamma
\end{aligned} \tag{A.63}$$

The first inequality is implied by inequalities A.62 and A.61 and since $\exp\{x\}$ is an increasing function. The $(K+1)$ appears because we are bounding $K+1$ terms with the same bound. The second equality holds due to the fact that $\sigma_{jf,t}$ is a probability mass function over $k = 0,1,\ldots,K$. Further simplification yields

$$\Rightarrow V_{it}(p_t) \leq \ln\left(\exp\left\{\beta\overline{V}_{t+1}\right\} + K\times\exp\left\{-c_i+\beta\overline{V}_{t+1}\right\}\right) + \gamma = \ln(1+Ke^{-c_i}) + \gamma + \beta\overline{V}_{t+1} \tag{A.64}$$

Hence proved.

Lemma A.9 *If* $E\left[V_{it+1}(p_t)\middle|a_{Rt},a_{Dt},p_t\right] \geq \underline{V}_{t+1}$ *for all* $a_{Rt} = 0,1,\ldots,K$, $a_{Dt} = 0,1,\ldots,K$ *and* $p_t \in \mathbb{R}^K$ *then*

$$V_{it}(p_t) \geq \ln(1+Ke^{-c_i}) + \gamma + \beta\underline{V}_{t+1} \tag{A.65}$$

Consider period t, note the following holds:

$$\mathbb{E}_{p_{t+1}}\left[V_{R,t+1}(p_{t+1})\middle|p_t,a_{RKt},a_{DKt}\right] \geq \underline{V}_{t+1} \tag{A.66}$$

Consider u_{ist} as given by proposition A.2.

$$u_{isT}(k;l,p_T) = -c_i\mathbb{1}\{k\neq 0\} + \beta\mathbb{E}_{p_{T+1}}\left[V_{R,T+1}(p_{T+1})\middle|p_T,a_{RKT},a_{DKT}\right] \geq -c_i\mathbb{1}\{k\neq 0\} + \gamma + \beta\underline{V}_{t+1} \tag{A.67}$$

Similar lemma A.8 it is possible to show the following holds:

$$u_{ift}(k;p_t) \geq -c_i\mathbb{1}\{k\neq 0\} + \beta\underline{V}_{t+1} \tag{A.68}$$

First consider the following log-sum-exp or smooth max:

$$\begin{aligned}
\ln\left\{\exp\left(\beta\underline{V}_{t+1}\right) + K\exp\left(-c_i+\beta\underline{V}_{t+1}\right)\right\} &= \ln\left\{\exp\left(\beta\underline{V}_{t+1}\right)(1+Ke^{-c_i})\right\} \\
&= \beta\underline{V}_{t+1} + \ln(1+Ke^{-c_i})
\end{aligned} \tag{A.69}$$

Going back to $V_{it}(p_t)$ note the following will hold

$$\begin{aligned}V_{i,t}(p_t) =& f_i \times \ln\left(\sum_{k=0}^{K}\exp\left\{u_{if,t}(k;p_t)\right\}\right) + (1-f_i)\times\sum_{k=0}^{K}\left[\sigma_{jf,t}(k;p_t)\ln\left(\sum_{l=0}^{K}\exp\left\{u_{is,t}(l;k,p_t))\right\}\right)\right] + \gamma \\ \geq& f_i \times \ln\left\{\exp\left(\beta\underline{V}_{t+1}\right) + K\exp\left(-c_i + \beta\underline{V}_{t+1}\right)\right\} \\ &+ (1-f_i)\times\sum_{k=0}^{K}\left[\sigma_{jf,t}(k;p_t)\ln\left\{\exp\left(\beta\underline{V}_{t+1}\right) + K\exp\left(-c_i+\beta\underline{V}_{t+1}\right)\right\}\right] + \gamma \\ =& f_i \times \ln\left\{\exp\left(\beta\underline{V}_{t+1}\right) + K\exp\left(-c_i + \beta\underline{V}_{t+1}\right)\right\} + (1-f_i)\ln\left\{\exp\left(\beta\underline{V}_{t+1}\right) + K\exp\left(-c_i+\beta\underline{V}_{t+1}\right)\right\} + \gamma \\ =& \ln\left\{\exp\left(\beta\underline{V}_{t+1}\right) + K\exp\left(-c_i+\beta\underline{V}_{t+1}\right)\right\} + \gamma \\ =& \ln(1+Ke^{-c_i}) + \gamma + \beta\underline{V}_{t+1}\end{aligned} \tag{A.70}$$

Hence proved.

Proposition A.3 *The following inequality holds:*

$$\frac{\left(1-\beta^{T-t+1}\right)(\ln(1+Ke^{-c_i})+\gamma)}{1-\beta} \leq \mathbb{E}\left[V_{it+1}(p_{t+1})\middle|a_{Rt},a_{Dt},p_t\right] \leq \frac{\left(1-\beta^{T-t+1}\right)(\ln(1+Ke^{-c_i})+\gamma)}{1-\beta} + \beta^{T-t+1}\sum_{k=1}^{K}E_k \tag{A.71}$$

Proof: Notice, that given $V_{it}(p_t) \leq \ln(1+Ke^{-c_i}) + \gamma + \beta\overline{V}_{t+1}$, one can easily show that $\mathbb{E}[V_{it}(p_t)|a_{Rt-1},a_{Dt-1},p_{t-1}] \leq \ln(1+Ke^{-c_i}) + \gamma + \beta\overline{V}_{t+1}$. I will use lemma A.8 to find the upper bound inA.3. Note, that the following holds:

$$\mathbb{E}_{p_{T+1}}\left[V_{R,T+1}(p_{T+1})\middle|p_T,a_{RKT},a_{DKT}\right] = \sum_{k=1}^{K}E_k\Phi\left(\frac{\alpha_R a_{Rk,t} + \alpha_D a_{Dk,t} + \tilde{\alpha}a_{Rk,t}a_{Dk,t} + \rho p_{kt} + \delta_k}{\sigma_\nu}\right) \leq \sum_{k=1}^{K}E_k \tag{A.72}$$

Clearly, we can choose $\overline{V}_{T+1} = \sum_k E_k$ and therefore $\overline{V}_T = \ln(1+Ke^{-c_i}) + \gamma + \beta\sum_k E_k$. Given this note that the following holds

$$\mathbb{E}\left[V_{iT}(p_T)|a_{RT-1},a_{DT-1},p_{T-1}\right] \leq \ln(1+Ke^{-c_i}) + \beta\sum_k E_k = \overline{V}_T \tag{A.73}$$

Similarly, note that the candidate for $\overline{V}_{T-1} = \ln(1+Ke^{-c_i})+\gamma+\beta V_T = \ln(1+Ke^{-c_i})+\gamma+\beta\left(\ln(1+Ke^{-c_i})+\gamma\right)+\beta^2\sum_{k=1}^{K}E_k$. Following this one can use this relation iteratively to construct the sequence:

$$V_{T-s} = \sum_{j=0}^{s}\beta^j\Big(\ln(1+Ke^{-c_i})+\gamma\Big) + \beta^{s+1}\sum_{k=1}^{K}E_k \tag{A.74}$$

Then we can finally derive the upper bound

$$\begin{aligned}\overline{V}_t = \overline{V}_{T-(T-t)} &= \sum_{j=0}^{T-t}\beta^j\Big(\ln(1+Ke^{-c_i})+\gamma\Big) + \beta^{T-t+1}E_k \\ &= \frac{\left(1-\beta^{T-t+1}\right)(\ln(1+Ke^{-c_i})+\gamma)}{1-\beta} + \beta^{T-t+1}E_k\end{aligned} \tag{A.75}$$

Let $t = T$, then $\underline{V}_{T+1} = 0$ as $\mathbb{E}[V_{T+1}(p_{T+1})\big|a_{RT},a_{DT},p_T] \geq 0$ then given Lemma A.9 $\underline{V}_T = \ln(1+Ke^{-c_i})+\gamma$. Applying the lemma again $\underline{V}_{T-1} = \ln(1+Ke^{-c_i})+\gamma+\beta\Big(\ln(1+Ke^{-c_i})+\gamma\Big)$, therefore for $T-j$ we have

$$\underline{V}_{T-j} = \sum_{s=0}^{j}\beta^s\Big(\ln(1+Ke^{-c_i})+\gamma\Big) = \frac{\left(1-\beta^{j+1}\right)(\ln(1+Ke^{-c_i})+\gamma)}{1-\beta} \tag{A.76}$$

Then for a period t, we will have:

$$\underline{V}_t = \underline{V}_{T-(T-t)} = \frac{\left(1-\beta^{T-t+1}\right)\left(\ln(1+Ke^{-c_i})+\gamma\right)}{1-\beta} \tag{A.77}$$

Hence proved.

Proposition A.4 *For $k > 0$ let $\Delta_k \mathbb{E}(V_{it})$ be defined as:*

$$\Delta_k \mathbb{E}(V_{it}) = \mathbb{E}\left[V_{it+1}(p_{t+1})\middle|a_{Rt} = k, a_{Dt}, p_t\right] - \mathbb{E}\left[V_{it+1}(p_{t+1})\middle|a_{Rt} = 0, a_{Dt}, p_t\right] \tag{A.78}$$

Then

$$\left|\Delta_k \mathbb{E}(V_{it})\right| \leq \beta^{T-t+1} \sum_{k=1}^{K} E_k \tag{A.79}$$

Proof: This is a direct implication of proposition A.3, as

$$\begin{aligned}
\Delta_k \mathbb{E}(V_{it}) =& \mathbb{E}\left[V_{it+1}(p_{t+1})\middle|a_{Rt} = k, a_{Dt}, p_t\right] - \mathbb{E}\left[V_{it+1}(p_{t+1})\middle|a_{Rt} = 0, a_{Dt}, p_t\right] \\
\leq& \frac{\left(1-\beta^{T-t+1}\right)\left(\ln(1+Ke^{-c_i})+\gamma\right)}{1-\beta} + \beta^{T-t+1}E_k - \mathbb{E}\left[V_{it+1}(p_{t+1})\middle|a_{Rt} = 0, a_{Dt}, p_t\right] \\
\leq& \frac{\left(1-\beta^{T-t+1}\right)\left(\ln(1+Ke^{-c_i})+\gamma\right)}{1-\beta} + \beta^{T-t+1}E_k - \frac{\left(1-\beta^{T-t+1}\right)\left(\ln(1+Ke^{-c_i})+\gamma\right)}{1-\beta} \\
=& \beta^{T-t+1} \sum_{k=1}^{K} E_k
\end{aligned} \tag{A.80}$$

The first inequality is true because of the upper bound inequality in proposition A.3 and the second is true by the lower bound inequality of A.3. Similarly,

$$\begin{aligned}
\Delta_k \mathbb{E}(V_{it}) =& \mathbb{E}\left[V_{it+1}(p_{t+1})\middle|a_{Rt} = k, a_{Dt}, p_t\right] - \mathbb{E}\left[V_{it+1}(p_{t+1})\middle|a_{Rt} = 0, a_{Dt}, p_t\right] \\
\geq& \frac{\left(1-\beta^{T-t+1}\right)\left(\ln(1+Ke^{-c_i})+\gamma\right)}{1-\beta} - \mathbb{E}\left[V_{it+1}(p_{t+1})\middle|a_{Rt} = 0, a_{Dt}, p_t\right] \\
\geq& \frac{\left(1-\beta^{T-t+1}\right)\left(\ln(1+Ke^{-c_i})+\gamma\right)}{1-\beta} - \frac{\left(1-\beta^{T-t+1}\right)\left(\ln(1+Ke^{-c_i})+\gamma\right)}{1-\beta} + \beta^{T-t+1}E_k \\
=& -\beta^{T-t+1} \sum_{k=1}^{K} E_k
\end{aligned} \tag{A.81}$$

The first inequality is true because of the lower bound inequality in proposition A.3 and the second is true by the upper bound inequality of A.3. Hence proved.

Corollary A.1 *For $\Delta_k \mathbb{E}(V_{it}) \approx 0$ for large $T - t$.*

Proof: The proof follows from proposition A.4 and applying squeeze theorem.

$$-\beta^{T-t+1} \sum_{k=1}^{K} E_k \leq \Delta_k \mathbb{E}(V_{it}) \leq \beta^{T-t+1} \sum_{k=1}^{K} E_k \tag{A.82}$$

Since $\beta < 1$, $\beta^{T-t+1} \to 0$ as $T - t \to \infty$. Therefore upper and lower bounds both tend to 0, as a result $\Delta_k \mathbb{E}(V_{it})$. Hence proved.

B Numerical Approximation

B.1 Primitives

Let $\mathbf{P} = \{(\mathbf{p}_1^r, \ldots, \mathbf{p}_K^r)\}_{r=1}^R$ be the state variable grid. Let $\mathbf{T}(\mathbf{p}) = \left(\mathbf{T}_1(\mathbf{p}), \mathbf{T}_2(\mathbf{p}), \ldots, \mathbf{T}_R(\mathbf{p})\right)$ be a vector collecting Chebyshev polynomial terms corresponding to an arbitrary grid point $\mathbf{p}$. The approximated values all value functions in the model take for a $\mathbf{p} \in \mathbf{P}$ be given by:

$$\left\{\tilde{V}_{R,t}(\mathbf{p}), \tilde{V}_{D,t}(\mathbf{p}), \left\{\tilde{u}_{R,f,t}(k;\mathbf{p}), \tilde{u}_{D,f,t}(k;\mathbf{p}), \{\tilde{u}_{R,s,t}(k;l,\mathbf{p}), \tilde{u}_{D,s,t}(k;l,\mathbf{p})\}_{l=0}^K\right\}_{k=0}^K\right\}_{t=1}^T$$

The approximated values all conditional choice probabilities in the model take for a $\mathbf{p} \in \mathbf{P}$ be given by:

$$\left\{\left\{\tilde{\sigma}_{R,f,t}(k;\mathbf{p}), \tilde{\sigma}_{D,f,t}(k;\mathbf{p}), \{\tilde{\sigma}_{R,s,t}(k;l,\mathbf{p}), \tilde{\sigma}_{D,s,t}(k;l,\mathbf{p})\}_{l=0}^K\right\}_{k=0}^K\right\}_{t=1}^T$$

I approximate the value functions by Chebyshev polynomials. In order to interpolate a value functions I only need to know the value of the coefficients of these polynomials. Let the coefficients of the polynomial terms approximating $V_{i,t}(.)$ be denoted by $\gamma_{i,t}^V$, $u_{i,f,t}(k;.)$ be denoted by $\gamma_{i,t,k}^f$ and $u_{i,s,t}(k;l,.)$ by $\gamma_{i,t,k,l}^s$. Apart from these, I also need a Gaussian quadrature for calculating conditional expectation. Let $\nu = \{(\nu_1^s, \ldots, \nu_K^s, \omega^s)\}_{s=1}^S$ be a Gaussian quadrature.

B.2 Last Period

For period T, we do not require coefficients in period $T+1$, nor the Gaussian quadrature because the conditional expectation of the value function can be computed. The following equations describe how to evaluate all value function values over the grid $\mathbf{P}$. Note here the approximated values are equal to true values.

$$\tilde{u}_{i,s,T}(a_i;a_j,\mathbf{p}^r) = -c_i(1-a_{i,0}) + \beta\sum_{k=1}^K E_k\Phi\left(\frac{\alpha_i a_{i,k} + \alpha_j a_{j,k} + \tilde{\alpha} a_{i,k} a_{i,k} + \rho\mathbf{p}_k^r + \delta_k}{\sigma_\nu}\right) \tag{B.1}$$

$$\tilde{\sigma}_{i,s,T}(a_j;a_i,\mathbf{p}^r) = \frac{\exp\left(\tilde{u}_{i,s,T}(a_i;a_j,\mathbf{p}^r) - \tilde{u}_{i,s,T}(0;a_j,\mathbf{p}^r)\right)}{1+\sum_{l=1}^K \exp\left(\tilde{u}_{i,s,T}(l;a_j,\mathbf{p}^r) - \tilde{u}_{i,s,T}(0;a_j,\mathbf{p}^r)\right)} \tag{B.2}$$

$$\tilde{u}_{i,f,T}(a_i;a_j,\mathbf{p}^r) = -c_i(1-a_{i,0}) + \beta\sum_{a_j=0}^K\sum_{k=1}^K E_k\Phi\left(\frac{\alpha_i a_{i,k} + \alpha_j a_{j,k} + \tilde{\alpha} a_{i,k} a_{i,k} + \rho\mathbf{p}_k^r + \delta_k}{\sigma_\nu}\right)\times\tilde{\sigma}_{j,s,T}(a_j;a_i,\mathbf{p}^r) \tag{B.3}$$

$$\tilde{\sigma}_{i,f,T}(a_j;\mathbf{p}^r) = \frac{\exp\left(\tilde{u}_{i,f,T}(a_i;\mathbf{p}^r) - \tilde{u}_{i,s,T}(0;\mathbf{p}^r)\right)}{1+\sum_{l=1}^K \exp\left(\tilde{u}_{i,f,T}(l;\mathbf{p}^r) - \tilde{u}_{i,s,T}(0;f\mathbf{p}^r)\right)} \tag{B.4}$$

$$\tilde{V}_{i,T}(\mathbf{p}^r) = f_i\log\left(\sum_{a_i=0}^K \exp\left(\tilde{u}_{i,f,T}(a_i;\mathbf{p}^r)\right)\right) + (1-f_i)\sum_{a_j=0}^K\log\left(\sum_{a_i=0}^K \tilde{u}_{i,s,T}(a_i;a_j,\mathbf{p}^r)\right)\times\tilde{\sigma}_{j,s,T}(a_j;a_i,\mathbf{p}^r) \tag{B.5}$$

Where $a_{i,k} = \mathbb{1}\{a_i = k\}$ for all $i \in \{R,D\}$ and $k \in \{0,1,\ldots,K\}$.

B.3 Period $\mathbf{t}$: Interpolating Polynomials

In an arbitrary period t, suppose we have computed the values of the approximated value functions. Define $\mathbf{T}$ as the matrix obtained by collecting transpose of all Chebyshev polynomial terms at each $\mathbf{p} \in \mathbf{P}$:

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}(\mathbf{p}^1)^\top \\ \mathbf{T}(\mathbf{p}^2)^\top \\ \vdots \\ \mathbf{T}(\mathbf{p}^R)^\top \end{bmatrix}_{R\times R} = \begin{bmatrix} \mathbf{T}_1(\mathbf{p}^1) & \mathbf{T}_2(\mathbf{p}^1) & \dots & \mathbf{T}_R(\mathbf{p}^1) \\ \mathbf{T}_1(\mathbf{p}^2) & \mathbf{T}_2(\mathbf{p}^2) & \dots & \mathbf{T}_R(\mathbf{p}^2) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{T}_1(\mathbf{p}^R) & \mathbf{T}_2(\mathbf{p}^R) & \dots & \mathbf{T}_R(\mathbf{p}^R) \end{bmatrix}_{R\times R} \tag{B.6}$$

We can collect the approximated values of the value function $\tilde{V}_{i,t}$ for each i as a vector and pre-multiply by $\mathbf{T}^{-1}$ to obtain the interpolating polynomial coefficients specific to $\tilde{V}_{i,t}$.

$$\gamma^V_{i,t} = \mathbf{T}^{-1} \begin{bmatrix} \tilde{V}_{i,t}(\mathbf{p}^1) \\ \tilde{V}_{i,t}(\mathbf{p}^2) \\ \vdots \\ \tilde{V}_{i,t}(\mathbf{p}^R) \end{bmatrix} \qquad i \in \{R, D\} \tag{B.7}$$

Similarly collect the approximated values of the option specific value function for the first mover, $\tilde{u}_{i,f,k,t}$, for each i, k as a vector and pre-multiply by $\mathbf{T}^{-1}$ to obtain the interpolating polynomial coefficients specific to $\tilde{u}_{i,f,k,t}$.

$$\gamma^f_{i,k,t} = \mathbf{T}^{-1} \begin{bmatrix} \tilde{u}_{i,f,t}(k; \mathbf{p}^1) \\ \tilde{u}_{i,f,t}(k; \mathbf{p}^2) \\ \vdots \\ \tilde{u}_{i,f,t}(k; \mathbf{p}^R) \end{bmatrix} \qquad i \in \{R, D\}, k \in \{0, 1, \dots, K\} \tag{B.8}$$

Lastly, collect the approximated values of the option specific value function for the second mover, $\tilde{u}_{i,s,k,l,t}$, for each i, k, l as a vector and pre-multiply by $\mathbf{T}^{-1}$ to obtain the interpolating polynomial coefficients specific to $\tilde{u}_{i,s,k,l,t}$.

$$\gamma^s_{i,k,l,t} = \mathbf{T}^{-1} \begin{bmatrix} \tilde{u}_{i,s,t}(k; l, \mathbf{p}^1) \\ \tilde{u}_{i,s,t}(k; l, \mathbf{p}^2) \\ \vdots \\ \tilde{u}_{i,s,t}(k; l, \mathbf{p}^R) \end{bmatrix} \qquad i \in \{R, D\}, k \in \{0, 1, \dots, K\}, l \in \{0, 1, \dots, K\} \tag{B.9}$$

Once, we have obtained these coefficients, it allows us to interpolate the value functions and conditional choice probabilities at any given popularity standing p. The following expressions need to be evaluated for the interpolation exercise:

$$V_{i,t}(p) \approx \hat{V}_{i,t}(p) = \sum_{r=1}^{R} \gamma^V_{i,t;r} \mathbf{T}_r(p) \tag{B.10}$$

$$u_{i,f,t}(k; p) \approx \hat{u}_{i,f,t}(k; p) = \sum_{r=1}^{R} \gamma^f_{i,k,t;r} \mathbf{T}_r(p) \tag{B.11}$$

$$u_{i,s,t}(k; l, p) \approx \hat{u}_{i,s,t}(k; l, p) = \sum_{r=1}^{R} \gamma^s_{i,k,l,t;r} \mathbf{T}_r(p) \tag{B.12}$$

Here, $\gamma^V_{i,t;r}$, $\gamma^f_{i,k,t;r}$, $\gamma^s_{i,k,l,t;r}$ are r^{th} components of the vectors $\gamma^V_{i,t}$, $\gamma^f_{i,k,t}$, $\gamma^s_{i,k,l,t}$. Moreover, we can also evaluate conditional choice probabilities as followed:

$$\sigma_{i,s,t}(a_i;a_j,p) \approx \hat{\sigma}_{i,s,t}(a_i;a_j,p) = \frac{\exp\Big(\sum_{r_2=1}^{R}\big(\gamma^s_{i,a_i,a_j,t;r} - \gamma^s_{i,0,a_j,t;r}\big)\mathbf{T}_r(p)\Big)}{1+\sum_{k=1}^{K}\exp\Big(\sum_{r_2=1}^{R}\big(\gamma^s_{i,k,a_j,t;r} - \gamma^s_{i,0,a_j,t;r}\big)\mathbf{T}_r(p)\Big)} \tag{B.13}$$

$$\sigma_{i,f,t}(a_i;p) \approx \hat{\sigma}_{i,f,t}(a_i;p) = \frac{\exp\Big(\sum_{r_2=1}^{R}\big(\gamma^f_{i,a_i,t;r} - \gamma^f_{i,0,t;r}\big)\mathbf{T}_r(p)\Big)}{1+\sum_{k=1}^{K}\exp\Big(\sum_{r_2=1}^{R}\big(\gamma^f_{i,k,t;r} - \gamma^f_{i,0,t;r}\big)\mathbf{T}_r(p)\Big)} \tag{B.14}$$

Based on $\hat{\sigma}_{i,s,t}(a_i;a_j,p)$ and $\hat{\sigma}_{i,f,t}(a_i;p)$ we can define the approximation of $\sigma_t(a_R,a_D;p)$ as followed:

$$\begin{aligned}\sigma_t(a_R,a_D;p) &\approx \hat{\sigma}_t(a_i,a_j;p)\\ &= f\hat{\sigma}_{R,f,t}\left(a_R;p\right)\hat{\sigma}_{D,s,t}\left(a_R;a_D,p\right) + (1-f)\,\hat{\sigma}_{D,f,t}\left(a_D;p\right)\hat{\sigma}_{R,s,t}\left(a_D;a_R,p\right)\end{aligned} \tag{B.15}$$

This property will be used extensively in the next subsection.

B.4 Period $\mathbf{t}$: Approximate Value Functions and CCPs on the grid

We can obtain period $t+1$ interpolating polynomial coefficients by following steps in the previous subsection. Now we will build over that in this subsection with the objective of obtaining period t values of the value functions over the grid $\mathbf{P}$. First, by following Judd et al. (2017), define vectors $I_{r,k,l}$ for each r,k,l that collects the integrated Chebyshev polynomial terms as followed:

$$I^{r'}_{r,k,l} = \sum_{s=1}^{S}\mathbb{T}_{r'}\left(\begin{bmatrix}\alpha_R\mathbb{1}\{k==1\}+\alpha_D\mathbb{1}\{l==1\}+\tilde{\alpha}\mathbb{1}\{k==1,l==1\}+\rho\mathbf{p}^r_1+\delta_1+\sigma_\nu\nu^s_1\\ \vdots\\ \alpha_R\mathbb{1}\{k==K\}+\alpha_D\mathbb{1}\{l==K\}+\tilde{\alpha}\mathbb{1}\{k==K,l==K\}+\rho\mathbf{p}^r_K+\delta_K+\sigma_\nu\nu^s_K\end{bmatrix}\right)\omega^s \tag{B.16}$$

Here, $\nu^s_1,\dots,\nu^s_K$ are Gaussian shocks and w^s is the weight of these shocks. The choice of the Gaussian quadrature is discussed later. Consequently define $I_{r,k,l} = (I^1_{r,k,l}, I^2_{r,k,l},\dots,I^R_{r,k,l})$. Note none of the terms used in this calculation depends upon the period t and therefore this calculation needs to be done once outside the value iteration loop. The integrated Chebyshev polynomial, $I_{r,k,l}$, can be used to calculate $\tilde{u}_{i,s,t}$ as followed:

$$\tilde{u}_{i,s,t}(a_i;a_j,\mathbf{p}^r) = -c_i(1-a_{i,0}) + \beta\sum_{r_2=1}^{R}\gamma^V_{i,t+1;r_2}I^{r_2}_{r,a_i,a_j} \tag{B.17}$$

Here $\gamma^V_{i,t+1;r_2}$ is the coefficient of the r_2^{th} term of the polynomial interpolating $V_{i,t+1}()$. The sum "$\sum_{r=1}^{R}\gamma^V_{i,t+1,r_2}I^{r_2}_{r,a_i,a_j}$" approximates $\mathbb{E}[V_{i,t+1}(p)|a_i,a_j,\mathbf{p}^r]$. Moreover the choice of the Gaussian quadrature ensures that the error in this approximation depends on the degree of the Chebyshev polynomial. The conditional choice probability for the second mover is given by:

$$\tilde{\sigma}_{i,s,t}(a_i;a_j,\mathbf{p}^r) = \frac{\exp\Big(-c_i(1-a_{i,0}) + \beta\sum_{r_2=1}^{R}\gamma^V_{i,t+1,r_2}\big(I^{r_2}_{r,a_i,a_j} - I^{r_2}_{r,0,a_j}\big)\Big)}{1+\sum_{k=1}^{K}\exp\Big(-c_i + \beta\sum_{r_2=1}^{R}\gamma^V_{i,t+1,r_2}\big(I^{r_2}_{r,k,a_j} - I^{r_2}_{r,0,a_j}\big)\Big)} \tag{B.18}$$

Provided the above we can calculate the first mover's pay-offs as followed:

$$\tilde{u}_{i,f,t}(a_i;a_j,\mathbf{p}^r) = -c_i(1-a_{i,0}) + \beta\sum_{a_j=0}^{K}\left(\sum_{r_2=1}^{R}\gamma^V_{i,t+1;r_2}I^{r_2}_{r,a_i,a_j}\right)\times\tilde{\sigma}_{j,s,t}(a_j;a_i,\mathbf{p}^r) \tag{B.19}$$

Here there are two expectations, the outer expectation is with respect to the opponent's second mover conditional choice probabilities. The inner expectation is over next period popularity. Provided this, it is possible to compute $\tilde{\sigma}_{i,f,t}(a_i;\mathbf{p}^r)$.

$$\tilde{\sigma}_{i,f,t}(a_i;\mathbf{p}^r) = \frac{\exp\Big(-c_i(1-a_{i,0}) + \beta\sum_{r_2=1}^{R}\gamma^V_{i,t+1,r_2}\Big(\sum_{a_j=0}^{K}(I^{r_2}_{r,a_i,a_j}\tilde{\sigma}_{j,s,t}(a_j;a_i,\mathbf{p}^r) - I^{r_2}_{r,0,a_j}\tilde{\sigma}_{j,s,t}(a_j;0,\mathbf{p}^r)\Big)\Big)}{1+\sum_{k=1}^{K}\exp\Big(-c_i + \beta\sum_{r_2=1}^{R}\gamma^V_{i,t+1,r_2}\Big(\sum_{a_j=0}^{K}(I^{r_2}_{r,k,a_j}\tilde{\sigma}_{j,s,t}(a_j;k,\mathbf{p}^r) - I^{r_2}_{r,0,a_j}\tilde{\sigma}_{j,s,t}(a_j;0,\mathbf{p}^r)\Big)\Big)} \tag{B.20}$$

We can compute the approximated value function at period t for an arbitrary $\mathbf{p}^r\in\mathbf{P}$ as followed:

$$\tilde{V}_{i,t}(\mathbf{p}^r) = f_i\log\Big(\sum_{a_i=0}^{K}\exp\big(\tilde{u}_{i,f,t}(a_i;\mathbf{p}^r)\big)\Big) + (1-f_i)\sum_{a_j=0}^{K}\log\Big(\sum_{a_i=0}^{K}\tilde{u}_{i,s,t}(a_i;a_j,\mathbf{p}^r)\Big)\times\tilde{\sigma}_{j,s,t}(a_j;a_i,\mathbf{p}^r) \tag{B.21}$$

B.5 Sparse Grid, Polynomial and Gaussian Quadrature

I follow Judd et al. (2014) for constructing a Smolyak Grid for approximation level μ and the corresponding Chebyshev polynomial. I construct Smolyak Grid, $\mathbf{U} = \{(\mathbf{u}^r_1,\dots,\mathbf{u}^r_K)\}_{r=1}^{R}$ over $[-1,1]^K$ and its corresponding Chebyshev polynomial $\mathbf{\Psi}(u)$. Then the grid $\mathbf{P} = \{(\mathbf{p}^r_1,\dots,\mathbf{p}^r_K)\}_{r=1}^{R}$ for a given set of parameters $\rho,\sigma_\nu,\delta_1,\dots,\delta_K$ is constructed as followed:

$$\begin{aligned}
\mathbf{p}^r &= \underline{p}_k + (\bar{p}_k - \underline{p}_k)\frac{\mathbf{u}^r_k+1}{2}\\
\text{where}\quad \bar{p}_k &= \frac{\delta_k}{1-\rho} + \lambda\left(\frac{\alpha_R}{1-\rho} + \frac{3\sigma_\nu}{\sqrt{1-\rho^2}}\right)\\
\underline{p}_k &= \frac{\delta_k+\alpha_D}{1-\rho} + \lambda\left(\frac{\alpha_D}{1-\rho} - \frac{3\sigma_\nu}{\sqrt{1-\rho^2}}\right)
\end{aligned} \tag{B.22}$$

Here λ is a tuning parameter which controls the size of the grid. I chose $\lambda=1$ after conducting extensive Monte-Carlo Experimentsrefer to Section D. The value, $\lambda=1$, provided me with the least bias and adequate MSE. The Chebyshev polynomial $\mathbf{T}(p) = (\mathbf{T}_1(p),\mathbf{T}_2(p),\dots,\mathbf{T}_R(p))$ is defined as followed:

$$\mathbf{T}_r(p) = \mathbf{\Psi}_r\left(2\Big(\frac{p_1-\underline{p}_1}{\bar{p}_1-\underline{p}_1}\Big)-1, 2\Big(\frac{p_2-\underline{p}_2}{\bar{p}_2+\underline{p}_2}\Big)-1,\dots,2\Big(\frac{p_K-\underline{p}_K}{\bar{p}_K+\underline{p}_K}\Big)-1\right) \tag{B.23}$$

Where $\mathbf{\Psi}_r(.)$ is the r^{th} Chebyshev polynomial term. The Gaussian quadrature, denoted by $\nu = \{(\nu^s_1,\dots,\nu^s_K,\omega^s)\}_{s=1}^{S}$, is obtained from http://www.sparse-grids.de/. I choose KPN for K dimensions and degree $2^\mu+1$. This quadrature can compute exact integral of a $K-$dimensional complete polynomial of maximal degree $2^\mu+1$.

B.6 Algorithm

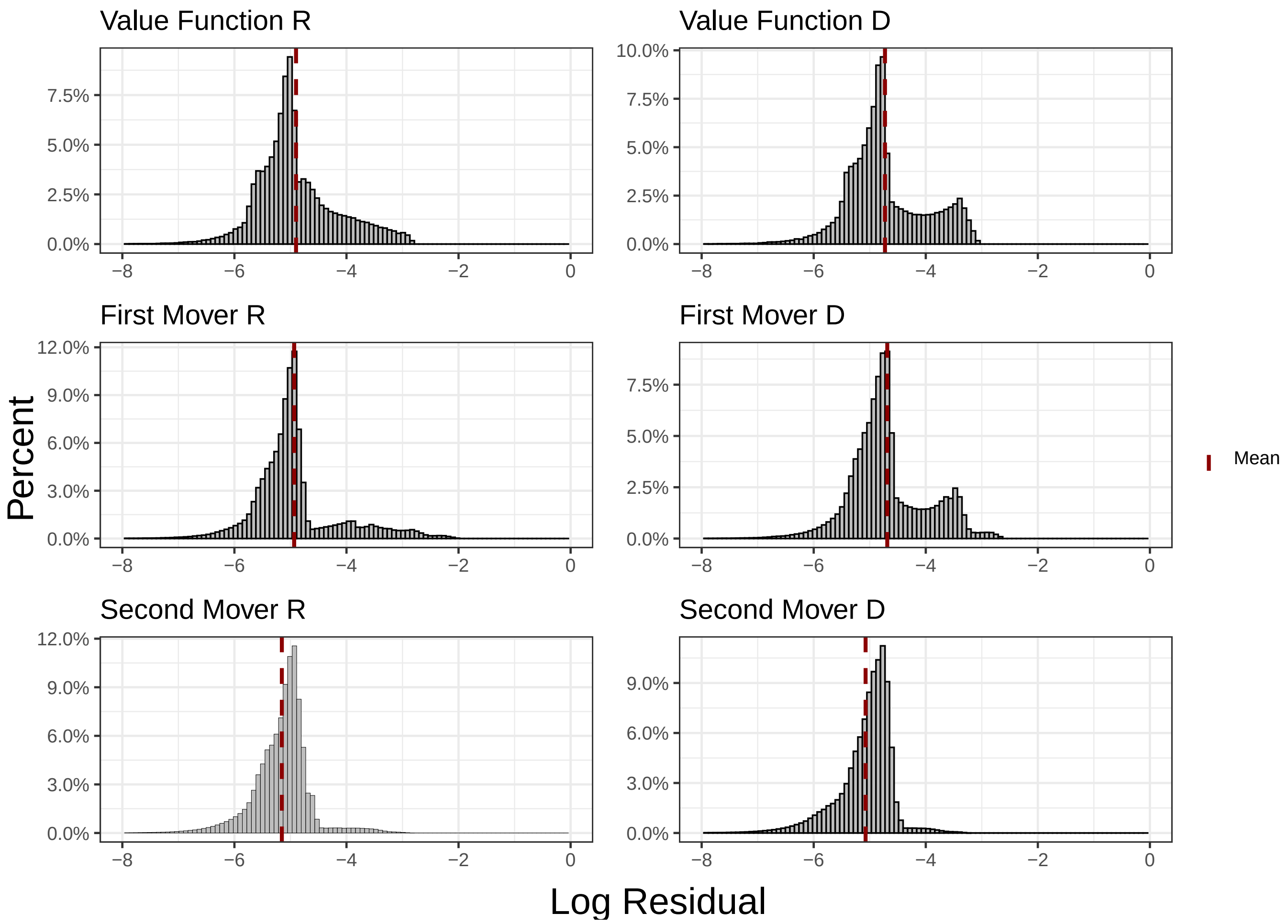
Here I will describe the algorithm step-by-step using the equations discussed above. The algorithm will be defined for a given parameter values, $\theta_{\text{Popularity}} = \{\alpha_R,\alpha_D,\tilde{\alpha},\rho,\sigma_\nu,\delta_1,\delta_2,\dots,\delta_K\}$ and $\theta_{\text{Cost}} = \{c_R,c_D,c_1,c_2,\dots,c_K\}$ and the approximation level μ.

Step 0 Generate the Smolyak pair, $\mathbf{U},\mathbf{\Psi}$ for K dimensions and approximation level μ by following Judd et al. (2014). Obtain KPN Gaussian quadrature $\nu = \{(\nu^s_1,\dots,\nu^s_K,w^s)\}_{s=1}^{S}$ from http://www.sparse-grids.de/ for K dimensions and approximation level $2^\mu+1$.

Step 1 Compute the parameter-specific $\mathbf{P},\mathbf{T}$ using equations B.22 and B.23.

Step 2 Pre-Compute integrals of Chebyshev terms contingent on current popularity and candidate decisions by equation B.16.

Figure 10: Residual Equation Errors



Notes: The histograms report the residual equation errors in decimal log basis. The dashed line marks the mean of residual euqation error.

Step 3 Approximate Backward Induction.

1 Carry-out the following two steps:

- Compute $\{\tilde{V}_{i,T}, \tilde{u}_{i,f,T}, \tilde{u}_{i,s,T}, \tilde{\sigma}_{i,f,T}, \tilde{\sigma}_{i,s,T}\}$ for candidate $i = R, D$ by following equations B.5, B.3, B.1,B.4, B.2 respectively.
- Obtain coefficients of the interpolating polynomials, $\{\gamma^V_{i,T}, \gamma^f_{i,k,T}, \gamma^s_{i,k,l,T}\}$ for candidate $i = R, D$ and $k, l = 0, \ldots, K$ by following equations B.7, B.8 and B.9 for period $t = T$ respectively.

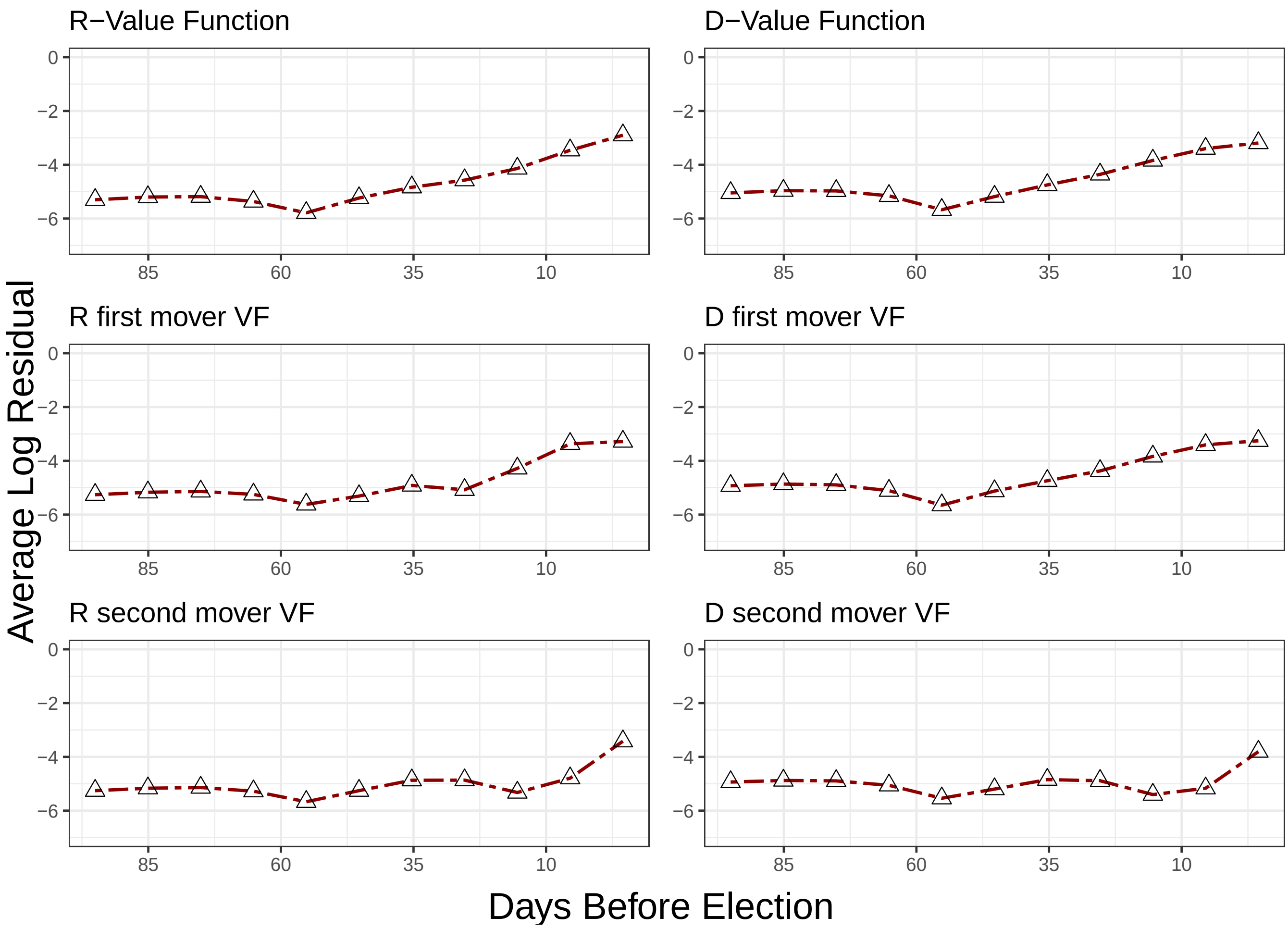
2 For $t = 1, 2, \ldots, T-1$ do the following:

- Compute $\{\tilde{V}_{i,T-t}, \tilde{u}_{i,f,T-t}, \tilde{u}_{i,s,T-t}, \tilde{\sigma}_{i,f,T-t}, \tilde{\sigma}_{i,s,T-t}\}$ for candidate $i = R, D$ by following equations B.21, B.19, B.17,B.20, B.18 respectively.
- Obtain coefficients of the interpolating polynomials, $\{\gamma^V_{i,T-t}, \gamma^f_{i,k,T-t}, \gamma^s_{i,k,l,T-t}\}$ for candidate $i = R, D$ and $k, l = 0, \ldots, K$ by following equations B.7, B.8 and B.9 for period $t = T$ respectively.

B.7 Accuracy of Numerical Approximation

I evaluate the accuracy of the numerical approximation by computing the errors of the residual equations (Judd, 1992). I simulate the model 400 times. This produce a set of popularity values, $\left\{\{(p_{1,t,i}, \ldots, p_{K,t,m})\}_{t=1}^{T}\right\}_{m=1}^{400}$. For each $p_{t,m} = (p_{1,t,m}, \ldots, p_{K,t,m})$,

Figure 11: Day-Specific Residual Equation Errors



let $Gp_{k,l}(p) = E[p_{t+1}|a_{R,t} = k, a_{D,t} = l, p]$ and define

$$u^{t+1}_{i,s,t}(k; l, p_{t,m}) = -c_i\,(1 - a_{i,0}) + \beta \sum_{r=1}^{R} \gamma^{V}_{i,t+1;r} \left(\sum_{s=1}^{S} \mathbf{T}_r\left(Gp_{k,l}(p_{t,m} + \sigma_\nu \nu^s)\right) \omega^s \right) \tag{B.24}$$

I calculate the residuals of equilibrium equation defining second mover value function, $\mathcal{R}_{i,s,k,l,t}(\gamma; p_{t,m})$ as followed:

$$\mathcal{R}_{i,s,k,l,t}(\gamma; p_{t,m}) = 1 - \frac{u^{t+1}_{i,s,t}(k; l, p_{t,m})}{\hat{u}_{i,s,t}(k; l, p_{t,m})} \quad \text{for all } i, k, l, t \tag{B.25}$$

Similarly define $u^{t+1}_{f,s,t}$ as followed:

$$u^{t+1}_{i,f,t}(k; p_{t,m}) = \sum_{l=1}^{k} u^{t+1}_{i,s,t}(k; l, p_{m,t}) \left(\frac{\exp\left(u^{t+1}_{j,s,t}(l; k, p_{m,t}) - u^{t+1}_{j,s,t}(0; k, p_{m,t})\right)}{1 + \sum_{l'=1}^{K} \exp\left(u^{t+1}_{j,s,t}(l'; k, p_{m,t}) - u^{t+1}_{j,s,t}(0; k, p_{m,t})\right)} \right) \tag{B.26}$$

Then define $\mathcal{R}_{i,f,k,t}(\gamma; p_{t,m})$

$$\mathcal{R}_{i,f,k,t}(\gamma; p_{t,m}) = 1 - \frac{u^{t+1}_{i,f,t}(k; p_{t,m})}{\hat{u}_{i,f,t}(k; p_{t,m})} \quad \text{for all } i, k, l, t \tag{B.27}$$

In similar fashion one can define $V^{t+1}_{i,t}(p_{t,m})$ and then calculate the corresponding residuals, denoted by $\mathcal{R}_{i,t}(\gamma; p_{t,m})$ for all i, t. Note by construction these residual values are all zero at the collocation points $\mathbf{P}$. These residual equations calculate the discrepancy between value functions derived by the numerical algorithm ($\hat{u}_{i,f,t}$, $\hat{u}_{i,s,t}$ and $\hat{V}_{i,t}$) and the ones obtained from the equilibrium conditions ($u_{i,f,t}^{t+1}$, $u_{i,s,t}^{t+1}$ and $V_{i,t}^{t+1}$) in points of the state space which are different from the collocation points. I report the decimal log of absolute values of these residuals errors. In Figure 10 I show the histogram of those errors.

The average residual equation errors are in the order of -4.9, -4.73 for R and D's value functions (resp.); -4.93 and -4.69 for R's and D's first mover value function; and -5.15 and -5.07 for R and D's second mover value functions. Given the complexity of the model these discrepancies are in a reasonable range.

C Data Appendix

The group of activities I am interested in involves a candidate (i) holding a rally, (ii) giving a speech, or (iii) organizing a special event. I call these activities a political rally. In various media reports, most of these organized special events (such *Focus* events, *Early Vote* events, *Get out Vote* events, etc.) are reported as rallies, or there is evidence that the candidate delivered a speech to voters. For instance, consider 2004 elections— even though currently not part of my empirical application. There are events called *Focus Events* used by George W. Bush's presidential campaign. An example of such entry in the calendar is regarding October 20, 2014, Rochester Airport rally. The entry in *Democracy in Action* is given as '*GWB participates in a "Focus on the Economy with President Bush" event at Rochester Aviation hanger in Rochester, MN*'. The same event was also reported as a rally http://news.minnesota.publicradio.org/features/2004/10/20_ap_bushrochester/. Another example for a set of entries that are akin to rallies but were entered as campaign events are *Early Vote Events.* Consider the October 21 *Early Vote Event* in Cleveland, Ohio by Hilary R. Clinton. In the recording— link: https://www.youtube.com/watch?v=abbxQn-9DBY— of the event Hilary R. Clinton can be seen delivering a speech to a large gathering of voters.

In the model, candidates can hold at most one rally in a given period, but empirically there are days when candidates visit multiple states for holding rallies. Therefore, I define a period as a quarter of the day and assign periods to observed rallies by using the chronological information for all activities. First, every day is divided into 4 sub-periods. To achieve this, I need to make sure there are at most four rallies in a day. I had to remove nine rallies for being a) late-night/post-midnight rally on the last day b) rallies in the same state consecutively[53]. I also removed rallies from states where less than two rallies were held at most. States with rare rallies are stronghold states. These rallies do not influence electoral outcomes within those states.

Second I assign periods to each rally by carrying out the following steps. I calculate the total number of appearances a candidate makes in a day (let us say n). If a rally was i^{th} appearance made by the candidate, then it received a score of i/n. Once all scores, then periods within a day are assigned in the following manner: 1) If $i/n \leq 0.25$, it is considered the first period within the day. 2) If $0.25 < i/n \leq 0.5$, it is considered the second period within the day. 3) If $0.5 < i/n \leq 0.75$, it is considered the third period within the day. 4) If $0.75 < i/n \leq 1$, it is considered the fourth period within that day. If two rallies receive the same periods, the one with lower i/n receives a lower period if the lower period is available otherwise the higher i/n receives a higher period. Whenever, such ties occurred one of the periods were available. Finally the periods in the model are calculated as *'model period'* $= 4 \cdot$ *'days before election'* + *'period within the day'*.

Table 9 shows the total number of activities that were available for 120 days before the election. These activities are categorized into groups. The category "*Rally/Event/Speech*" are of interest to this paper. The number of rallies retained after removing stronghold state rallies and counting consecutive rallies in the same state as one rally is also shown here.

[53]Not all such rallies were removed only four such rallies were removed to ensure at most four rallies in a day

Table 9: Candidates Appearances

	Romney	Obama	Trump	Clinton
	2012	2012	2016	2016
Address/Church Visit	13	15	17	17
Debate Related	12	9	8	8
Fundraiser	51	31	39	39
Interview/Meet/Discuss	12	12	34	12
Rally/Event/Speech	106	92	130	81
Stop/Tour	41	58	21	29
Travel	17	11	2	3
Number of Rallies retained	99	89	119	71
Number of Rallies dropped	7	3	11	10

Note: The table shows summary statistics for raw data obtained from Democracy in Action. Here I show the categories into which candidate appearances were categorized. This data contains classification for last 120 days rather than 100 days. The category Rally/Event/Speech is the largest category that I define as Rallies. I also show the number of rallies that were dropped and retained.

D Simulated Likelihood Procedure

D.1 Simulated Likelihood

In order to calculate $\lambda_d^\theta\left(X_d \middle| X_{d-1}\right)$ we need to execute an integration over $\mathbb{R}^{3K}$. This exercise is not feasible analytically and therefore we use a Quasi Monte-Carlo scheme that relies on Sobol sequence. We use $M = 2^{10} \times K$ points to evaluate $\lambda_d^\theta(X_d|X_{d-1})$. Lets denote the set of probability integral transforms of $3K$ dimensional Sobol sequence, till M, by $\zeta = \left\{\zeta^m = (\zeta^m_{1,1}, \ldots, \zeta^m_{1,K}, \ldots, \zeta^m_{3,1}, \ldots, \zeta^m_{3,K})\right\}_{m=1}^M$ [54]. Based on ζ we can define the following set of plausible popularity values:

$$\hat{p}^{m,d}_{1,k} = P_{d,1,k} \tag{D.1}$$

Here $P_{d,1,k}$ is the observed popularity on day d in state k. For $l = 1, 2, 3$ I define the following

$$\hat{p}^{m,d}_{l+1,k} = \alpha_R \mathbb{1}\{A_{R,d,l} == k\} + \alpha_D \mathbb{1}\{A_{D,d,l} == k\} + \tilde{\alpha}\mathbb{1}\{A_{R,d,l} == k, A_{D,d,l} == k\} + \rho \hat{p}^{m,d}_{l,k} + \delta_k + \sigma_\nu \zeta^m_{l,k} \tag{D.2}$$

Lastly call $\hat{p}^{m,d}_{5,k}$ as the predicted popularity on day d at sub period 1 for the Sobol draw m. For each draw m we can construct a predicted popularity value conditioned on $P_{d,1}, A_{d,1}, \ldots A_{d,4}$. This gives us a plausible mean for observed popularity on day $d+1$ sub-period 1. This predicted popularity is given by:

$$\hat{p}^{m,d}_{5,k} = \alpha_R \mathbb{1}\{A_{R,d,4} == k\} + \alpha_D \mathbb{1}\{A_{D,d,4} == k\} + \tilde{\alpha}\mathbb{1}\{A_{R,d,4l} == k, A_{D,d,4} == k\} + \rho \hat{p}^{m,d}_{4,k} + \delta_k \tag{D.3}$$

Therefore we have a set of plausible popularity values $\mathcal{P}_d = \left\{\hat{p}^{m,d} = (\hat{p}^{m,d}_{1,1}, \ldots, \hat{p}^{m,d}_{1,K}, \ldots, \hat{p}^{m,d}_{5,1}, \ldots, \hat{p}^{m,d}_{5,K})\right\}_{m=1}^M$ for each d and it can be used to approximate $\lambda_d^\theta\left(X_d \middle| X_{d-1}\right)$ as followed:

$$\begin{aligned}\lambda_d^\theta\left(X_d \middle| X_{d-1}\right) \approx& \hat{\lambda}_d^\theta\left(X_d \middle| X_{d-1}\right)\\ \approx& \frac{1}{M}\sum_{m=1}^M \left\{\left(\prod_{l=1}^4 \hat{\sigma}_{4(d-1)+l}\left(A_{d,l}; \hat{p}^{m,d}_l\right)\right) \times \frac{1}{\sigma_\nu^K}\left(\prod_{k=1}^K \phi\left(\frac{P_{d+1,1,k} - \hat{p}^{m,d}_{5,k}}{\sigma_\nu}\right)\right)\right\}\end{aligned} \tag{D.4}$$

Where $\hat{p}^{m,d}_l = (\hat{p}^{m,d}_{l,1}, \ldots, \hat{p}^{m,d}_{l,K})$, the function $\hat{\sigma}_{4(d-1)+l}(A_{d,l}; \hat{p}^{m,d}_l)$ has been defined in equation B.15 and $\phi(.)$ is the p.d.f. of standard normal distribution. The density $\hat{\lambda}^\theta\left(X_d \middle| X_{d-1}\right)$ provides a close approximation of $\lambda^\theta\left(X_d \middle| X_{d-1}\right)$. If ζ were drawn from a standard normal distribution instead, call this density $(\tilde{\lambda}^\theta\left(X_d \middle| X_{d-1}\right))$ then it is not hard to see that $\tilde{\lambda}^\theta\left(X_d \middle| X_{d-1}\right) \to \tilde{\lambda}^\theta\left(X_d \middle| X_{d-1}\right)$ as $M \to \infty$. The error of this integral would vanish to zero with a rate of $\sqrt{M}$. However, we are using QMC, which in practice is known to provide better convergence rate as long as the variation of $\lambda_d^\theta(.|.)$ is finite. This will be true as long as $\sigma_\nu > \frac{1}{\Delta}$ and $1 - \rho > \frac{1}{\Delta}$ for a large $\Delta >> 0$. Therefore, the approximate log-likelihood is given by:

$$\begin{aligned}\ell\ell(\theta; X_0, X_1, \ldots, X_{\bar{D}}) \approx& \hat{\ell\ell}(\theta; X_0, X_1, \ldots, X_{\bar{D}})\\ \approx& \frac{1}{\bar{D}}\sum_{d=1}^{\bar{D}} \log\left[\frac{1}{M}\sum_{m=1}^M \left\{\left(\prod_{l=1}^4 \hat{\sigma}_{4(d-1)+l}\left(A_{d,l}; \hat{p}^{m,d}_l\right)\right) \times \frac{1}{\sigma_\nu^K}\left(\prod_{k=1}^K \phi\left(\frac{P_{d+1,1,k} - \hat{p}^{m,d}_{5,k}}{\sigma_\nu}\right)\right)\right\}\right]\end{aligned} \tag{D.5}$$

D.2 Algorithm

Here I will describe the algorithm used for computing $\hat{\ell\ell}(\theta; X_0, X_1, \ldots, X_{\bar{D}})$ step-by-step. We will use the equations discussed above in sections B and D. The algorithm will be defined for a given parameter values, $\theta = \{\alpha_R, \alpha_D, \tilde{\alpha}, \rho, \sigma_\nu, \delta_1 \ldots, \delta_K, c_R, c_D, c_1, \ldots, c_K\}$ and the approximation level μ. The steps of the algorithm are below:

[54] Here $\zeta^m_{l,k}$ is $\Phi^{-1}\left(u^m_{\text{Sobol},l,k}\right)$, where $u^m_{\text{Sobol},l,k}$ is the $(3(l-1)+k)^{\text{th}}$ component of the m^{th} point of $3K$ dimensional Sobol sequence. Note that Φ^{-1} is the probability integral transform for the standard normal distribution.

Step 0 Generate the Smolyak pair, $\mathbf{U}, \mathbf{\Psi}$ for K dimensions and approximation level μ by following Judd et al. (2014). Obtain KPN Gaussian quadrature $\nu = \{(\nu_1^s, \ldots, \nu_K^s, w^s)\}_{s=1}^S$ from http://www.sparse-grids.de/ for K dimensions and approximation level $2^\mu + 1$. Generate a $3K$ dimensional Sobol sequence upto M points, call this set $\zeta = \left\{\zeta^m\right\}_{m=1}^M$ [55].

Step 1 Compute the parameter-specific $\mathbf{P}, \mathbf{T}$ using equations B.22 and B.23.

Step 2 Pre-Compute integrals of Chebyshev terms contingent on current popularity and candidate decisions by equation B.16.

Step 3 Approximate Backward Induction.

1 Carry-out the following two steps:

- Compute $\{\tilde{V}_{i,T}, \tilde{u}_{i,f,T}, \tilde{u}_{i,s,T}, \tilde{\sigma}_{i,f,T}, \tilde{\sigma}_{i,s,T}\}$ for candidate $i = R, D$ by following equations B.5, B.3, B.1,B.4, B.2 respectively.
- Obtain coefficients of the interpolating polynomials, $\{\gamma_{i,T}^V, \gamma_{i,k,T}^f, \gamma_{i,k,l,T}^s\}$ for candidate $i = R, D$ and $k, l = 0, \ldots, K$ by following equations B.7, B.8 and B.9 for period $t = T$ respectively.

2 For $t = 1, 2, \ldots, T-1$ do the following:

- Compute $\{\tilde{V}_{i,T-t}, \tilde{u}_{i,f,T-t}, \tilde{u}_{i,s,T-t}, \tilde{\sigma}_{i,f,T-t}, \tilde{\sigma}_{i,s,T-t}\}$ for candidate $i = R, D$ by following equations B.21, B.19, B.17,B.20, B.18 respectively.
- Obtain coefficients of the interpolating polynomials, $\{\gamma_{i,T-t}^V, \gamma_{i,k,T-t}^f, \gamma_{i,k,l,T-t}^s\}$ for candidate $i = R, D$ and $k, l = 0, \ldots, K$ by following equations B.7, B.8 and B.9 for period $t = T$ respectively.

Step 4 For $d = 1, 2, \ldots, \bar{D}$ do the following:

- Calculate $\hat{p}_{l,k}^{m,d}$ for $k = 1, \ldots, K$, $l = 1, \ldots, 5$ and $m = 1, \ldots, M$ using equations D.1, D.2 and D.2.
- For each $m \in \{1, 2, \ldots, M\}$ and $l \in \{1, 2, \ldots, 5\}$ calculate $\hat{\sigma}_{4(d-1)+1}(A_{d,l}; \hat{p}_l^{m,d})$, where $\hat{p}_l^{m,d} = (\hat{p}_{l,1}^{m,d}, \ldots, \hat{p}_{l,K}^{m,d})$, using equation B.15.
- Calculate $\hat{\lambda}_d^\theta \left(X_d \middle| X_{d-1}\right)$ using equation D.4.

Step 5 Calculate $\hat{\ell\ell}\left(\theta; X_0, X_1, X_2, \ldots, X_{\bar{D}}\right)$ using equation D.5.

D.3 Monte Carlo

This section demonstrates the Monte-Carlo performance of the estimator in the previous sub-section. Here I use the total electoral college votes to be equal 540. For each exercise I generate the data while considering approximation level μ_{dgp} and for the log likelihood the approximation level will be μ_{ll}. This is to allow for poorer approximation by the likelihood than the DGP which holds true in the real world. The electoral college votes are distributed similar to the real data wherever applicable. I document bias, MSE, proportional bias and proportion MSE wherever possible

[55] Recall $\zeta^m = (\zeta_{1,1}^m, \ldots, \zeta_{1,K}^m, \ldots, \zeta_{3,1}^m, \ldots, \zeta_{3,K}^m)$

Table 10: Monte Carlo for $K = 2$ when $\mu_{dgp} = 4 > \mu_{ll} = 3$

Parameter	value	bias	MSE	$\frac{\text{bias}}{\text{value}}$	$\frac{\text{MSE}}{\text{value}^2}$	μ_{dgp}
α_R	0.0600	0.0002	0.0001	0.0027	0.0297	4
	0.0600	-0.0001	0.0002	-0.0022	0.0443	5
α_D	-0.0600	-0.0017	0.0001	0.0278	0.0329	4
	-0.0600	-0.0024	0.0002	0.0402	0.0529	5
ρ	0.9900	0.0089	0.0001	0.0090	0.0001	4
	0.9900	0.0070	0.0001	0.0071	0.0001	5
c_R	2.8000	0.1336	0.0745	0.0477	0.0095	4
	2.8000	0.3818	0.1999	0.1364	0.0255	5
c_D	2.8000	0.1582	0.0745	0.0565	0.0095	4
	2.8000	0.4218	0.2318	0.1507	0.0296	5
σ	0.0600	-0.0009	0.0000	-0.0148	0.0027	4
	0.0600	-0.0004	0.0000	-0.0067	0.0026	5

Notes: This accounts for bias arising due to numerical approximation since $\mu_{dgp} > \mu_{ll}$. Number of Monte Carlo simulations is 400.

Table 11: Monte Carlo for $K = 2$ when $\mu_{dgp} = \mu_{ll}$

Parameter	value	bias	MSE	$\frac{\text{bias}}{\text{value}}$	$\frac{\text{MSE}}{\text{value}^2}$	Days
α_R	0.06000	-0.00101	0.00010	-0.01680	0.02862	D=100
	0.06000	-0.00035	0.00006	-0.00576	0.01535	D=200
α_D	-0.06000	0.00159	0.00011	-0.02657	0.03028	D=100
	-0.06000	0.00074	0.00005	-0.01231	0.01461	D=200
ρ	0.99000	-0.00134	0.00001	-0.00136	0.00001	D=100
	0.99000	-0.00100	0.00001	-0.00101	0.00001	D=200
c_R	2.40000	0.02011	0.02827	0.00838	0.00491	D=100
	2.40000	0.01388	0.01343	0.00578	0.00233	D=200
c_D	2.40000	0.00616	0.02923	0.00257	0.00507	D=100
	2.40000	0.00748	0.01159	0.00312	0.00201	D=200
σ	0.06000	0.00006	0.00001	0.00094	0.00264	D=100
	0.06000	0.00028	0.00000	0.00470	0.00128	D=200

Notes: This doesn't accounts for bias arising due to numerical approximation since $\mu_{dgp} = \mu_{ll}$. Number of Monte Carlo simulations is 400.

Table 12: Monte Carlo for $K = 4$ when $\mu_{dgp} > \mu_{ll}$

Parameter	value	bias	MSE	$\frac{\text{bias}}{\text{value}}$	$\frac{\text{MSE}}{\text{value}^2}$
α_R	0.0600	0.0057	0.0001	-0.0955	0.0417
α_D	-0.0600	-0.0056	0.0002	-0.0941	0.0432
ρ	0.9900	-0.0073	0.0001	0.0074	0.0001
c_R	3.5000	-0.2660	0.1298	0.0760	0.0106
c_D	3.5000	-0.2613	0.1283	0.0747	0.0105
σ	0.0600	0.0005	0.0000	-0.0085	0.0014
d_1	-0.0010	-0.0016	0.0000	-1.6269	17.5230
d_2	0.0010	0.0019	0.0000	-1.9447	15.8239
d_3	0.0000	0.0001	0.0000	-Inf	Inf
d_4	0.0000	-0.0001	0.0000	Inf	Inf
c_1	0.3000	0.1202	0.1332	-0.4008	1.4797
c_2	0.3000	0.1208	0.1339	-0.4027	1.4882
c_3	0.3000	-0.0259	0.1044	0.0863	1.1599

Notes: This accounts for bias arising due to numerical approximation since $\mu_{dgp} > \mu_{ll}$. Number of Monte Carlo simulations is 400.

Table 13: Monte Carlo Experiments for $K = 2$, $D = 100$ $\mu_{dgp} > \mu_{ll}$
Interaction term specification

Parameter Configuration 1					
Parameter	Value	Bias	MSE	Bias (prop.)	MSE (prop.)
α_D	-0.060	0.004	0.000	0.062	0.083
α_R	0.050	-0.009	0.000	0.188	0.145
$\tilde{\alpha}$	0.000	0.015	0.003	-	-
c_D	2.600	0.046	0.185	0.018	0.027
c_R	2.600	0.120	0.140	0.046	0.021
ρ	0.990	0.003	0.000	0.003	0.000
σ_v	0.060	0.001	0.000	0.018	0.006
Parameter Configuration 2					
Parameter	Value	Bias	MSE	Bias (prop.)	MSE (prop.)
α_D	-0.040	-0.001	0.000	0.014	0.063
α_R	0.050	0.001	0.000	0.017	0.026
$\tilde{\alpha}$	-0.020	-0.006	0.000	0.321	0.285
c_D	2.600	-0.005	0.063	0.002	0.009
c_R	2.600	-0.140	0.124	0.054	0.018
ρ	0.990	0.004	0.000	0.004	0.000
σ_v	0.060	0.002	0.000	0.037	0.005
Parameter Configuration 3					
Parameter	Value	Bias	MSE	Bias (prop.)	MSE (prop.)
α_D	-0.020	0.004	0.000	0.176	0.183
α_R	0.050	-0.005	0.000	0.102	0.075
$\tilde{\alpha}$	-0.040	-0.003	0.000	0.063	0.169
c_D	2.600	-0.174	0.097	0.067	0.014
c_R	2.600	-0.168	0.123	0.065	0.018
ρ	0.990	0.003	0.000	0.003	0.000
σ_v	0.060	0.002	0.000	0.028	0.003

Notes: This accounts for bias arising due to numerical approximation since $\mu_{dgp} > \mu_{ll}$. Number of Monte Carlo simulations is 400. The specification includes the interaction term. The take-away from the table is bad Monte-Carlo performance for the range of parameters that I find in the data when using this specification

E Additional Robustness Tests

E.1 Robustness to Popularity Definition and Calibrated Parameters

E.1.1 Raw Polls and $E = 157$

Here I estimate the model after relaxing two assumptions. A first assumption is that popularity is defined as the deviation of poll margins from the aggregate mean of poll margins across all state groups and days. I estimate the model by considering the raw polls directly. A second assumption is that the terminal pay-offs are given by the total electoral college votes the U.S. has. I set the electoral pay-offs as 157 electoral college votes. This is the total number of electoral college votes states in my sample have.

I display the results from this exercise in columns (1) and columns(2) in the Table 14. Here rally effectiveness of all candidates is significant. The point estimates of effectiveness for rallies do not significantly change from the baseline.

E.1.2 De-meaned Polls and $E = 157$

Here I estimate the model by retaining the de-meaned poll margins, but I set the electoral pay-offs as 157 electoral college votes. This is the total number of electoral college votes states in my sample have. The results from this exercise are shown in columns (3) and columns(4) in Table 14. Here rally effectiveness of all candidates is significant. The point estimates of effectiveness for rallies do not significantly change from the baseline.

E.2 First Mover Probability

I calibrate the f at 0.5 in the baseline specification. This choice is to ensure that both candidates have the same informational advantage over each other. In this game a second mover observes the first mover's action and therefore makes a more informed decision. This implication ensures that the second mover has a more advantageous position over the first mover exogenously. By choosing a value of $f = 0.5$, I ensure that the game is ex-ante fair along this dimension to both the players.

Here, I test the sensitivity of my estimates to this assumption. For this purpose, I re-estimate the model for the values of $f = 0.33$ and $f = 0.67$. In the first case, R has the informational advantage over the second mover while in the second case D has an information advantage. From the estimation exercise I find that my estimates of rally effectiveness do not change significantly. The results $f = 0.33$ are given in columns (5) and (6) of Table 14. The results for $f = 0.67$ are given in columns (7) and (8) of Table 14.

Table 14: Additional Robustness Tests

	Raw Polls $E = 157$		De-meaned Polls $E = 157$		First Mover Prob $f = 0.33$		First Mover Prob $f = 0.67$		"Most Swing" State Polls Within Group	
Parameters	2012 (1)	2016 (2)	2012 (3)	2016 (4)	2012 (5)	2016 (6)	2012 (7)	2016 (8)	2012 (9)	2016 (10)
α_R	0.106	0.0906	0.101	0.0867	0.0741	0.0835	0.0732	0.0841	0.105	0.0834
	0.027	0.0197	0.0265	0.0228	0.0153	0.0154	0.0176	0.0155	0.0218	0.0145
α_D	-0.0231	-0.0776	-0.0354	-0.0768	-0.0506	-0.0743	-0.0652	-0.0748	-0.0613	-0.0809
	0.0113	0.0165	0.013	0.0217	0.0144	0.0151	0.018	0.0153	0.0155	0.0129
ρ	0.991	0.989	0.99	0.988	0.988	0.991	0.989	0.991	0.983	0.991
	0.002	0.003	0.002	0.002	0.002	0.001	0.002	0.001	0.003	0.001
σ	0.146	0.16	0.146	0.161	0.147	0.16	0.147	0.16	0.21	0.201
	0.0142	0.0149	0.0143	0.0151	0.014	0.0148	0.0141	0.0148	0.0171	0.0166
c_R	2.46	2.03	2.43	2.01	2.76	2.36	2.69	2.36	2.78	2.4
	0.232	0.204	0.237	0.212	0.266	0.208	0.263	0.208	0.276	0.21
c_D	2.59	2.86	2.59	2.85	2.86	3.26	2.9	3.26	2.84	3.28
	0.174	0.248	0.175	0.251	0.194	0.259	0.191	0.259	0.198	0.264
Fixed Effects:										
Cost	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Poll Margins	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
LL	-681.95	-676.93	-681.3	-678.44	-657.03	-654.57	-658.74	-654.53	-782.57	-745.64
Observations	100	100	100	100	100	100	100	100	100	100

[a] Note: The table shows estimates for model parameters under 6 modifications. Columns (1) and (2) consider raw poll margins as popularity and $E = 157$ as electoral pay-off. Columns (3) and (4) consider de-meaned poll margins as popularity and $E = 157$ as electoral pay-off. Here the standard errors have been computed by using observation wise gradient and likelihood hessian. I use HAC estimation for this purpose. Columns (5) and (6) consider a different calibrated value for first mover probability, $f = 0.33$. This values gives R a second mover advantage or information advantage over the opponent. Columns (7) and (8) consider a different calibrated value for first mover probability, $f = 0.67$. This values gives R a first mover disadvantage or information disadvantage over D. Columns (9) and (10) consider an alternative definition of state group popularity. Here candidates consider the most swing state's polls instead of weighted average.

E.3 Poll Aggregation Measure

I use a weighted mean for the aggregating the polls at the state level. One concern is that I am combining a moderately R-leaning state with a moderately D-leaning state which essentially leads to creating a group which appears more purple[56] than it should be otherwise. To test how much of this aggregation step may distort my estimates, I propose a different aggregation measure. Here I pick the polls of the state which is consistently most purple within the group. Formally,

$$\tilde{l}_k = \arg\min_{l \in G_k} \left\{ \frac{1}{D} \sum_{d=1}^{D} (P_{ld})^2 \right\} \tag{E.1}$$

Here the term G_k denotes the set of states in group k. The daily polls for state l are denoted by P_{ld}. The state $\tilde{l}_k$ will be the most purple state within each group. Then I define the state group poll-margin as $P_{kd} = P_{\tilde{l}_k d}$. The model is estimated after this step and the results from this exercise are reported in columns (9) and (10) of Table 14.

[56]For the lack of a good term, I am using "purple". By this I mean the easiness of swinging the state/state group.

F Additional Figures

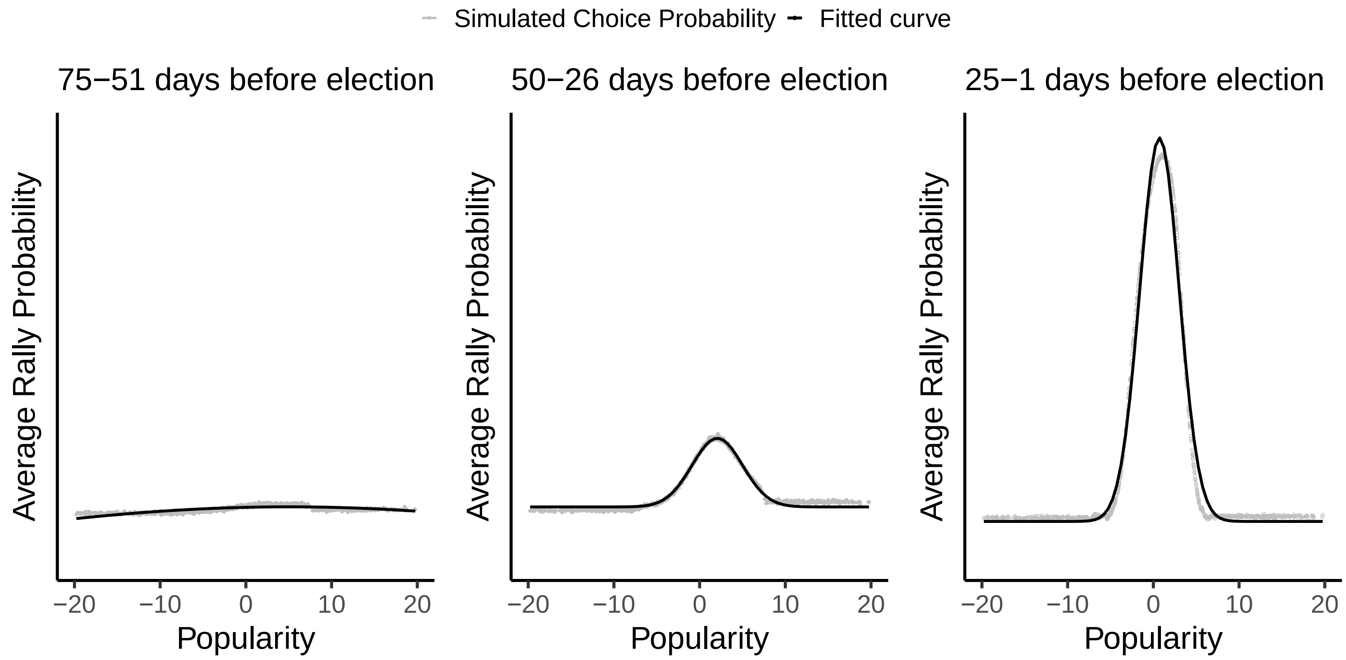


Figure 12: *One state case:* $K = 1$ This figure shows that the second empirical pattern is supported by the model even if there is one state.

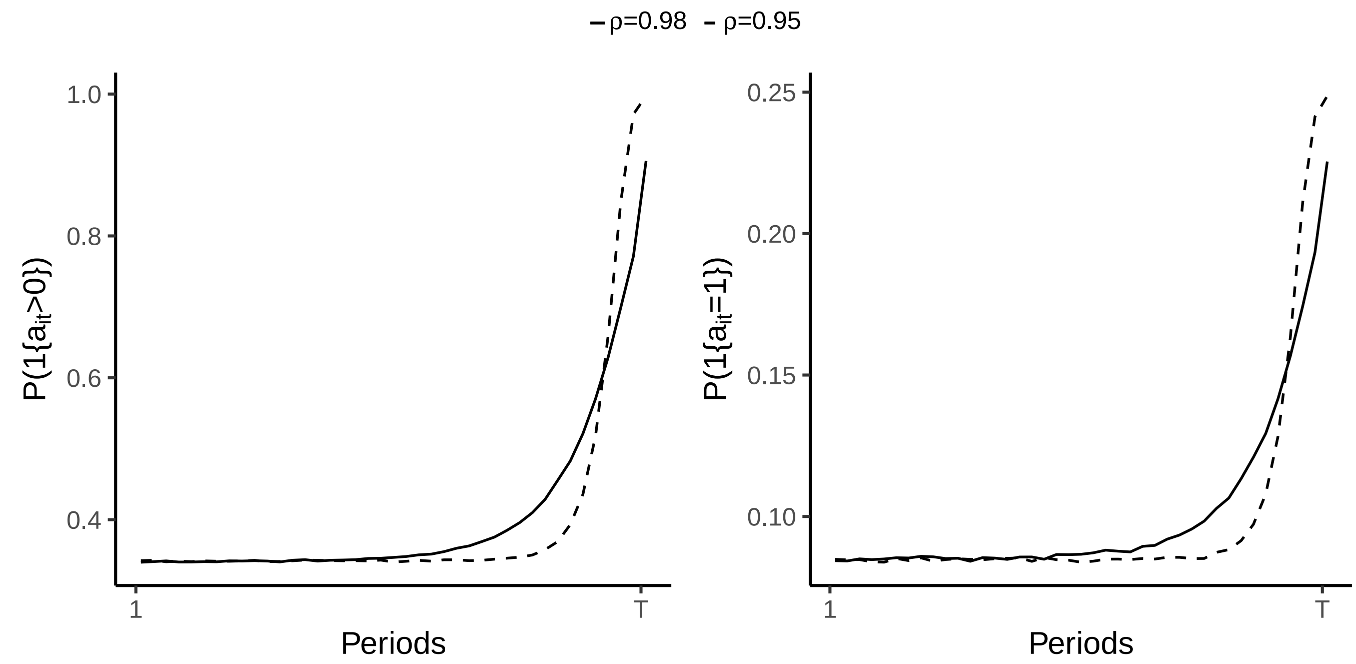


Figure 13: *The plot demonstrates increase in earlier rallies due to an increase in persistence parameter, ρ. The increase in early rallies can be explained by the fact that at any given period t, rally effectiveness on election day popularity, ceteris paribus, is given by $\rho^{T-t}\alpha_R$. This effectiveness increases with ρ, and therefore, returns from rallying has a higher effect.*

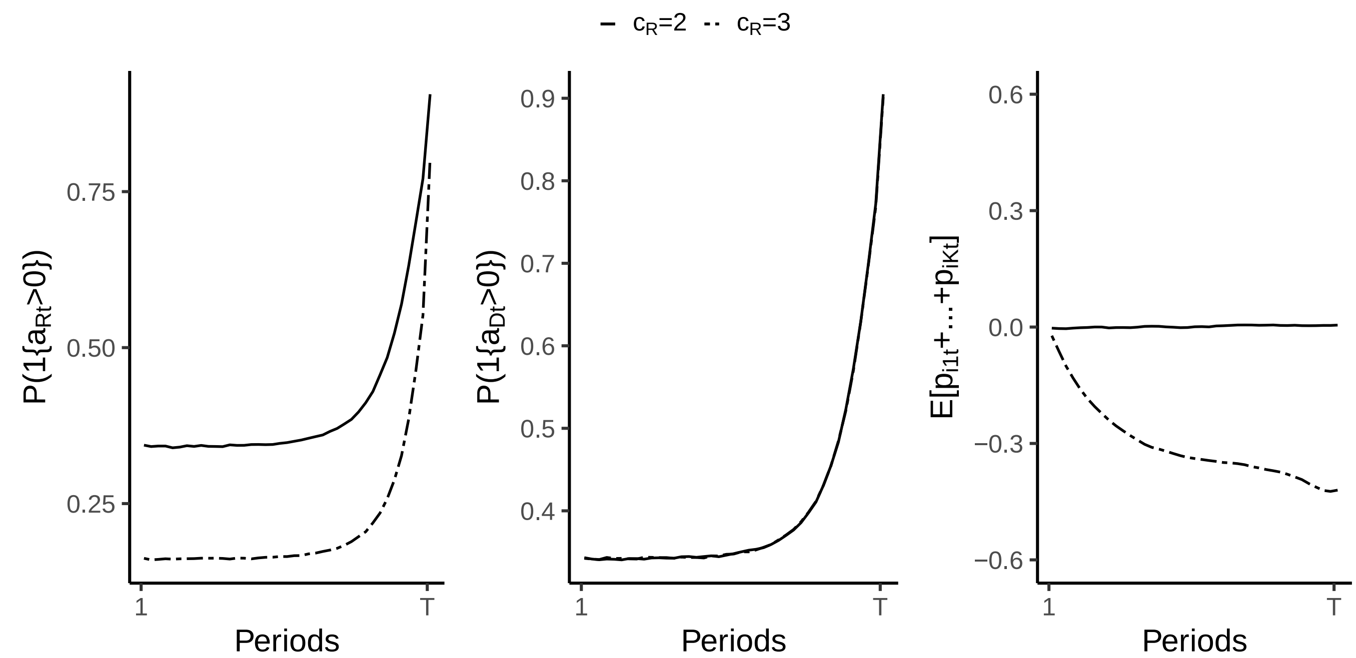


Figure 14: *The plot demonstrates level shift in probability of rally, for both candidates and also changes in expected popularity due to increase in cost of rallying, c_R. The level shift can be expplained by proposition A.4 which shows that in earlier periods the gain from rallying is negligible and therefore probability of rally is determined by cost of rallying alone. Higher cost leads to a decrease in probability of rally. The change in expected popularity is due to decreased rallying by R.*

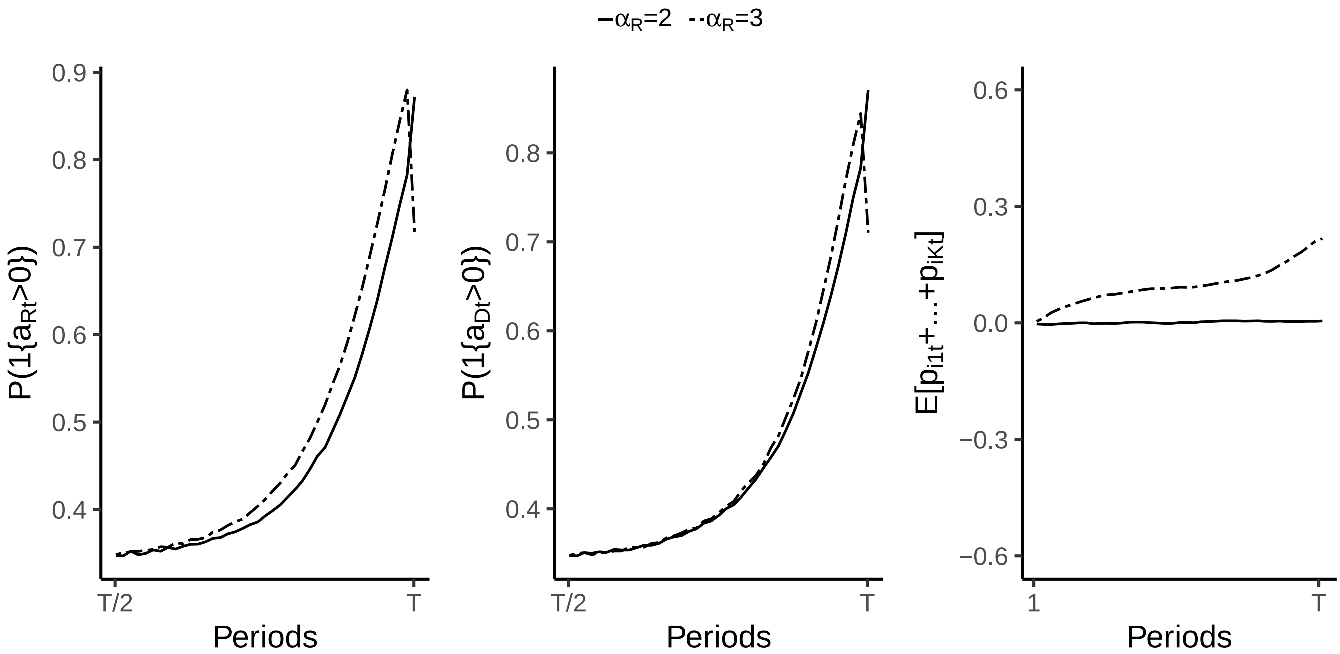
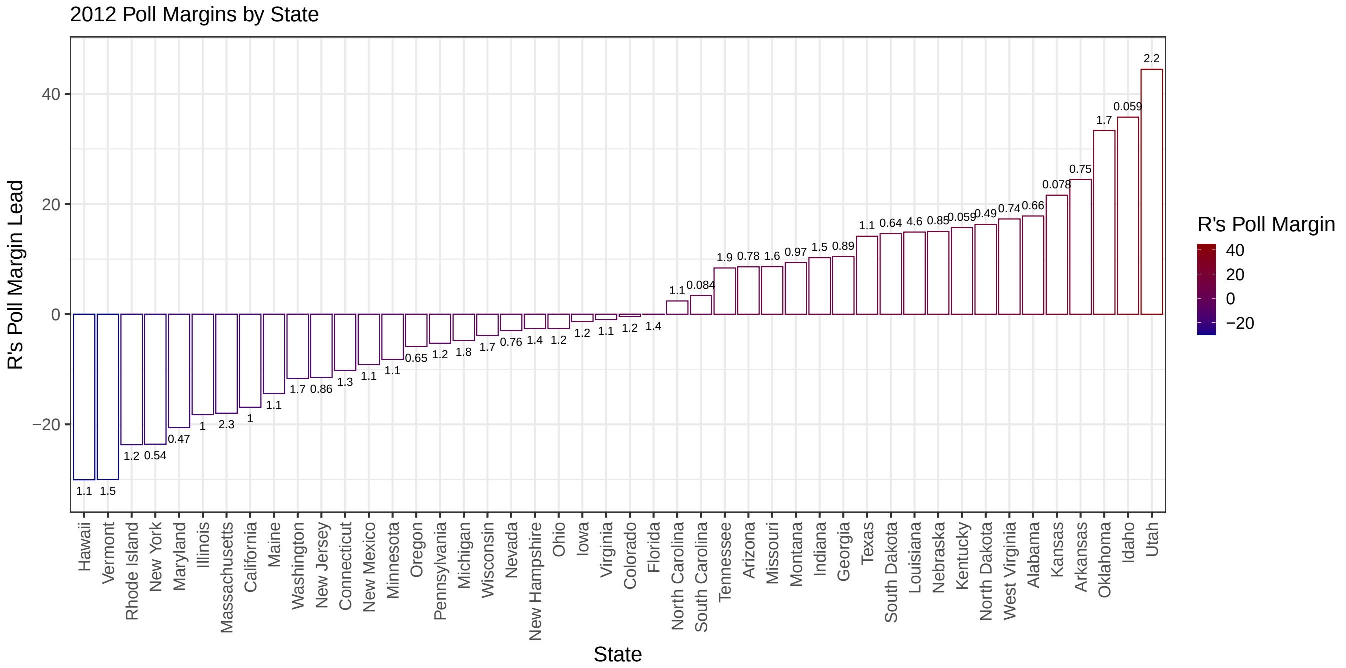
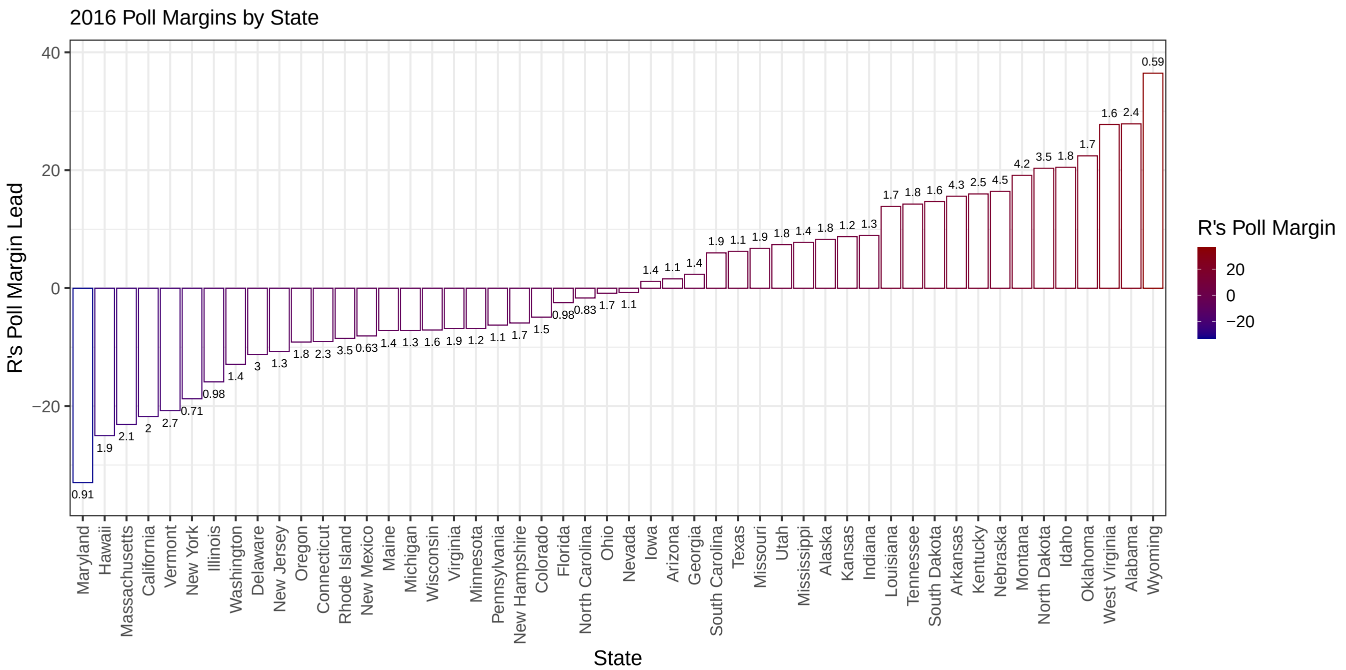


Figure 15: *The plot demonstrates changes in probability of rally, for both candidates, and expected popularity due to increase in rally effectiveness, α_R. Higher effectiveness increases returns from rallying therefore rallying increase. The change in expected popularity is in R's favor as his rallies are more effective. The drop in rallying at the end can be explained by the fact that popularity is less pivotal towards the end than in the case with symmetric effectiveness. Therefore, due to the bell-shaped nature of rally probability curve, there is a drop in rallies.*



(a) *2012 R's poll margin lead distribution across states*



(b) *2016 R's poll margin lead distribution across states*

Figure 16: *This Figure shows mean and standard deviation of R's poll margin for all states. For this figure, I have removed District of Columbia, NE-1, NE-2, NE-3, ME-1, ME-2. The means are shown on the y-axis and the standard deviation are shown on top/bottom of the state specific bars.*

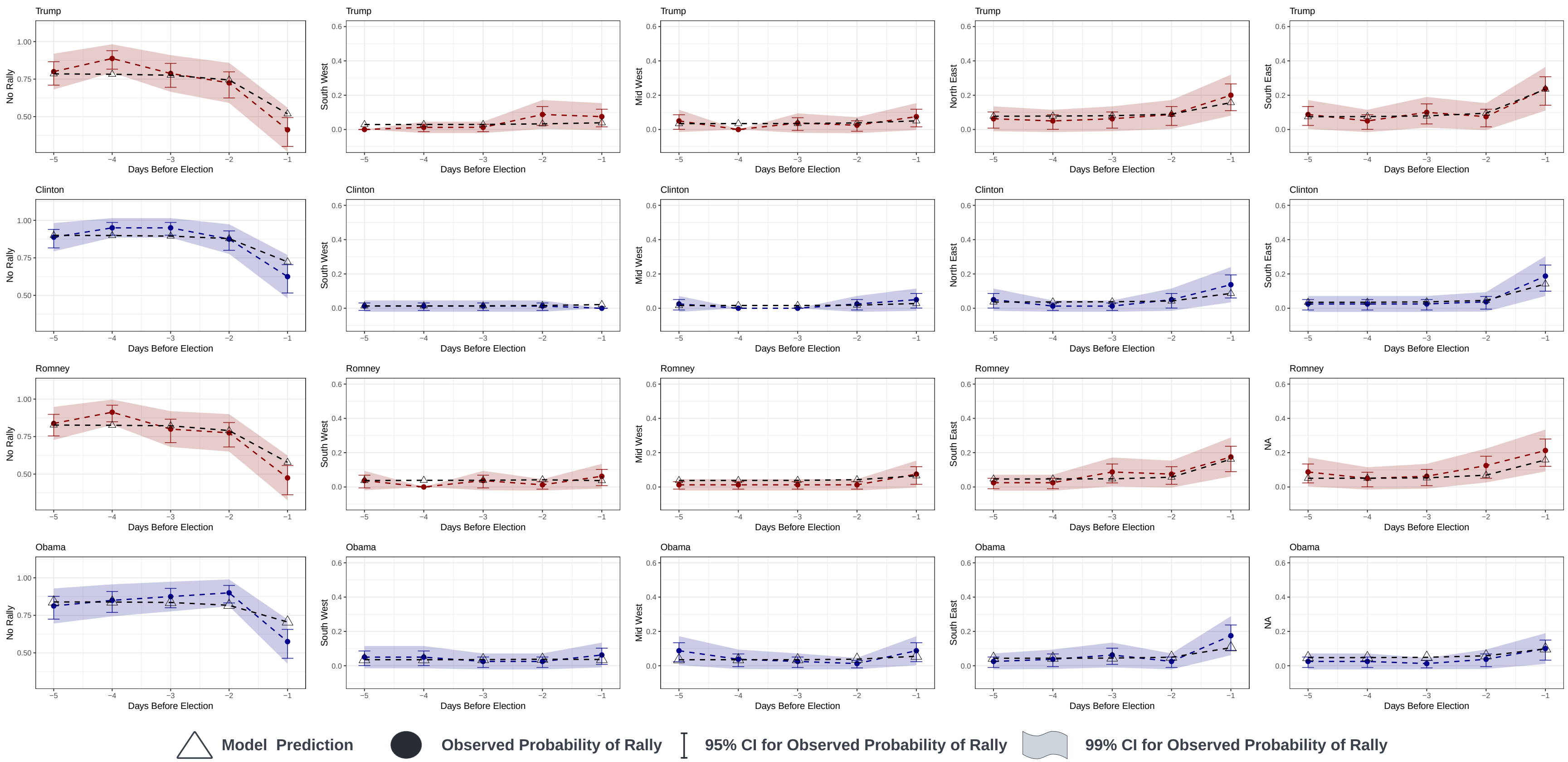


Figure 17: *This figure shows the cumulative effect a candidate's rallies had on their vote margin lead and winning probability. For each candidate, first I draw 400 draws parameter values from the asymptotic distribution of the model parameter estimates. Then for each draw I simulate the model outcomes for the cases of (i) only the candidate rallies and (ii) none rally. Then I take the differences of these outcomes across (i) and (ii). The variance of the distribution of these differences are used to formulate the confidence intervals.*

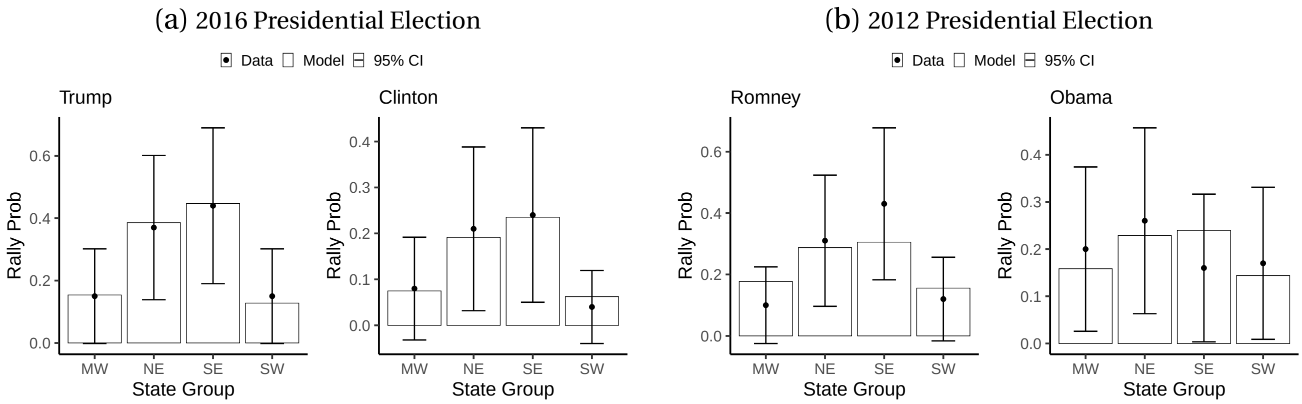


Figure 18: *This figure plots the model performance within the sample. Here, I plot the predicted probability of rally for each state group against observed probability of rally. The columns represent the model prediction, the solid points show observed probability of rallies for these groups and error-bars represent the 95% confidence intervals. Most of the predicted values lie in 95% confidence intervals which gives a sense of model performance within sample fit.*

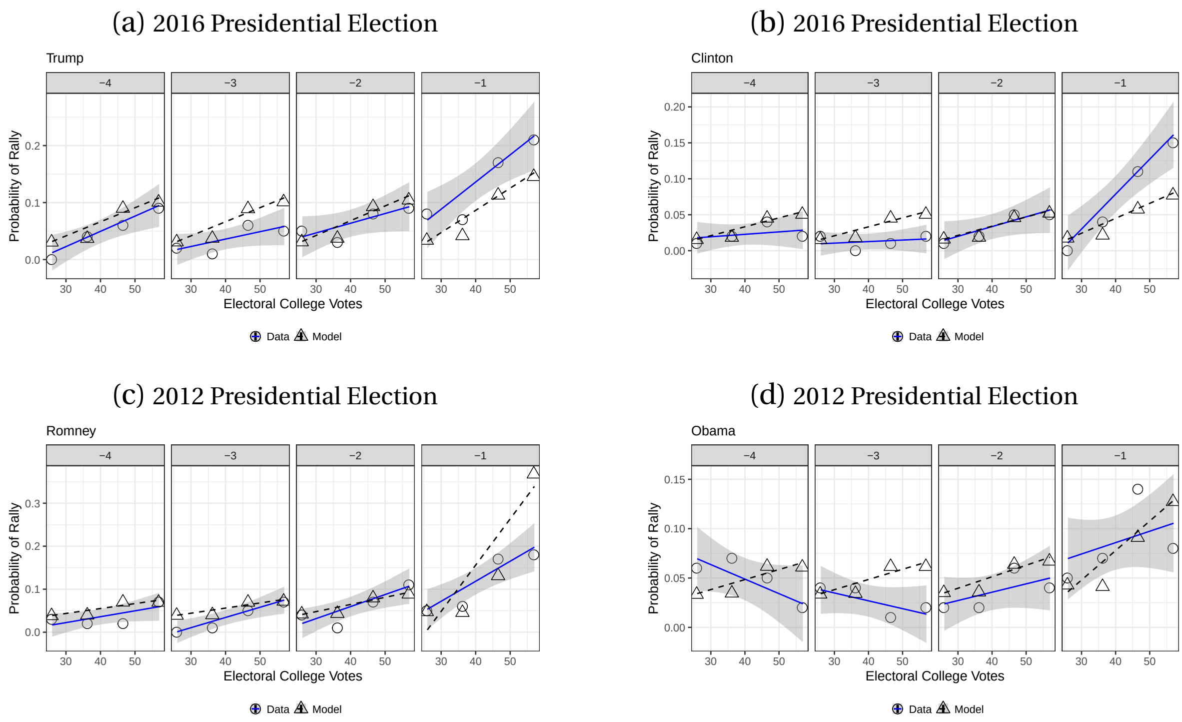


Figure 19: *This figure presents that the model supports the increasing correlation between rallies and electoral college vote pattern. The bin* -4 *corresponds to* $100-76$, -3 *corresponds to* $75-51$, -2 *corresponds to* $50-26$, *and finally* -1 *corresponds to* $25-1$ *days before election. For each candidate and each bin I provide a bin scatter plot along with a fitted line for the data and the model. The blue line, circular points and the confidence regions correspond to the data and the black lines and the triangle points correspond to the model.*

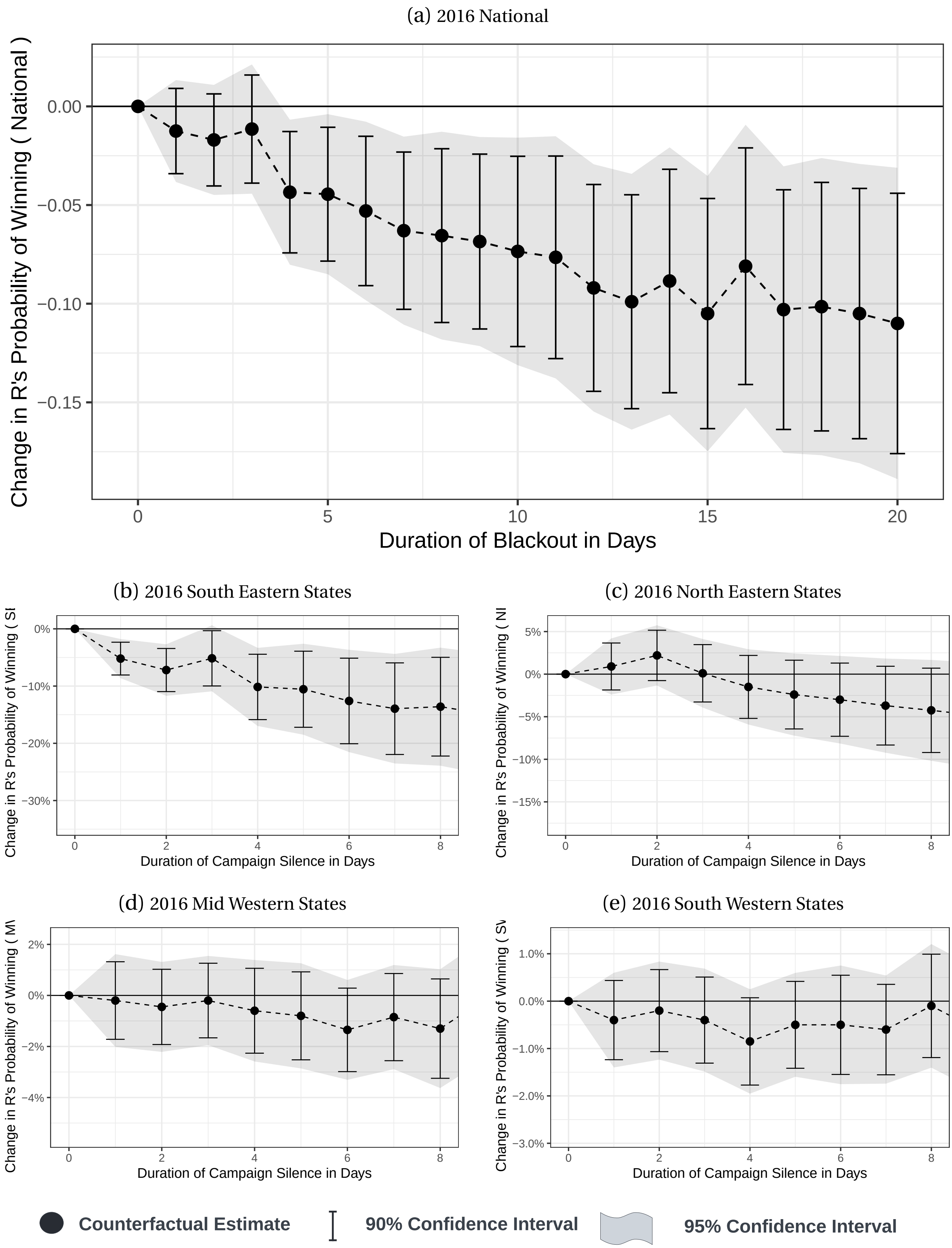


Figure 20: *Blackout Duration and Electoral Outcome for 2012 Presidential Elections. This figure provides estimates for changes in electoral outcomes when blackouts of varying duration lengths are imposed. For each blackout duration, I calculate the R's probability of winning along with the corresponding confidence intervals for these probabilities.*

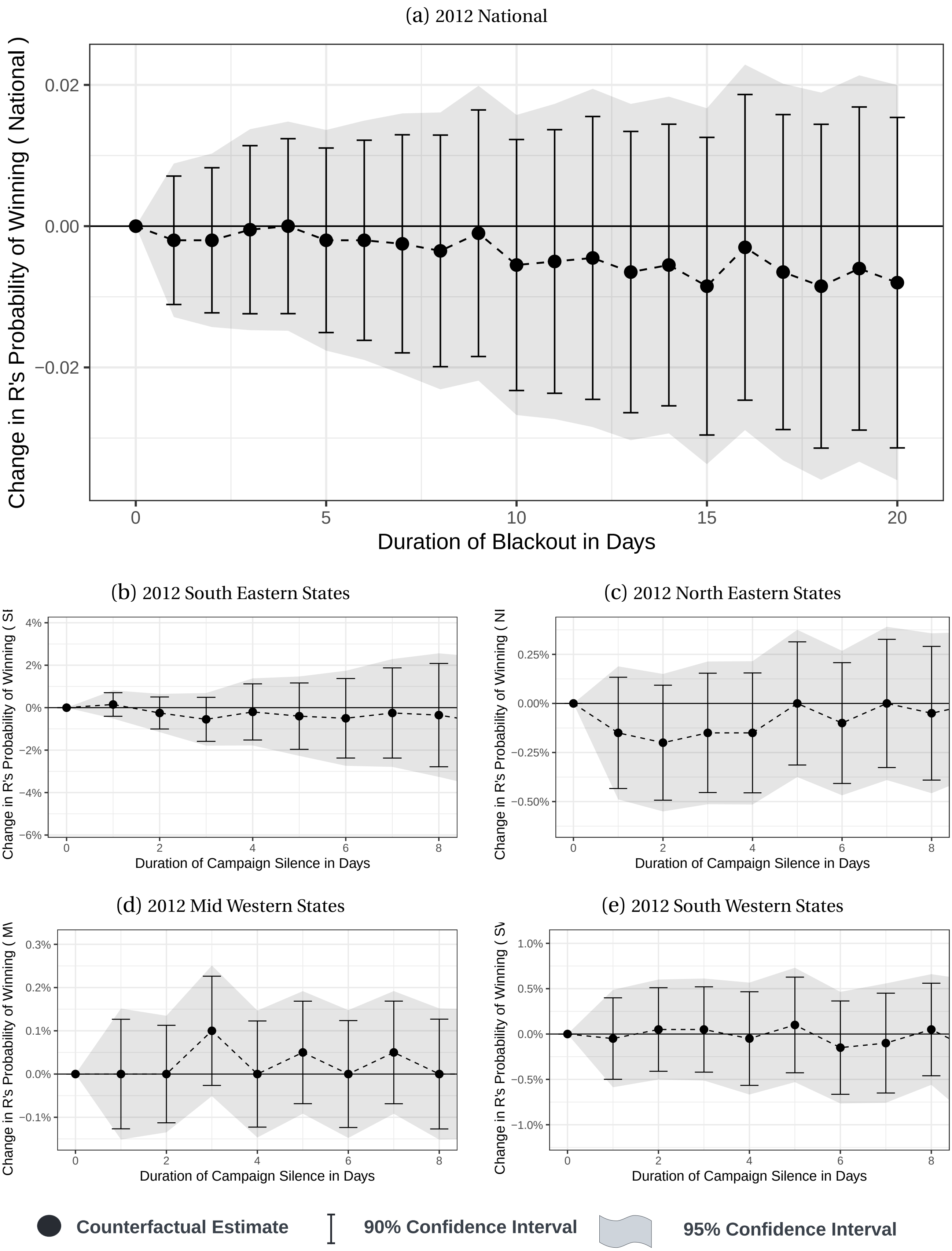


Figure 21: *Blackout Duration and Electoral Outcome for 2012 Presidential Elections. This figure provides estimates for changes in electoral outcomes when blackouts of varying duration lengths are imposed. For each blackout duration, I calculate the R's probability of winning along with the corresponding confidence intervals for these probabilities.*

G Additional Tables

G.1 Reduced form evidence for rally and poll margin relation

Table 15: Estimates for Regression G.1

Dependent Variable:	$A_{i,d,k,y}$		
Model:	Full Sample (1)	Only Republican (2)	Only Democrat (3)
Variables			
$\mathbb{1}\{-80 \leq d \leq -61\}$	-0.010***	-0.011**	-0.009**
	(0.004)	(0.005)	(0.005)
$\mathbb{1}\{-60 \leq d \leq -41\}$	0.002	0.003	-0.002
	(0.003)	(0.004)	(0.005)
$\mathbb{1}\{-40 \leq d \leq -21\}$	0.010**	0.012*	0.008
	(0.004)	(0.006)	(0.005)
$\mathbb{1}\{-20 \leq d \leq -1\}$	0.043***	0.047***	0.039***
	(0.009)	(0.013)	(0.012)
$P_{i,d-1,k,y}$	6.92×10^{-5}	0.002**	-0.002**
	(0.0006)	(0.0010)	(0.0009)
$P^2_{i,d-1,k,y}$	1.01×10^{-5}	-9.94×10^{-6}	-7.74×10^{-6}
	(1.1×10^{-5})	(2.14×10^{-5})	(1.86×10^{-5})
$P_{i,d-1,k,y} \times \mathbb{1}\{-80 \leq d \leq -61\}$	1.39×10^{-5}	7.44×10^{-6}	1.25×10^{-5}
	(5.71×10^{-5})	(8.65×10^{-5})	(6.53×10^{-5})
$P_{i,d-1,k,y} \times \mathbb{1}\{-60 \leq d \leq -41\}$	-6.27×10^{-6}	-2.7×10^{-5}	4.21×10^{-5}
	(5.39×10^{-5})	(9.05×10^{-5})	(7.55×10^{-5})
$P_{i,d-1,k,y} \times \mathbb{1}\{-40 \leq d \leq -21\}$	-4.54×10^{-5}	-2.82×10^{-5}	-4.44×10^{-5}
	(8.83×10^{-5})	(0.0001)	(0.0001)
$P_{i,d-1,k,y} \times \mathbb{1}\{-20 \leq d \leq -1\}$	-4.1×10^{-5}	-0.0002	0.0002
	(0.0002)	(0.0002)	(0.0002)
$P^2_{i,d-1,k,y} \times \mathbb{1}\{-80 \leq d \leq -61\}$	1.42×10^{-5}***	1.67×10^{-5}**	1.46×10^{-5}**
	(5.35×10^{-6})	(8.24×10^{-6})	(7.23×10^{-6})
$P^2_{i,d-1,k,y} \times \mathbb{1}\{-60 \leq d \leq -41\}$	-2.26×10^{-6}	-3.01×10^{-6}	4.75×10^{-6}
	(5.01×10^{-6})	(6.58×10^{-6})	(7.08×10^{-6})
$P^2_{i,d-1,k,y} \times \mathbb{1}\{-40 \leq d \leq -21\}$	-2.09×10^{-5}**	-1.99×10^{-5}	-1.43×10^{-5}
	(9.54×10^{-6})	(1.27×10^{-5})	(1.04×10^{-5})
$P^2_{i,d-1,k,y} \times \mathbb{1}\{-20 \leq d \leq -1\}$	-5.34×10^{-5}***	-5.32×10^{-5}**	-4.45×10^{-5}**
	(1.48×10^{-5})	(2.03×10^{-5})	(1.86×10^{-5})
Fixed-effects			
$i \times k \times y$	Yes	Yes	Yes
Fit statistics			
Observations	18,274	9,137	9,137
R^2	0.09332	0.09614	0.09008
Within R^2	0.00847	0.00922	0.00854

Clustered ($i \times k \times y$) standard-errors in parentheses
*Significant Codes: ***: 0.01, **: 0.05, *: 0.1*

Consider the following specification

$$A_{i,d,k,y} = \sum_{s=1}^{5} \Big(\beta_{0,s} B_s(d) + \beta_{1,s} P_{i,d-1,k,y} B_s(d) + \beta_{2,s} P^2_{i,d-1,k,y} B_s(d)\Big) + \gamma_{i,k,y} + \epsilon_{i,d,k,y} \tag{G.1}$$

Here, i is a candidate, d is day, k is a state and y is a year. Moreover, $B_s(d) = \mathbb{1}\{-120 + 20s \leq d \leq -101 + 20s\}$. Note for $s = 1$ this is $\mathbb{1}\{-100 \leq d \leq -81\}$, for $s = 5$ this is $\mathbb{1}\{-20 \leq d \leq -1\}$. Poll margin is denoted by $P_{i,d,k,y}$ and $\gamma_{i,k,y}$ represent the candidate i, state k and year y fixed effect. Finally $\epsilon_{i,d,k,y}$ is the error term. The estimates for this regression are given in the table G.1. In this table it important to note that the coefficient on $P^2_{i,d-1,k,y} \times B_s$ becomes more negative as s increases from 1 to 5. This supports the model prediction for the hump shaped[57] relation exists only close to election.

[57]bell shaped relation in the main body of the paper

G.2 In-Sample Model Fit

Table 16: In Sample Model Fit

Panel (A): Comparison of Means

	Romney		Obama		Trump		Clinton	
	Model	Data	Model	Data	Model	Data	Model	Data
South West	0.158	0.12	0.132	0.17	0.128	0.15	0.0625	0.04
		0.069		0.081		0.076		0.04
Mid West	0.161	0.1	0.136	0.2	0.154	0.15	0.0748	0.08
		0.063		0.088		0.076		0.056
North East	0.313	0.31	0.256	0.26	0.386	0.37	0.191	0.21
		0.11		0.099		0.12		0.09
South East	0.329	0.43	0.264	0.16	0.447	0.44	0.235	0.24
		0.12		0.079		0.13		0.096

Panel (B): Measures of Fit

	Romney	Obama	Trump	Clinton
Correlation	0.7076	0.7530	0.6950	0.8378
Mean Squared Error	0.3996	0.3466	0.4138	0.2385
Correct Predictions	0.7600	0.8025	0.7375	0.8600

[a] This table shows the in-sample model fit. The average number of rallies per day lie in 95% confidence intervals of the observed in the data. The worst correlation is 0.69. For each period, I define prediction as the option with the highest probability of choosing. I compare these predictions with the data and calculate the proportion of correct predictions. Using this metric for prediction I find that worst correct predictions is 73%.

G.3 Cumulative Effect of Rallies: State-Wise Effects

Table 17: Effect of Rallies on Electoral Outcomes

	2012		2016	
	Romney	Obama	Trump	Clinton
Voter Share				
National	294.525***	93.0554***	483.896***	92.6859***
	(67.125)	(20.739)	(94.296)	(23.692)
South West	19.4427***	9.00424**	20.0117***	3.91761
	(4.9946)	(3.684)	(6.2449)	(2.9525)
Mid West	32.0433***	14.518***	33.2474***	9.62969**
	(9.6481)	(5.1558)	(11.221)	(4.6813)
North East	87.055***	34.3326***	128.058***	43.5479***
	(27.028)	(7.4425)	(32.869)	(9.4602)
South East	155.984***	35.2006***	302.58***	35.5907**
	(44.464)	(12.289)	(71.062)	(14.467)
Probability of Winning				
National	0.0065	5e-04	0.4025***	0.036
	(0.0211)	(0.0027)	(0.118)	(0.0458)
South West	0	0.002	0.0155	-0.002
	(0.00856)	(0.0043)	(0.0176)	(0.00864)
Mid West	0	0	0.05	0.0095
	(0.00213)	(0.000916)	(0.0487)	(0.0156)
North East	0.006	0.001	0.357***	0.0865**
	(0.0205)	(0.00217)	(0.0755)	(0.0377)
South East	0.015	0.0135	0.4945***	0.051
	(0.0307)	(0.019)	(0.0962)	(0.0348)

[a] Note: This table shows the cumulative effect a candidate's rallies had on their vote margin lead and winning probability. For each candidate, first I draw 400 draws parameter values from the asymptotic distribution of the model parameter estimates. Then for each draw I simulate the model outcomes for the cases of (i) only the candidate rallies and (ii) none rally. Then I take the differences of these outcomes across (i) and (ii). The variance of the distribution of these differences are used to formulate the confidence intervals.