Rally the Vote: Electoral Competition with Direct Campaign Communication

Anubhav Jha*

Abstract

Political rallies constitute a large part of electoral campaigning in the U.S. and in modern democracies since the 19th century and remain a salient politico-economic phenomenon today. This paper accounts for candidates' strategic decisions to rally as a finite-horizon dynamic game of electoral competition and applies it to structurally estimate rally spatial and temporal choices by candidates. For the 2012 and 2016 U.S. presidential elections, we show that rallies substantially increase poll margin leads in targeted constituencies over non-rallying opponents and trigger systematic dynamic responses by opponents. In terms of magnitudes, rallies by presidential candidates are more persuasive than television ads, and estimates of the gross effect show that President Trump's rallies were in fact electorally pivotal. Instead, rallies by all other candidates did not change their win probability. Counterfactual policy experiments reveal that the effects of short-term campaign silences (i.e., electoral blackouts) are limited

*Department of Politics, Princeton University. Email: aj7954@princeton.edu

I would like to thank Francesco Trebbi, Matilde Bombardini, Vitor Farinha Luz, Matias Iaryczower, and Paul Schrimpf for their comments and suggestions and the Dev/PE, Econometrics and Theory group at the Vancouver School of Economics, the University of British Columbia, Princeton University, and BPER lab at the University of California Berkeley. 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.

since candidates can time their rallies and gain sufficient support from the electorate before campaign silences begin.

JEL Codes: C73, D72, P0, M37

Keywords: Advertising, Elections, Political Economy, Stochastic and Dynamic Games

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.[1] However, it was only in the late 19th century that political rallies were first used as an electioneering tool in a large-scale election. 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). Harry Truman and Thomas Dewey later utilized this practice 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.[2] Nine of these rallies had more than 10,000 attendees. In the fall campaigns of 2012 and 2016, political rallies constituted 44.5%[3] 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 advertis-

[1] See van der Blom (2016) for details on *contios*, informal public meetings where Roman magistrates addressed the people. See Johnstone and Graff (2018) for details on *bouleutêria*, auditoriums dedicated to oratorical performances.

[2] This figure is calculated using news reports on individual rallies from multiple news providers. Complete details on the news sources for each of Trump's rally are available upon request.

[3] I used candidate calendars made available by Appleman (2012, 2016) to calculate this figure.

ing (Iaryczower et al., 2022; Spenkuch et al., 2018; Gordon and Hartmann, 2013; Hill et al., 2013; Gerber et al., 2011), 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.[4] Theoretical work is challenging due to multiple equilibria associated with such discrete action games.

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. To this end, this paper presents a game of perfect information with a finite time horizon. In this game, office-seeking candidates are subject to electoral competition while facing regional differences and dynamic uncertainty in their popularity. Dynamic uncertainty accommodates popularity shocks, i.e., unforeseen circumstances in electoral races that lead to a candidate jumping ahead of 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.

In this model, a candidate's objective is to stay popular in as many states as possible on election day. The locally popular candidate receives a payoff proportional to the state's electoral college votes. Candidates can hold rallies across states to increase their local popularity. I allow campaigning effects to decay by modeling local popularity as an AR(1) process. Decay in campaigning effects has been documented by Hill et al. (2013) and Gerber et al. (2011). The decay rate forms one of the critical components in the model and has a one-to-one relationship with the autocorrelation parameter of the AR(1) process. I call this parameter persistence in popularity. This parameter influences the urgency with which candidates

[4]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 found that Trump was more effective than his opponents in gaining support through political rallies. Due to a 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 carries over into their estimates. Moreover, the stable unit treatment value assumption (SUTVA) is harder to maintain at media market-level analysis as there can be geographical spillovers.

increase their rally frequency over time.[5]

Candidates in the model must pay a cost to hold a rally. This cost is allowed to vary across candidates and states. Candidates are assumed to move in a stochastic order, thereby ensuring that the game is one of perfect information. I further assume that candidates are equally likely to be the first or the second mover, which implies that no candidate has an ex-ante first (or second) mover advantage. The perfect information structure provides us with rally decision probabilities, which are uniquely solvable using backward induction.

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 earlier rallies. An increase in the cost of rallies introduces a downward level shift in the probability of rallying. An increase in the effectiveness parameter also leads to a higher number of rallies for each period and popularity.

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 downwards because candidates may be more likely to rally in states where they need to boost their popularity. This downward bias would make rallies appear ineffective, which is a common finding in this literature. On the other hand, in this game, factors like the contemporaneous rally decisions 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 into the rally decisions of candidates. These factors shape the expected benefit of holding a rally in a given state at a given time, thereby providing one with the required identifying variation.

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 the election approaches. This pattern holds individually for all candidates and it is predicted by the model. To estimate the model parameters, I leverage the Markov property obeyed by local popularity and prove that daily rally decisions and poll margins must also obey the Markov property.

[5] A similar intuition and pattern can be found for campaign spending in Acharya et al. (2022).

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, which is used for estimation.

This paper finds that Trump's rallies increased his poll margin lead by 0.084 pp[6] in a state, while Clinton's rallies increased her poll margin lead by 0.075 pp. For the 2012 election, I find that Romney's and Obama's rallies increased their poll margin lead by 0.1 pp and 0.04 pp, respectively. 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 TV ad. The persuasion rate of a Trump rally is 0.168%, and that of a Republican TV ad is 0.01% (Spenkuch and Toniatti, 2018). This implies that to compensate for one MAGA rally, Trump would have required 17 ad spots.

The third contribution is to estimate political rallies' electoral effects and quantify how many rallies are strategic responses to opponents. I answer these two questions by executing two counterfactual experiments. The first counterfactual experiment focuses on the cumulative effect of rallies on electoral outcomes. The cumulative effect of rallies is due to the fact that candidates hold multiple rallies in each state. Cumulative effects differ from contemporaneous effects on popularity since the effects of a rally decay with time. These two forces are in tension, and their net effect on election results is ambiguous. To estimate the effect of a candidate's total rallies, I compare electoral outcomes under (i) "No one Rally" with (ii) "Only one candidate rallies." I find that Trump's rallies increased his chances of winning by 33%. By contrast, other candidates did not increase their chances of winning significantly. In the second counterfactual experiment, I focus on the number of rallies a candidate held or did not hold as a campaign response to the opponent. For this case, I compare the number of rallies when (i) "Both candidates rally" with (ii) "Only one candidate rallies." In case (ii), candidates are in a decision-theoretic environment and optimize while considering cross-sectional and dynamic uncertainty alone. In case (i), an additional layer of complexity is added as candidates are in a game theoretic environment. The equilibrium play by a candi-

[6]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.

date inherently incorporates opponent actions either by anticipation (when first mover) or by observation (when second mover). The counterfactual experiment uncovers that Obama and Clinton held 12 and 13 additional rallies as a campaign response, while Romney did not change his number of rallies, and Trump held 3.5 fewer rallies.

The fourth contribution is to analyze campaign silence laws and spending limits using counterfactual experiments. Campaign silence laws ban campaigning for a given number of days in the run-up to an election, and they are enforced in many countries for different durations. For instance, France imposes a campaign silence that lasts one day (Pickles, 1960), while Cyprus, Indonesia, and Brazil impose campaign silences that last two or more days (Knews, 2022; IFES, 2012; Globo, 2020). It remains unclear what is the minimal length of campaign silence that effectively reduces the influence of campaigning on election results. I impose campaign silences of distinct durations and find that short campaign silences (<= 4 days) do not change the outcome of an election because they cannot induce sufficient decay to reduce the dependence of electoral results on campaigning.

I also analyze spending limits by enforcing limits on total number of rallies that can be held. For this experiment, I assume that monetary cost of a rally is same across candidates and years.[7] I find that even seemingly nonbinding limits[8] on the number of rallies can induce a reduction in overall campaigning by candidates and change electoral outcomes significantly. For instance, none of the candidates held more than 110 rallies during the fall campaign of 2012 and 2016 elections, however a limit of 110 rallies significantly reduce the total number of rallies held. This is because candidates optimally choose to allow for contingencies that may require intense rallying in the future. This behavior is absent is static environments.

This paper makes contributions to two bodies of literature. 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 when and where to hold a 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

[7]Cost parameters in the model reflect opportunity cost of holding a rally.

[8]I am using the term “seemingly nonbinding limits” to refer to those limits that are higher than observed number of rallies.

a dynamic model with candidate-specific strategies where campaign effects decay. Iaryczower et al. (2022) model dynamics of an incumbent senator's advertising and platform choices contingent on their lead in the polls. The authors do not model the opponent's problem and focus on incumbent senators. In this paper, I model the dynamic electoral competition between two office seeking candidates.

Acharya et al. (2022) study political spending within an election and identify candidates' perceived decay rate associated with campaigning. However, the authors characterize optimal spending ratios rather than candidate-specific spending levels, while I use a perfect information setup that allows one to study candidate-specific strategies. Kawai and Sunada (2022) study spending across elections, whereas I study rally decisions within an election.

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 the effects of rallies on poll margins and on the electoral outcomes. The literature finds mixed evidence on the effects 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 rally effectiveness across candidates and attempted to 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. Whereas the authors in this study used a difference-in-differences specification at the media market level to address the selection bias,[9] I directly address the selection bias by modeling these rally decisions. Grosjean et al. (2022) found that Trump rallies led to more racial discrimination by police officers. The effect is not explained by changed behavior of drivers, as it is stronger when officers were racially biased, and when Trump directly or indirectly mentioned racial issues.

The paper proceeds as follows, Section 2 discusses the model, equilibrium, and comparative statics. Section 3 discusses data sources, summary statistics, and empirical patterns. Section 4 discusses parameterization and the 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 counterfactual experiments. Section 8 con-

[9]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.

cludes.

2 Model

The model analyzes the interaction between presidential candidate rally decisions and their popularity levels. 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 the popularity measure 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 stage game, the order of play between the candidates is random. Each candidate, at their turn, must choose at most one state out of K states. In the election period $T+1$, every state chooses the popular candidate as the winner.

2.1 Preliminary

The set of states is given by $K = \{1, 2, \ldots, K\}$. Let $i \in \{R, D\}$ and $t \in \{1, 2, \ldots, T\}$ be an arbitrary candidate and period, respectively. A candidate can hold at most one rally in a state and pay a cost c_i if a rally is held. A rally decision is denoted by a_{ikt}. State-level popularity is denoted by p_{kt}, and the next-period popularity, $p_{k,t+1}$, follows an AR(1):

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

where α_i is i's effectiveness in influencing popularity, the parameter ρ is the persistence in popularity, $\nu_{k,t}$ is a popularity shock, and δ_k is a state-specific drift. I assume the following.

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

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

Where $\sigma_\nu^2 I_K$ is a positive definite matrix and σ_ν is the volatility in popularity.[10] Let the density of popularity in period $t+1$ given period t primitives be denoted by $f(p_{t+1}|a_{R,t}, a_{D,t}, p_t)$.

[10] In 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 states are allowed.

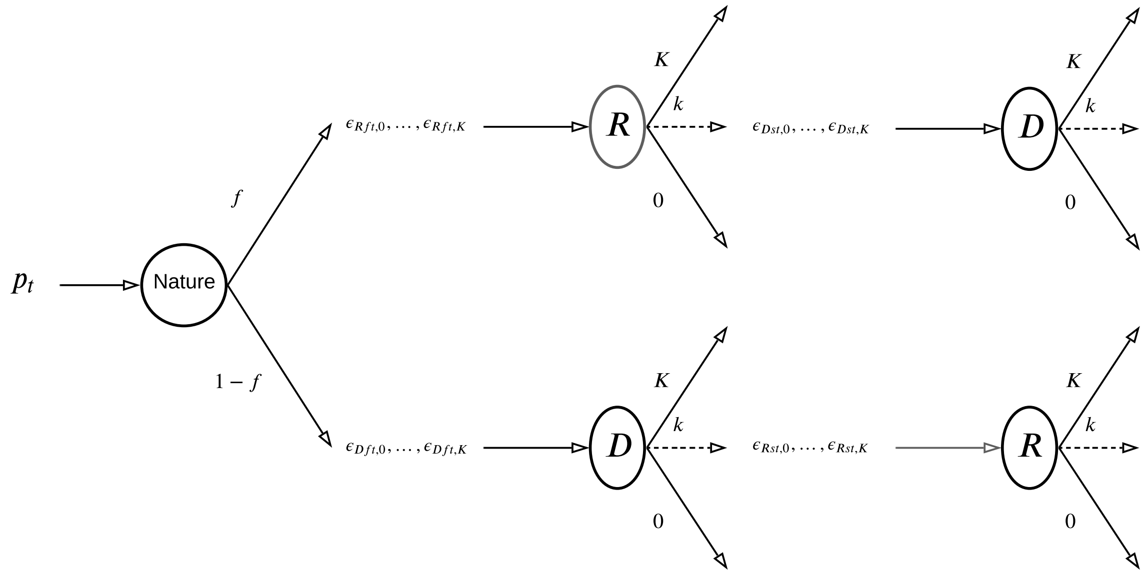


Figure 1: Stage Game

The stage game for each period $t = 1, 2, \dots, T$ is depicted in this figure. In each period, both candidates observe their 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 from the distribution and observed by both candidates. The second mover makes their rally decision and then the game proceeds to period $t + 1$.

Note that this density takes the form of the normal pdf under Assumption 2.1. The popularity evolution equation 2.1 and Assumption 2.1 can also be statistically founded by considering a mean-reverting process similar to that of Acharya et al. (2022).[11]

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

$$V_{R,T+1}(p_{T+1}) = \sum_{k=1}^{K} e_k E \times \mathbb{1}\Big\{p_{k,T+1} > 0\Big\}. \tag{2.3}$$

[11]The authors also micro-founded the mean-reverting process by considering a set of impressionable voters (Andonie and Diermeier, 2019) who vote on the basis of good will.

2.2 Timing of Decisions and Information

In $t \in \{1, 2, \dots, T\}$, a stage game, as described in Figure 1, is played. The stage game unfolds over six subperiods, $\tau_1, \tau_2, \dots, \tau_6$, which are described below:

- τ_1 The popularity-level vector $p_t = \left(p_{1t}, p_{2t}, \dots, p_{Kt}\right)$ is observed by the candidates.
- τ_2 Nature draws the R's and D's order of play for the stage game. The probability i is chosen as the first mover is denoted by f_i.
- τ_3 The first mover i's cost shocks, $\epsilon_{if,t} = \left(\epsilon_{if,t,0}, \epsilon_{if,t,1}, \dots, \epsilon_{if,t,K}\right)$, are realized. These cost shocks capture the effect of unforeseen events on rally decisions.[12].
- τ_4 The first mover, i, makes a rally decision by solving the following bellman equation:

$$V_{ift}(p_t, \epsilon_{if,t}) = \max_{k \in \{0,1,\dots,K\}} \Bigg\{ -c_i \times \mathbb{1}\{k \neq 0\} - \epsilon_{if,t,k} + \beta \sum_{l=0}^{K} \mathbb{E}_p \left[V_{i,t+1}(p) \middle| a_{it} = k, a_{jt} = l, p_t \right] \times \sigma_{js,t}\left(l; k, p_t\right) \Bigg\}, \tag{2.4}$$

  where the flow utility consists of the cost c_i and the random cost shock $\epsilon_{if,t,k}$. 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 actions, $a_{it} = k$, $a_{jt} = l$, and current-period popularity p_t. The outer expectation is with respect to second mover j's action given i's action k and current-period 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 j's cost shocks, $\epsilon_{js,t} = \left(\epsilon_{js,t,0}, \epsilon_{js,t,1}, \dots, \epsilon_{js,t,K}\right)$, are realized.
- τ_6 The second mover j makes a rally decision. The second mover, j, also observes the first mover's action. Therefore, $a_{if,t}$ is also a state variable for the second mover in addition to popularity and cost shocks. After observing these state variables, second mover j solves the following bellman equation:

$$V_{jst}(l, p_t, \epsilon_{js,t}) = \max_{k \in \{0,1,\dots,K\}} \left\{ -c_j \times \mathbb{1}\{k \neq 0\} - \epsilon_{js,t,k} + \beta \mathbb{E}_p \left[V_{j,t+1}(p) \middle| a_{jt} = k, a_{it} = l, p_t \right] \right\}, \tag{2.5}$$

[12]The interpretation of cost shocks using unforeseen events in this context is not just a convenience construct. For instance, Hurricane Sandy made campaigning on the Atlantic seaboard very difficult in the 2012 presidential election.

where c_i and $\epsilon_{is,t,k}$ are the flow costs from choosing option k. The continuation value is the expectation of i's value in the next period given a_{it}, a_{jt}, p_t. Let a_{ist} be the associated policy function.

Candidate i's value function at popularity vector p_t prior to the order of play in period t is given by:

$$V_{i,t}(p_t) = f_i \times \mathbb{E}_{\epsilon_{if,t}}\Big(V_{ift}(p_t, \epsilon_{ift})\Big) + (1 - f_i)\sum_{k=0}^{K}\Big[\sigma_{jft}(k; p_t) \times \mathbb{E}_{\epsilon_{is,t}}\Big(V_{ist}(k, p_t, \epsilon_{ist})\Big)\Big]. \quad (2.6)$$

With probability f_i, i is the first mover and the term $\mathbb{E}_{\epsilon_{if,t}}\Big(V_{ift}(p_t, \epsilon_{ift})\Big)$ is the expected payoff to the first mover. The second term is the expected payoff of i when they are the second mover. This term is a nested expectation: the outer expectation is taken with respect to j's rally decisions when j is the first mover and the inner expectation is taken with respect to cost shocks.

2.3 Equilibrium

Recall that this is a game of perfect information as only one candidate moves at a given time and all past actions and shocks are common knowledge. Therefore, the game is solved using backward induction.

Assumption 2.2 (Independent Cost Shocks) *Cost shocks are independent across all information nodes and actions.*

This assumption implies that the current cost shocks are payoff relevant, but past cost shocks are not. I also assume that cost shocks are drawn from the type-1 extreme value distribution.

Assumption 2.3 (Distribution of Cost Shocks) *Cost shocks are drawn from the type-1 extreme value distribution.*

Assumptions 2.3 and 2.2 ensure that the subgame perfect equilibrium exists and is essentially unique, i.e., multiplicity can exist with probability zero.[13] Moreover, we can also show that best responses, in probability space, are functions of current popularity, cost shocks,

[13]The only way multiple equilibria will exist is when a candidate is indifferent between two actions. However, under these assumptions the random variable formed by adding ϵ_{imtk} to $-\epsilon_{imtl}$ for each i, m, t ,k, l is a continuous random variable. Therefore, the indifference conditions hold with probability zero.

and, in the case of the second mover, first mover action. Proposition 2.1 lays out this characterization.

Proposition 2.1 *Given Equation 2.3, which defines the electoral payoff, and Assumptions 2.2 and 2.3, the following equations hold for all* $t = 1, 2, ..., T$

$$\begin{aligned} V_{i,t}(p_t) =& f_i \times \ln\left(\sum_{k=0}^{K} \exp\left\{u_{if,t}(k;p_t)\right\}\right) \\ &+ (1-f_i) \times \sum_{k=0}^{K}\left[\sigma_{jf,t}(k;p_t)\ln\left(\sum_{l=0}^{K}\exp\left\{u_{is,t}(l;k,p_t))\right\}\right)\right] + \gamma \end{aligned} \tag{2.7}$$

$$\sigma_{if,t}(k;p_t) = P\left(k = a^*_{if,t}(p_t, \epsilon_{if,t})\right) = \frac{\exp\left(u_{if,t}(k;p_t) - u_{if,t}(0;p_t)\right)}{1 + \sum_{l=1}^{K}\exp\left(u_{if,t}(l;p_t) - u_{if,t}(0;p_t)\right)} \tag{2.8}$$

$$\sigma_{is,t}(k;l,p_t) = P\left(k = a^*_{is,t}(a_{jft} = l, p_t, \epsilon_{is,t})\right) = \frac{\exp\left(u_{is,t}(k;l,p_t) - u_{is,t}(0;l,p_t)\right)}{1 + \sum_{q=1}^{K}\exp\left(u_{is,t}(q;l,p_t) - u_{is,t}(0;l,p_t)\right)}, \tag{2.9}$$

where $V_{i,t}$ *is equilibrium period* t *value of candidate* i*. Moreover,* $\sigma_{if,t}$ *and* $\sigma_{is,t}$ *are the equilibrium choices of the first and second mover, respectively. Lastly,* $u_{if,t}$ *and* $u_{is,t}$ *are the option-specific value functions of the first and the second mover, respectively.*[14]

The proof of Proposition 2.1 is given in Section A.1. The proof uses induction. Starting with the second mover I show that if period $t+1$ value function, $V_{i,t+1}$, is bounded then the optimal actions given the cost shocks are unique a.s. and characterize the conditional choice probabilities and the value function. Then I repeat these steps for the first mover. For the inductive argument, I show that if period $t+1$ value function is bounded then period t value function is also bounded. Since period $T+1$ value function is bounded by definition, proposition holds for all periods.

I rely on simulations to analyze equilibrium behavior since equilibrium choices do not have reduced form expressions and the presence of non-linear functions[15] makes applying montone comparative static results infeasible.[16] Moreover, I analyze a candidate i's proba-

[14] Or deterministic part of period t payoffs.

[15] Such as the standard normal cumulative distribution function for the expected payoff in period $T+1$, logexp sum to evaluate $t+1$ value function, and multinomial logit functional forms for conditional choice probabilities.

[16] I follow Judd et al. (2014), Judd et al. (2017), and Heiss and Winschel (2008) for simulating the model. Online Appendix D provides the details.

bility of choosing k before nature chooses the order of play in a given period, call this quantity $\sigma_{it}(k; p_t)$.[17]

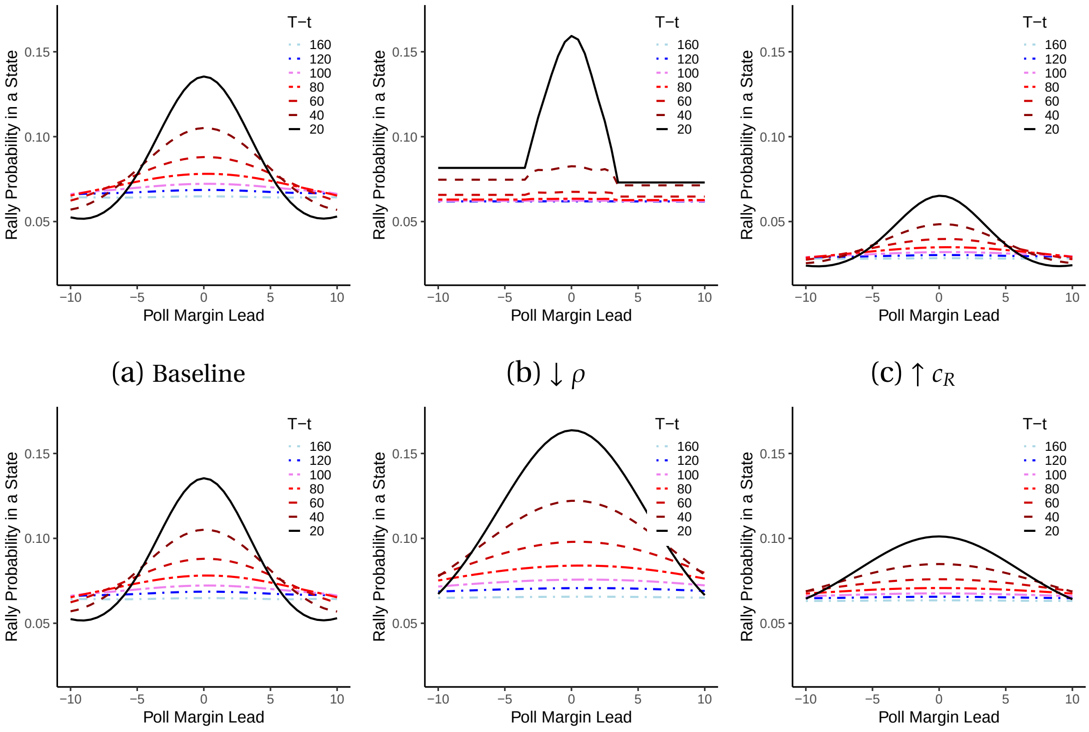
Three remarks are necessary regarding how the $\sigma_{it}(k, p_t)$ relates with p_{kt}, which is demonstrated in Figure 2a. First note, the conditional choice probabilities have a weak relationship with popularity in the respective states in earlier periods. The explanation behind this is twofold. For earlier periods (i.e., small t), electoral payoff incentives are repeatedly discounted and therefore have negligible influence. Therefore, conditional choice probabilities do not depend on popularity. Also, earlier rallies have negligible effects on election day popularity because these effects decay exponentially with time. As a consequence, earlier rallies are less beneficial for the candidates and therefore in equilibrium probability of holding a rally is low.

A second remark highlights what happens when we are closer to election day. As the election approaches, 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. The winner-takes-all payoff at the state level ensures that the change in the expected payoff due to a rally is maximal when a candidate's popularity is closer to zero and negligible when that candidate's popularity is lopsided. The third remark focuses on the transition shown in Figure 2a. As the elections approaches, the decay and discounting channel weakens, making rallying profitable but only in states where the candidate has close to zero popularity.

2.4 Comparative Statics

In this section, I discuss four comparative statics concerning rally effectiveness, cost of rallying, persistence, and volatility in popularity. A primary consequence of increased persistence in popularity (ρ) is heightened observed autocorrelation. Furthermore, greater persistence leads to earlier rallies exerting a stronger influence on election day popularity (p_{T+1}) across all current period popularity levels (p_t). This stronger effect prompts the emergence of a bell-shaped relationship between rallies and popularity earlier rather than later. To illustrate this, consider the comparison between Figures 2a and 2b. Figure 2a, which has a higher ρ value, displays higher, more distinctly bell-shaped rally probabilities in earlier pe-

[17]To calculate this, first I calculate the joint probability of candidates i and j choosing k and l respectively, call this $\sigma_t(k, l; p_t) = \frac{\sigma_{ift}(k,p_t)\sigma_{jst}(l;k,p_t)+\sigma_{jft}(l,p_t)\sigma_{ist}(k;l,p_t)}{2}$. Then $\sigma_{it}(k; p_t) = \sum_l \sigma_t(k, l; p_t)$.



(a) Baseline (b) $\downarrow \rho$ (c) $\uparrow c_R$

(d) Baseline (e) $\uparrow \alpha_R$ (f) $\sigma_v \uparrow$

Figure 2: *This figure illustrates how the probability of rallying in a state changes when the key parameters ρ, α_i, c_i, and σ_v change. Here $T - t$ denotes the periods before the election. For the baseline I fix parameter values $\alpha_R = -\alpha_D = 0.07$, $\rho = 0.99$, $c_R = c_D = 2.5$, and $\sigma_v = 0.2$. There are four states and they are symmetric except for the popularity values. A candidate's popularity in a state is 0, -20 and 20 for states 2, 3 and 4, respectively. The x-axis shows the variation in state 1's popularity value. Panels 2b, 2c, 2e, and 2f illustrates the case when $\rho = 0.95$, $c_R = 3.5$, $\alpha_R = 0.15$, and $\sigma_v = 0.5$, respectively while the other parameters are kept the same as in the baseline.*

riods compared to Figure 2b, where ρ is lower and the probabilities are flatter and lower.

Increasing the cost of rallying parameter (c_i) diminishes the likelihood of rallying as it becomes a more expensive option. This effect is consistent across all periods and levels of popularity, as demonstrated by the fact that c_i is higher in Figure 2c than in Figure 2a. Conversely, an increase in the rally effectiveness parameter (α_i) results in a higher frequency of rallies since they become more advantageous for the candidates, causing the bell-shaped relationship to manifest sooner, as shown in Figure 2e.

Increasing the volatility in popularity parameter (σ_v) introduces greater uncertainty regarding future popularity. This uncertainty has two effects. First, it reduces rally probabilities when popularity is near the threshold due to the risk of unfavorable outcomes. Second, at extreme popularity values, the likelihood of holding a rally increases, particularly in earlier periods. In situations where popularity is significantly negative, increased volatility increases the prospects of a comeback for trailing candidates, prompting more frequent rallies. Conversely, when candidates have a strong lead, increased volatility increases the risk of losing their advantage, motivating them to consolidate their position through more frequent rallies. These dynamics are evident when one compares the rally probabilities in Figures 2a and 2f.

3 Data

3.1 Data Sources and Summary Statistics

This paper uses two primary data sources. The first, Democracy in Action, provides the rally decisions. The second, FiveThirtyEight, provides state-specific poll margins (popularity in the model).

Rally decisions: The data on rally decisions is obtained from Democracy in Action, a website created by Eric M. Appleman. In particular, for the 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 reports the activities candidates undertook in chronological order. The website provides information not only on rallies but also on various other activities. To isolate political rallies from other activities, I search whether the words “rally”, “speech”, or “special events” are mentioned in the description. In the model, each candidate holds one rally in a period, while in the data, each candidate can hold multiple rallies in a day. I define periods in the model as each lasting a quarter of a day and allocate these periods based on chronological information provided in Democracy In Action.[18]

Poll Margins: I use FiveThirtyEight’s poll repository to obtain aggregate poll margins at the state level. FiveThirtyEight is an organization that focuses on opinion poll analysis, economics, politics, and sports blogging. Since its inception, FiveThirtyEight has focused

[18] I 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 B, I provide details on data cleaning and allocation of periods to rally decisions.

Table 1: Summary Statistics

	2012 Election				2016 Election				
	Poll Lead		Rallies		Poll Lead		Rallies		
State	Mean	Std. Dev.	Obama	Romney	Mean	Std. Dev.	Clinton	Trump	EC Votes
Arizona	8.59	0.785	0	0	1.58	1.1	0	3	11
Colorado	-0.394	1.200	11	8	-4.91	1.5	0	8	9
Florida	-0.0386	1.380	6	20	-2.47	0.977	15	23	29
Iowa	-1.32	1.230	15	10	1.160	1.41	4	5	6
Michigan	-4.78	1.820	0	0	-7.170	1.3	4	5	16
Nevada	-2.98	0.755	6	4	-0.731	1.1	4	4	6
New Hampshire	-2.59	1.4	6	3	-5.91	1.72	2	8	4
North Carolina	2.4	1.06	0	3	-1.66	0.828	9	15	15
Ohio	-2.58	1.2	20	26	-0.849	1.69	10	13	18
Pennsylvania	-5.26	1.21	0	2	-6.26	1.09	9	16	20
Virginia	-0.996	1.13	10	20	-6.87	1.91	0	6	13
Wisconsin	-3.88	1.67	5	0	-7.1	1.62	0	5	10

[a] Note: The table shows summary statistics for the Republican candidate's poll margin lead, the number of rallies held by all candidates across the two elections, and the number of electoral college votes by state.

on producing reliable forecasts for presidential general elections, primaries, congressional elections, and gubernatorial elections. In 2016, the organization produced one of the most accurate forecasts for the presidential election. As a poll aggregator, FiveThirtyEight 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. [19]

The aggregate number of activities in the raw data obtained after classifying the set of all activities is provided in Table 8.[20] The table also shows how many rallies were removed specifically after the cleaning process for each candidate. Detailed summary statistics for the states that had two or more rallies by a candidate are provided in Table 1. These states provide a sufficient 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

[19] For more details on the methodology visit https://fivethirtyeight.com/features/a-users-guide-to-fivethirtyeights-2016-general-election-forecast/

[20] These numbers correspond to last 125 days. For the analysis I focus on last 100 days.

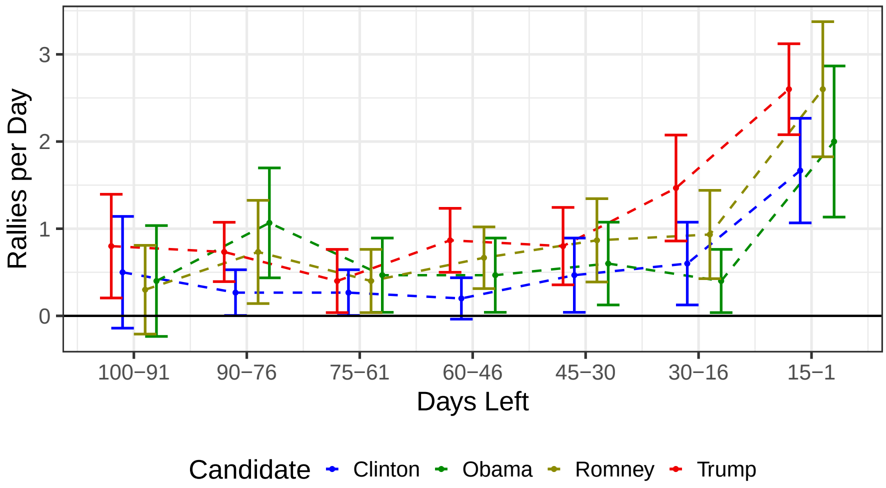


Figure 3: This figure shows average rallies per day for 15-day bins (and a 10-day bin for 100–91 days before the election). For each of these bins the corresponding 95% confidence interval for average rallies per day is also provided.

variation in poll margins is smaller. Most standard errors are between 0.6 to 1.9, suggesting that day-to-day variation in poll margin lead is moderate.

As can be seen in Table 1, Florida witnessed a significant number of rallies consistently in both elections (26 in 2012 and 38 in 2016). Arizona had no rallies in 2012, and only 3 in 2016. Ohio also witnessed high number of rallies in both years, with 46 rallies in 2012 and 23 in 2016. By contrast, Pennsylvania had an average of 2 rallies in 2012 but 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.

3.2 Dynamic Patterns of Political Rallies

Rally Ramp-up: For each candidate, rally intensity increases as election day approaches. This pattern is evident in Figure 3. To produce these plots, I consider daily rallies across all states. Then, I create 15-day bins for 90-1 days before an election and a 10-day bin for 100-91 day before an election and calculate average rallies per day, standard deviation, and the corresponding 95% confidence interval. The pattern in Figure 3 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 AR(1) process for modeling popularity. In my model, the

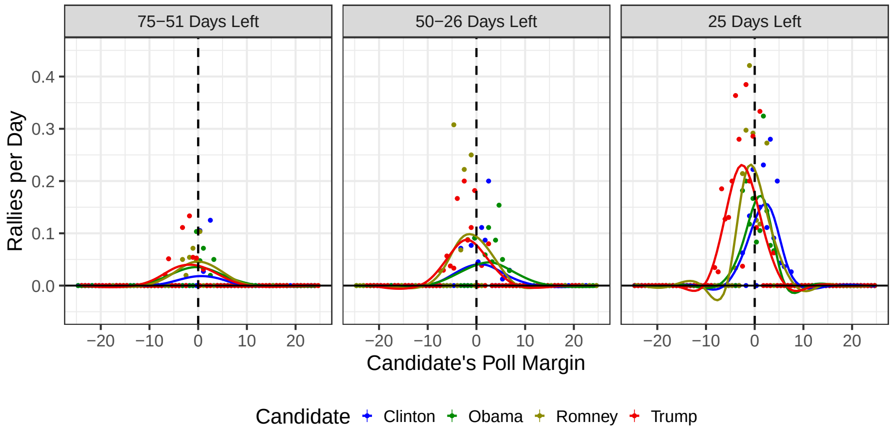


Figure 4: This figure shows a bin scatter of a candidate's number of rallies in a day and poll margin lead along with a generalized additive model fit line.

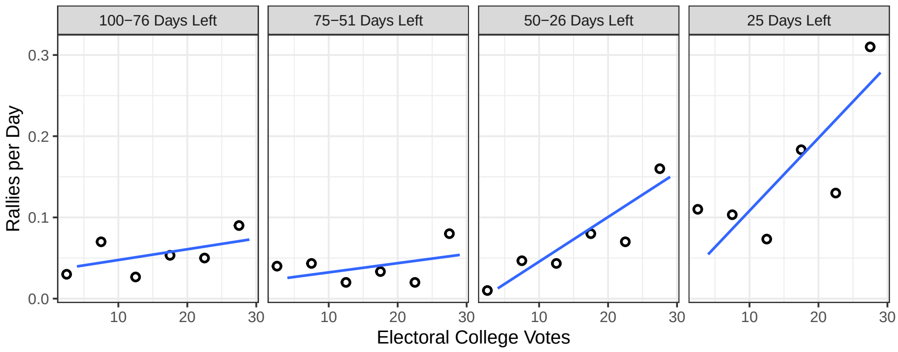


Figure 5: This figure shows a bin scatter of a candidate's number of rallies and electoral college votes at the state × day level for swing states.

persistence in popularity parameter has a one-to-one relation with the decay rate. The deviation from Acharya et al. (2022) is the candidate-level heterogeneity in these patterns that my model can support due to candidate-specific cost and rally effectiveness.[21]

Rallies and Poll Margin: Candidates rally in highly contested states as elections approach. This pattern relates to the qualitative prediction discussed in Section 2 (Figure 2a). 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 4. I also plot generalized additive model fits for each candidate and day bin. In Figure 4, it is evident that as the election approaches, candidates rally more intensely in states where candidate polls are neck and neck. A cross-sectional pattern for television advertising and vote share lead is documented in Gordon

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

and Hartmann (2016), and a cross-sectional pattern for campaign activity is documented in Strömberg (2008).

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.[22] Within this set of states, candidates prioritize states with higher electoral college votes over those with lower electoral college votes. From the Table 9 and Figure 5 show that for all candidates, the correlation between rallies and electoral college votes increases as the election day approaches.[23]

4 Identification and Estimation

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 $c_R, c_D, c_1, \ldots, c_K$. 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 and c_D are identified by the initial level of rallying in state K. Parameter c_k is identified by the average of the two candidates' probability of rallying in state k.

The second set of parameters that we are interested in is the set of parameters that govern the AR(1) process defined in equation 2.1. It include $\alpha_R, \alpha_D, \rho, \sigma_\nu, \delta_1, \ldots, \delta_K$. The identification of α_i relies on candidate i's strategy and the change in popularity, P_{ikt}, post a rally in state k at time $t-1$. The parameter ρ is identified jointly by the autocorrelation of poll margins and the gradual increase in the level of rallying as the election approaches. The dispersion in poll margins help in identifying σ_ν. State-specific drifts, δ_k are identified by long-run means of popularity.

The data used in this paper does not allow one to identify the parameters f (i.e., the probability of R moving first), and E (i.e., the maximal electoral payoff). To identify f, one would

[22] 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).

[23] The candidate level analysis reveals this pattern for Trump, Clinton, and Romney.In the case of Obama, the correlation starts negative and significant value and gradually becomes positive, but insignificant.

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 between rallying in a state and not rallying. However, multiple parameters influence this difference. Specifically, the parameters of rally effectiveness, cost of rallying, and persistence in popularity influence the difference in payoff. Consequently parameter E cannot be identified. I calibrate its value to 538, the total number of electoral college votes in the United States.

4.2 Likelihood

I use full solution method to estimate the model. Note that the sequential move assumption of this game brings it closer to Igami (2017). Therefore, I choose the full solution method to estimate this game (Rust, 1987).[24] This subsection presents the log likelihood that I use to estimate the model. Let A_t denote the realized rally decisions in a period t, for $t = 1, 2, \ldots, T$. Let P_t denote the realized poll margins in a given period t, for $t = 1, 2, \ldots, T + 1$. Given the assumptions 2.2, 2.3 and 2.1 we can characterize the transition density for random vectors $\tilde{X}_t = (A_t, P_{t+1})$ for $t = 1, 2, \ldots, T$. This states that $\tilde{X}_t$ is the vector containing rally decisions in period t and the popularity vector in period $t + 1$. 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. Lemma 4.1 defines this transition density for us.

Lemma 4.1 *Given Assumptions 2.2, 2.3, and 2.1 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:*

$$\psi_t(\tilde{X}_t|\tilde{X}_{t-1}) = f(P_t|A_{R,t}, A_{D,t}, P_t)\sigma_t\,(A_t; P_t) \tag{4.1}$$

where σ_t is joint probability of rallying and f is density of popularity shocks.

The proof of Lemma 4.1 follows straight from the equilibrium choice probabilities and the AR(1) process for modeling local popularity. The transition density, $\psi_t(\tilde{X}_t|\tilde{X}_{t+1})$ would be the ideal choice for estimation if all popularity values were observed. I have one observation of poll margins per day. Since I define periods in the model as each lasting a quarter of a

[24]There is an active literature that studies the estimation of dynamic discrete games without making assumptions on move order, such as Aguirregabiria and Mira (2007), Aguirregabiria and Marcoux (2021), Bajari et al. (2007), Egesdal et al. (2015).

day, we are limited to one observation of poll margins for every four periods. I assume that poll margins I observe at d are realized at the beginning of day $d+1$. In other words, the poll margin I observe on day d is isomorphic to the popularity candidates would observe in period $t = 4d + 1$, which one may also call the first subperiod of day $d+1$. The remaining popularity values for periods $4d+2$, $4d+3$, and $4d+4$ are missing. Therefore, each period t can be mapped to (d, l), where d is a day, and l is a subperiod of the day d. There will be four subperiods in each day. Therefore, for any period t, there exists a day d and subperiod, l such that $t = 4(d-1) + l$.

Let X_d be the observations for day d. I observe all chosen rally decisions taken by candidates on day d. These decisions 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 subperiod l (or period $4(d-1)+l$). I also observe popularity, or poll margin, for day d, which I assume to be realized in subperiod 1. Therefore, for a given day d, P_{4d-3} is observed, but P_{4d-2}, P_{4d-1} and P_{4d} are not. It is worth noting that $P_{4(d-1)+l} = (P_{4(d-1)+l,1}, P_{4(d-1)+l,2} \ldots, P_{4(d-1)+l,K})$ 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 $\{X_1, X_2, \ldots, X_{\bar{D}}\}$ obey the Markov property and that the day-to-day transition density of these observations, denoted by $\lambda_d^{\theta}(X_d|X_{d-1})$.

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

$$\lambda_d^{\theta}(X_d|X_{d-1}) = \int\limits_{(p_2,p_3,p_4)\in\mathbb{R}^{3K}} \left(\prod_{l=1}^{4} \sigma_{4(d-1)+l}(A_{4(d-1)+l}; p_l)\right) \times \left(\prod_{l=1}^{4} f\left(p_{l+1}|A_{4(d-1)+l}, p_l\right)\right) dp_2 dp_3 dp_4 \quad (4.2)$$

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

Proposition 4.1 is proved in Appendix A.1.2. The proof follows from applying Lemma 4.1 recursively. Based on this proposition I can formulate the likelihood of observing X_1, X_2, …, $X_{\bar{D}}$. It is worth pointing out that this Markov process is not time homogeneous, as the densities vary with day d. The key reason for having this density vary with day d is the candidates' equilibrium choice profiles. As seen in the data and also the model predictions, rally intensity increases as the election approaches and therefore a Markov process that is time homogeneous cannot support these features. Finally, the log likelihood is $\ell\ell\left(\theta; X_1, X_2, \ldots, X_{\bar{D}}\right) = \sum_{d=1}^{\bar{D}} \log\left(\lambda_d^{\theta}\left(X_d \middle| X_{d-1}\right)\right)$. The integration in Proposition 4.1 is not

Table 2: Summary Statistics for States Groups

State	Rallies Per Day				R's Poll Margin		Electoral College Votes
	Romney'12	Obama'12	Trump'16	Clinton'16	2012	2016	
Southeast	0.12	0.17	0.15	0.04	3.93	2.27	26
	(0.68)	(0.81)	(0.76)	(0.40)	(0.56)	(0.96)	
Midwest	0.10	0.20	0.15	0.08	-2.73	-2.08	32
	(0.62)	(0.87)	(0.76)	(0.56)	(1.49)	(1.36)	
Northeast	0.31	0.26	0.37	0.22	-2.72	-0.39	42
	(1.07)	(0.99)	(1.16)	(0.91)	(1.07)	(1.28)	
Southwest	0.43	0.16	0.44	0.24	1.52	0.21	57
	(1.24)	(0.78)	(1.25)	(0.95)	(1.11)	(0.96)	

**Standard deviations in parentheses wherever applicable.*

[a] Note: The table shows summary statistics for the number of daily rallies and the average Republican poll margin lead across state groups. These statistics are given for the last 100 days before the election. In 2012 Midwest and Northeast groups were Democrat leaning while Southeast group was Republican leaning. Relatively, the Southeast group was neither Republican leaning nor Democrat leaning. In 2016, similar patterns hold for Southeast and Midwest groups while Northeast and Southwest groups were relatively neutral.

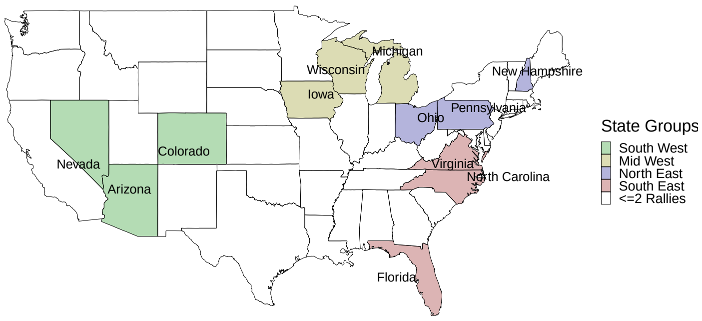


Figure 6: *The figure displays states considered in the analysis. The unlabeled states, which are omitted, had less than 2 rallies by each candidates. The states considered in the analysis are all swing states. To keep the dimensionality of the model tractable I group these swing states into 4 groups.*

feasible analytically and therefore I rely on Monte Carlo methods which are discussed in Appendix C.

4.3 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 sufficient 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 U.S. region and the swing states[25] in that region. For instance, this intersection is given by swing states Nevada, Arizona, and Colorado for the Southwest group. The Southeast group consists of Florida, Virginia, and North Carolina. The Midwest group consists of Michigan, Wisconsin, and Iowa. Finally, the Northeast group consists of New Hampshire, Pennsylvania, and Ohio. Ohio is a Midwest state but it is included in the Northeast group in order to obtain a state group with a similar number of electoral college votes as the Southwest group.[26]

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 weight of each state is proportional to its share of electoral college votes within the state group. The summary statistics for rallies and final poll margins are provided in Table 2. 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

Table 3 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. The estimates uncover that a Trump rally increased his poll margin lead over Clinton by 0.0839 pp, while a Clinton Rally increased her lead over Trump by 0.0745 pp. For 2012, I find a Romney rally increased his lead over Obama by 0.100 pp lead over Obama while an Obama rally increased his lead over Romney by 0.042 pp. These estimates are significant. The net effect of both candidates rallying in a state for 2016 is a gain of 0.0094 pp (0.012) in favor of

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

[26] I consider two alternative definitions of state groups for robustness tests. One considers the state groups used in Snyder and Yousaf (2020), and the other treats Florida as the sole member of the Southeast state group by including Virginia and North Carolina in the Northeast group as Florida is geographically isolated. The results from this exercise are discussed in Section 6.

Trump, which is insignificant, while for 2012 the net effect is a gain of 0.058 pp (0.014) in favor of Romney which is significant.

Table 3: Parameter Estimates

Main Parameters			Fixed Effects		
Parameters	2012	2016	Parameters	2012	2016
Popularity					
α_R	0.100	0.0839	δ_1	0.04	0.023
	0.0227	0.0155		0.011	0.008
α_D	-0.042	-0.0745	δ_2	-0.027	-0.02
	0.0135	0.0152		0.0082	0.0090
ρ	0.990	0.991	δ_3	-0.027	-0.0069
	0.002	0.001		0.0082	0.0073
σ	0.147	0.16	δ_4	0.00019	-0.0074
	0.0143	0.0148		0.0079	0.0073
c_R	2.9	2.36	c_1	0.367	0.943
	0.287	0.208		0.324	0.411
c_D	2.83	3.26	c_2	0.421	0.788
	0.206	0.259		0.277	0.308
			c_3	-0.0864	-0.0443
				0.281	0.22

	2012	2016
Observations	100	100
Log Likelihood	-656.26	-654.60

[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.

As mentioned above, 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). This paper's findings qualitatively agree with articles that have considered candidate-specific effects (Shaw and Gimpel, 2012). This paper also shows that rally effectiveness can differ significantly in magnitude, unlike the assumption of identical effectiveness made in Strömberg (2008).

The persistence in popularity parameter is approximately 0.99 across both election years. The weekly decay rate is 28%.[27] This decay rate lies in the right tail of the perceived de-

[27] To calculate the decay rate, λ, over Δ periods I use the relation $\rho = e^{-\frac{\lambda}{\Delta}}$. For weekly decay rate, $\Delta = 7 \times 4$

cay rate distribution estimated in Acharya et al. (2022). However, it is lower than Hill et al. (2013). The third-degree lagged polynomial specification considered in Gerber et al. (2011) estimated a decay rate of 25%, similar to this paper's estimate.

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. This estimate captures the asymmetry in rally decisions between the opponents, despite possessing similar levels of rally effectiveness. Meanwhile, this asymmetry is not found for the 2012 election, as the cost estimates of rallies are not significantly different for Obama and Romney.

To put rally effectiveness estimates into perspective, refer to Table 4. The table displays the persuasion rates of rallies and TV advertising. The persuasion rate measures the proportion of voters who switched their voting choice from candidate i to the opponent after being exposed to a tool of persuasion of candidate i. The persuasion rates of TV advertising, estimated in Spenkuch and Toniatti (2018), are also provided. I adopt the definition of persuasion rates of 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 is voter turnout in state k:

$$f_i^{\text{Rally}} = \mathbb{E}\left[\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{V_k}{T\cdot(\sum_l V_l)}\cdot\frac{2}{100+(-1)^{\mathbb{1}\{i=R\}}P_{kt}}\cdot|\alpha_i|\right] \tag{5.1}$$

The term $(100+(-1)^{\mathbb{1}\{i=R\}}P_{kt})/2$ calculates proportion of voters in support of the opponent of candidate i when no rallies are held in state k in period t.[28] When this term is multiplied by $|\alpha_i|$, it gives the estimate of the proportion of voters who switch from the opponent to candidate i.[29]

I find that the persuasion rates of rallies are higher than those of TV advertising. For instance, 17 TV ad spots are as persuasive as one Trump rally. These ratios for Romney, Obama,

and therefore $\lambda = -28\times\log(\rho)$, where log has the base e.

[28]Note that for k,t we have $P_{kt} = P_{Rkt} - P_{Dkt}$ and $P_{Rkt} + P_{Dkt} = 100$. Then $P_{Rkt} = (100+P_{kt})/2$ and $P_{Dkt} = (100-P_{kt})/2$. Since the denominator should be the number of voters who do not support a candidate, P_{Dkt} is used as denominator for $f\text{Rally}_R$ and vice-versa.

[29]I take average across all states and periods and calculate its expectation using simulations.

Table 4: Persuasion Rates

	2012		2016		Spenkuch and Toniatti (2018)	
	Romney	Obama	Trump	Clinton	Rep. Ads	Dem. Ads
Pers. Rate of **1** Rally/T.V. ad (%)	0.20	0.085	0.168	0.150	0.01	0.03
	0.001	0.0003	0.032	0.029	0.005	0.004
Agg. Switched Decisions (in Millions)	0.35		0.32		2.2	
	0.10		0.07		-	

[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).

and Clinton are given by 20, 3, and 5 TV ad spots, respectively. Contrary to the hypothesis that rallies are less important than TV ads in an election, these numbers show that rallies are an important tool of persuasion.[30] The cumulative effect of TV ads can outweigh that of rallies since rallies are time-constrained. For instance, the number of voters that changed their voting choice due to rallies in was 350K in 2012 and 320K in 2016 (see Section 7), whereas the number of voters who changed their vote due to TV ads was 2.2M in 2012, roughly 6 times higher than the effect of rallies.

I examine the in-sample and out-of-sample performance of the model in Tables 10 and 5, respectively. For the in-sample, the model's predicted average number of daily rallies lies in the 95% confidence intervals of the observed average number of daily rallies. The correlation for rally decisions predicted by the model and observed in the data varies from 69.5% (Trump) to 84% (for Clinton). Define prediction as the rally decision that has the maximum probability. Then, I find that the worst proportion of correct predictions is 74% for Trump and the highest proportion of correct predictions for Clinton is 86%.

To evaluate the out-of-sample model fit, refer to Table 5, I divide the data into two subsamples, training and validation. I randomly select (without replacement) 20% of the observations for the validation sample. I estimate the model on the remaining 80% of the observations for the training sample, and then calculate model fit metrics on the validation sample. The average number of daily rallies predicted by the model lies within one standard deviation from its observed counterpart in the validation sample. The worst correlation is

[30] However, the persuasion rates of rallies are still lower than that of TV news. For instance the persuasion rate for FOX News is $f = 11.6$ (DellaVigna and Kaplan, 2007).

Table 5: Out-of-Sample Fit

Panel (A): Comparison of Means

	Romney		Obama		Trump		Clinton	
	Model	Data	Model	Data	Model	Data	Model	Data
Southwest	0.173	0.1	0.18	0.1	0.116	0.2	0.0675	0
		0.14		0.14		0.2		0
Midwest	0.175	0.05	0.17	0.2	0.107	0.1	0.0697	0.05
		0.1		0.2		0.14		0.1
Northeast	0.273	0.15	0.238	0.1	0.336	0.45	0.199	0.25
		0.18		0.14		0.29		0.22
Southeast	0.376	0.25	0.28	0.1	0.323	0.2	0.206	0.3
		0.22		0.14		0.2		0.24

Panel (B): Measures of Fit

	Romney	Obama	Trump	Clinton
Correlation	0.8303	0.8412	0.7022	0.8157
Mean Squared Error	0.2583	0.2429	0.4060	0.2679
Correct Predictions	0.8375	0.8750	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. For each period, I define prediction as the option with the highest probability of choosing. Note that the worst correct predictions is 76%.

0.70 corresponding to Trump's rally decisions. I also calculate the correct predictions made by the model, which range from 76% for Trump to 87% for Obama.

6 Robustness

The model does not incorporate television ads, which might be correlated with political rallies. To address this concern, I follow Spenkuch and Toniatti (2018) and construct ad impression measures at the state level using data from Nielsen Ad Intel data and the Weselyan Media Project. Then I define popularity as

$$\text{Pop}_{dk} = \text{Poll Margin}_{dk} - \alpha_R^{\text{TV Ad}} \text{ R Ad Imp per capita}_{dk} + \alpha_D^{\text{TV Ad}} \text{ D Ad Imp per capita}_{dk}, \quad (6.1)$$

where Poll Margin_{dk} denotes R's lead in polls in state k on day d over D. $\text{R Ad Imp per capita}_{dk}$ and $\text{D Ad Imp per capita}_{dk}$ denote TV ad impressions per capita at the state level for R and D, respectively. The parameters $\alpha_R^{\text{TV Ad}}$ and $\alpha_D^{\text{TV Ad}}$ are the TV add effectiveness of estimates from the estimated values in Column (8) of Table VII in Spenkuch and Toniatti (2018). The resultant variable, Pop_{dk}, denotes poll margins for the estimation of the model, which is orthogonal to variation TV ads. The results from the estimation exercise are given in Columns (1) and (2) of Table 6. Note that these new estimates do not change significantly from the estimates reported in Table 3.

I also consider a different set of state groups. First, I reclassify North Carolina and Virginia with the Northeastern state group. This is entirely plausible, given that, North Carolina and Virginia are farther from Florida than from the Northeastern state group. Thus, for the Southeastern state group, I include only Florida. The estimates for this state group structure are given in Columns (3) and (4) of Table 6. For 2016, I do not find significant differences in the rally effectiveness estimates. For 2012, the rally effectiveness estimates decrease in magnitude, but remain significant. I also use the swing states used by Snyder and Yousaf (2020) in their event study of political rallies. The results of this exercise are given in Columns (5) and (6) in Table 6. For 2016, I do not find significant changes from the baseline rally effectiveness estimates. For 2012, these estimates decrease in magnitude but remain significant. I also test how robust my estimates are if I account for polling errors. Specifically, I calculate ex-post state-specific polling errors.[31] The results of this exercise are reported in Columns (11) and (12) of Table 6. Note that the estimates do not change significantly.

I also test how much estimates change when one considers spatial correlation and the presence of aggregate shocks. These features are accommodated by re-parameterizing the variance-covariance matrix of popularity shocks. For spatial autocorrelation, I assume that the popularity shocks of two state groups are correlated, and this correlation is inversely proportional to the distance between the state groups. In this case if Ω is the variance-covariance matrix of popularity shocks, then $\Omega_{ii} = \sigma_\nu^2$ and $\Omega_{ij} = \frac{\rho^2}{D_{i,j}}$ for $i \neq j$. Here parameter ρ_{corr} accounts for correlation and D_{ij} is the distance between state group i's and state group j's centroids in units of 1000 kilometers (KM). The results of this exercise are reported in Columns (9) and (10) of Table 6. I find that for Romney, Trump, and Clinton, the estimates do not change significantly compared to baseline. For Obama, rally effectiveness becomes insignificant but still preserves the same sign. Apart from spatial correlation, presence of

[31]This is calculated by comparing R's election-day poll margin obtained from FiveThirtyEight with the R's observed vote share margin (i.e., the election result). I calculate the differences for each state and correct for this error for each daily poll. I then use these corrected poll margins or estimation instead of the de-meaned poll margins in the baseline specification.

Table 6: Estimates

	Control fo T.V. Ads Spenkuch and Toniatti (2018)		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	2012	2016
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
α_R	0.0733	0.0862	0.0271	0.0796	0.037	0.0839	0.0724	0.0682	0.0817	0.065	0.0913	0.0745
	0.0127	0.0182	0.00458	0.0149	0.00628	0.0155	0.0108	0.0146	0.0136	0.0136	0.0224	0.0265
α_D	-0.0534	-0.0828	-0.021	-0.0593	-0.0322	-0.0745	-0.0496	-0.0582	-0.00436	-0.0575	-0.0569	-0.0789
	0.00923	0.019	0.00427	0.0113	0.00639	0.0152	0.00784	0.0127	0.00824	0.0129	0.0242	0.0156
ρ	0.987	1.01	0.991	0.988	0.987	0.991	0.986	0.989	0.994	0.99	0.994	0.992
	0.002	0.001	0.001	0.002	0.003	0.001	0.002	0.001	0.001	0.001	0.001	0.002
σ	0.157	0.166	0.0611	0.157	0.0901	0.16	0.128	0.121	0.143	0.154	0.147	0.16
	0.0133	0.0128	0.00729	0.0118	0.0062	0.0148	0.0112	0.00939	0.0132	0.0127	0.0146	0.0149
c_R	2.71	2.68	3.33	2.69	2.78	2.36	2.74	2.38	2.86	2.38	2.75	2.39
	0.268	0.241	0.271	0.235	0.291	0.208	0.275	0.198	0.277	0.201	0.275	0.208
c_D	2.83	3.55	3.44	3.49	2.92	3.26	2.85	3.19	2.63	3.21	2.84	3.33
	0.186	0.305	0.218	0.292	0.22	0.259	0.197	0.257	0.195	0.255	0.193	0.291
ρ_{corr}	-	-	-	-	-	-	-	-	0.005	0.01	-	-
	-	-	-	-	-	-	-	-	0.001	0.002	-	-
σ_{agg}	-	-	-	-	-	-	0.069	0.103	-	-	-	-
	-	-	-	-	-	-	0.0144	0.0138	-	-	-	-
Fixed Effects:												
Cost	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Poll Margins	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
LL	-694.05	-679.07	-285.64	-634.53	-468.54	-654.6	-642.4	-618.58	-653.34	-621.52	-670.18	-658.44
Observations	100	100	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) controls for T.V. ads by removing variation in poll margin data that can be explained by TV ad impressions. The effectiveness of T.V. ad impressions are calibrated using estimates from Spenkuch and Toniatti (2018). Columns (3) and (4) consider state groups where Florida constitutes Southeastern states and North Carolina along with Virginia are considered to be a part of Northeast states. Columns (5) and (6) consider states that are used by authors in Snyder and Yousaf (2020) for their event study. Columns (7) and (8) relax the assumption of uncorrelated popularity shocks and accommodates spatial autocorrelation in popularity. Columns (9) and (10) relaxes the assumption of uncorrelated popularity shocks and accommodates aggregate shocks in popularity. Columns (11) and (12) 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.

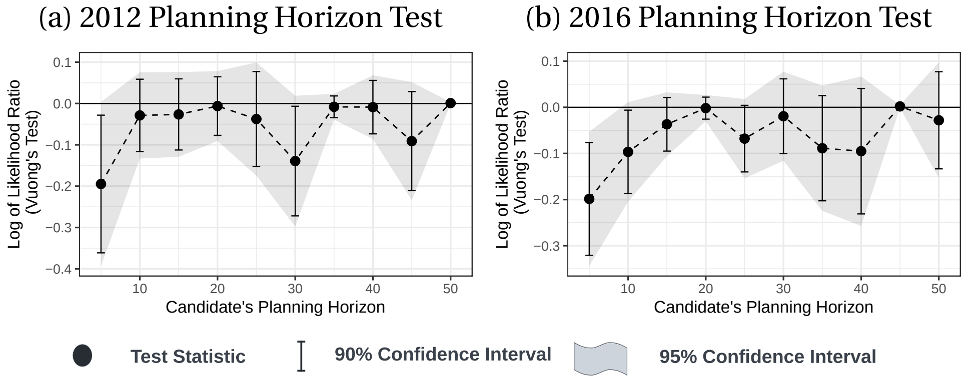


Figure 7: *This figure shows results from conducting Vuong's closeness test where the "Full Planning Horizon" model (baseline model) is compared with models that have shorter planning horizons. A negative statistic indicates that the baseline model performs better than the competing models. The x-axis shows the planning horizon length and the y-axis shows the corresponding Vuong's closeness test statistic.*

aggregate shocks can also lead to a correlation in popularity shocks. To mitigate this possibility, the variance-covariance matrix is parameterized to $\Omega_{ii} = \sigma_v^2 + \sigma_{agg}^2$ and $\Omega_{ij} = \sigma_{agg}^2$ for $i \neq j$. The estimates of this exercise are reported in Columns (7) and (8) of Table 6. I find that the estimates for the 2012 and 2016 elections are not significantly different from the baseline.

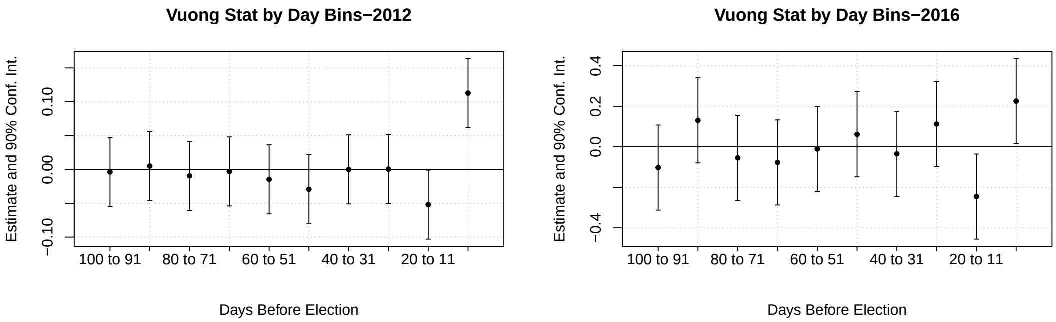
Table 7: Model Selection Tests

	Election Year		
	2012	2016	Pooled
Baseline Model vs Non-Strategic Candidates	0.000595	0.000359	0.000477
	0.010123	0.040219	0.0207
Baseline Model vs Max Win Prob	0.0545	0.0932	0.0739
	0.0613	0.0875	0.0533
Observations	100	100	200

[a] Note: The table shows the results from model selection tests between strategic and non strategic model. Moreover, It also shows the model selection test to infer whether candidates maximize winning probability or expected sum of electoral college votes. Positive values indicate the baseline model performs better.

I also test 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.[32] The second assumption is that candidates make rally decisions

[32]In this context (i) myopic candidates, (ii) a voter base that is more attentive closer to election, or (iii) a voter



(a) (b)

Figure 8: *Vuong Statistic by Day Bins. The figure shows that candidates are strategic especially 10 days prior to the election, while being non-strategic between 20 and 10 days before the election.*

strategically. Yet it is possible that candidates do not consider opponents' rally decisions and base their rally decisions only on a state's electoral payoff and their popularity in the state. I test whether this is true or not.[33]

Myopic campaign strategy models assume that candidates consider future payoffs 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 conduct Vuong's closeness test. The results of this exercise are reported in Figure 7. Myopic candidates will exhibit two key features: (1) their campaign strategies correlate with one another and with popularity only once the election period enters the planning horizon and not before it, and (2) a jump in the intensity of rallies when the election enters the planning horizon. These behavioral predictions allow us to determine whether candidates are myopic or not. Figure 7 reports the results of this exercise. A negative point estimate implies that the baseline model (full horizon model) performs better. The test statistics are not significantly positive. Moreover, horizons of 5 and 10 days are significantly lower (at 90% level of significance) than the baseline model.

base with limited memory span (do not recall earlier rallies) will yield observationally equivalent campaign strategy. Therefore, for the statistical test, it does not matter whether politicians are myopic or their target population is, we should see some jump in campaign activity as the election approaches.

[33]While researchers have considered strategic and myopic behavior in sequential voting settings, they have not studied 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. By contrast, in the present setting, these features are separable and can be individually tested.

I also test whether candidates make rally decisions strategically or not. For this purpose, I construct a model in which candidates neither anticipate nor observe the opponent's rally decisions and only consider popularity, time until the election, and electoral college votes. The results are reported in Table 7. The test statistic indicates that the strategic model performs better. However, it is not significant. Therefore, both models can explain the observations.[34] I also test whether the model where candidates maximize their probability of winning performs better than the one where candidates maximize the sum of electoral college votes. Vuong's test statistic fails to reject the sum of the electoral college votes model. The log-likelihood for the sum of electoral college votes model is higher than that for the winning probability of winning model.

7 Counterfactual Experiments

Cumulative Effect of Rallies: In this analysis, I evaluate electoral outcomes under two distinct scenarios. The first scenario, labeled "Only One Candidate Rallies," permits only one candidate to conduct political rallies, thereby isolating the influence of that candidate's campaigning efforts by eliminating the opponent's counter-campaigning response. The second scenario, "No One Rallies," predicts the electoral results when neither candidate holds rallies. Comparing these two scenarios helps isolate the specific effect of a candidate's total rally decisions on election outcomes, such as vote shares and winning probabilities.[35]

The findings, presented in Panels (a)–(f) of Figure 9, reveal significant increases in vote shares for all candidates. However, the impact of Trump's rallies notably exceeds that of Clinton's and Obama's. Specifically, Trump's rallies led to a substantial 33% increase in his probability of winning, whereas the rallies of other candidates did not significantly affect their probability of winning. This suggests that Trump's rallies were crucial, while those of

[34] A deeper look at how the test statistic varies with time shows that candidates are strategic 10 days before the election but non-strategic between 20 and 10 days before the election. Both models explain earlier pre-election bins with similar precision.

[35] To estimate the standard error of these counterfactual outcomes, I generate a sample (of size M) of parameter values from the asymptotic distribution, assumed to be normally distributed with a mean equal to the parameter estimates and a variance-covariance matrix provided by a consistent estimator of the parameters' variance-covariance. For each sample, I compute outcomes under both "None Rally" and "i Rallies", using the same methodology as before. The standard errors are then calculated using the differences $\left\{\Delta y^m : \Delta y^m = y^m_{i\,\text{Rallies}} - y^m_{\text{None rally}}\right\}_{m=1}^{M}$.

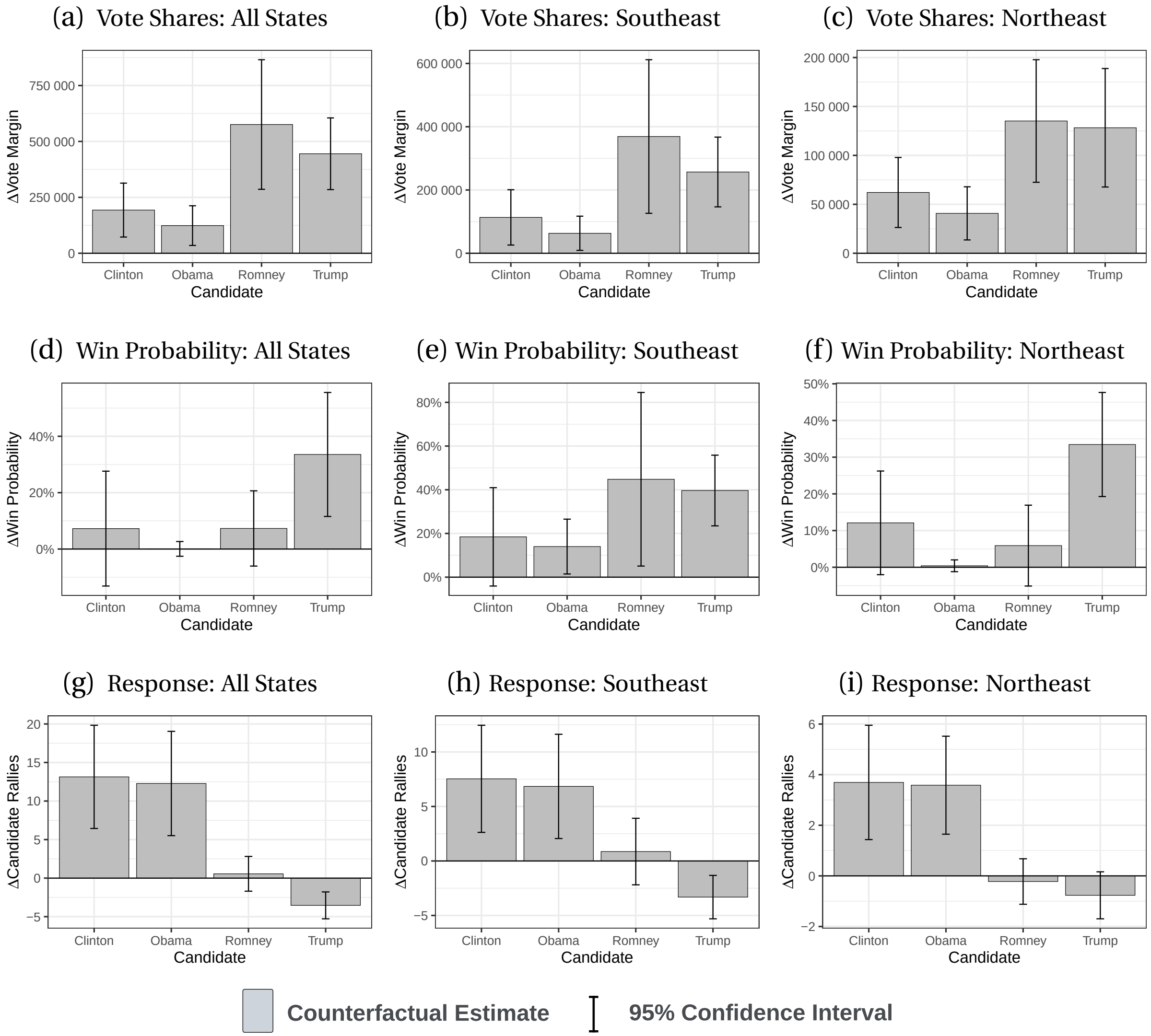


Figure 9: *This figure shows the cumulative effect of a candidate's rallies on their vote margin lead in Panels (a)–(c) and their probability of winning in Panels (d)–(f). Estimates of number of rallies held or not held as a campaign response to the opponent's rallies are shown in Panels (g)–(i).*

other candidates were not as impactful.

These findings contribute to the longstanding debate encapsulated by the question "Do campaigns matter?" as discussed in the seminal studies such as (Lazarsfeld et al., 1968; Berelson et al., 1986; Jacobson, 2015). My findings underscore that like widely studied campaign instruments like TV ads, political rallies can also be a decisive tool for presidential candidates, especially in competitive elections. Contrary to previous research suggesting a minimal impact of presidential campaigns on electoral outcomes (Franz and Ridout, 2010; Huber and Arceneaux, 2007; Jacobson, 2015), our findings align with political economy and quantitative marketing literature that suggests substantial effects of presidential campaigns on electoral results and voting behavior (Spenkuch and Toniatti, 2018; Gordon and Hartmann, 2013).

Campaign Response to Opponent: In this analysis, I estimate the number of rallies that were held, or not held, as a strategic response to an opponent's rallies. Specifically, I compare the total number of rallies conducted—either nationally or within certain state groups—when the opponent is also making rally decisions versus when the opponent does not hold any rallies. This comparison provides an insight into how campaign strategies adjust in response to the presence of an opponent. The results indicate significant strategic adjustments: Obama and Clinton conducted an additional 12.3 (3.45) and 13.14 (3.42) rallies, respectively, in response to rallies held by Romney and Trump. Conversely, Trump held 3.5 (0.89) fewer rallies in response to Clinton's campaigning, while Romney's campaign strategy did not alter the total number of his rallies. The findings, graphically illustrated in Panels (g)-(i) of Figure 9, quantify the extent to which strategic considerations shape campaign strategies.

Campaign Silence Laws: In this analysis, I introduce campaign silence periods ranging from 1 to 8 days.[36] This restriction modifies the continuation values for candidates during active campaigning periods, influencing their strategic behavior. The effectiveness of the campaign silence is contingent upon the total rallies held before its commencement, which determines the level of accumulated popularity. A short campaign silence results in a slight decay of this accumulated popularity, rendering the silence largely ineffective. Conversely, a prolonged silence leads to significant popularity decay, markedly altering electoral out-

[36]For simplicity, this discussion limits the campaign silence to 8 days, though initially, periods of up to 30 days were considered. The impact of campaign silence on election results tends to increase slightly for 2016 and remains relatively stable for 2012.

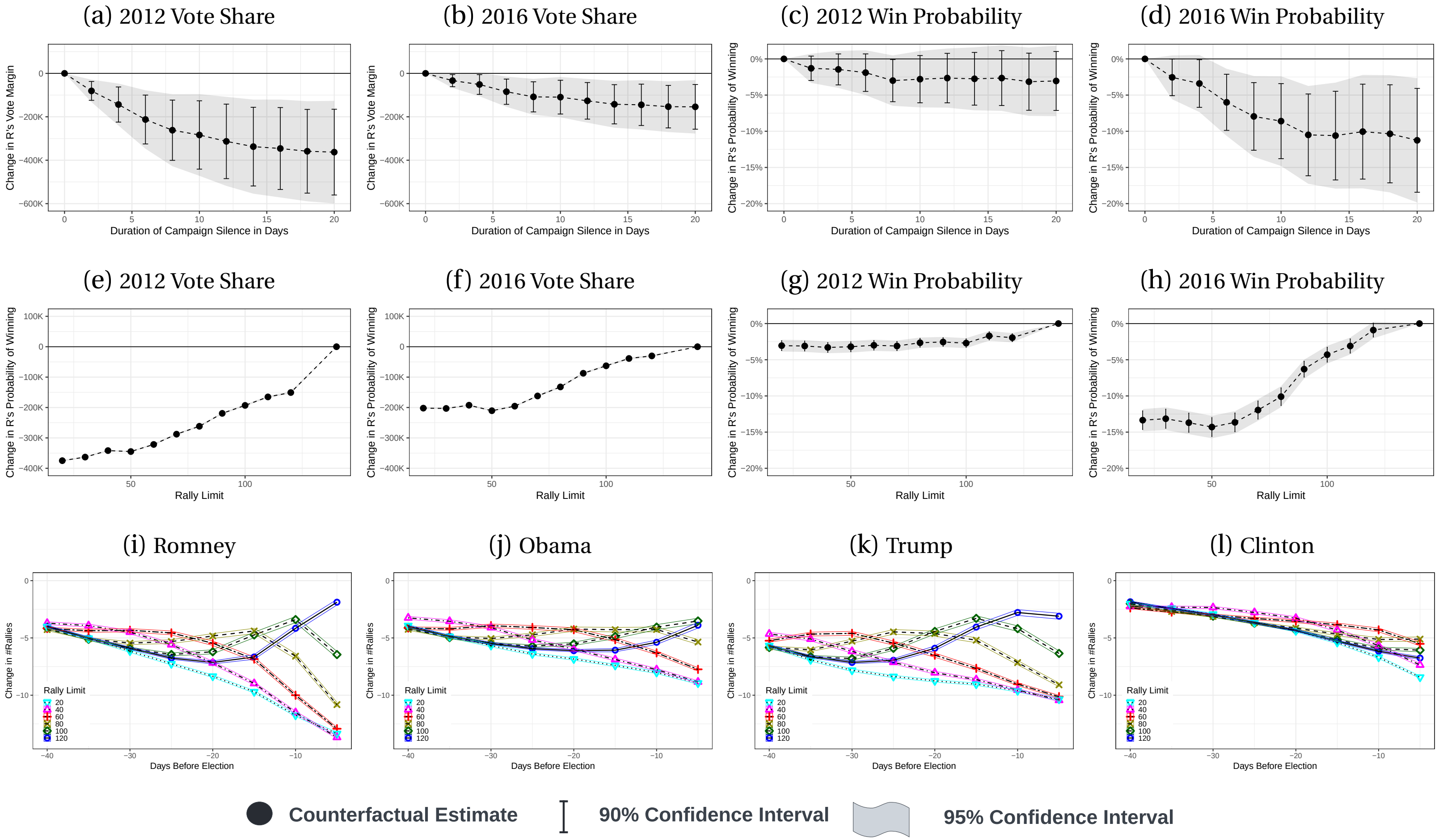


Figure 10: *Campaign Silence Duration and Election Day Poll Margin for 2012 Presidential Elections. This figure provides estimates for changes in electoral outcomes when campaign silences of varying duration are imposed. For each campaign silence duration, I calculate R's probability of winning along with the corresponding confidence intervals.*

comes.

For each duration of campaign silence, I compare the electoral outcomes with the electoral outcomes that would have been obtained if there had been no campaign silence. My findings suggest that campaign silences are largely ineffective in less competitive elections. However, in highly competitive elections, campaign silences can significantly impact outcomes. State-specific parameter estimates indicate that the 2012 election was less competitive than 2016. Notably, even in competitive elections, shorter periods of campaign silence prove ineffective, as shown in Panels (a)-(d) of Figure 10.[37]

Spending Limits: In this analysis, I impose limits on the number of rallies each candidate can hold, akin to spending limits in this context. I assume that the average monetary cost of rallies is consistent across candidates.[38] Limits are set within a range from 20 to 120. I then simulate the model and compare the equilibrium outcomes with a baseline scenario with no rally limits.

My findings indicate that even when the imposed limits on rallies exceed the actual number of rallies candidates hold, there is a noticeable decrease in both the probability of winning and vote shares across both election years. The decrease in these metrics is monotonic, eventually stabilizing at different levels for each year, as depicted in Panels (e)-(h) of Figure 10. Further analysis of equilibrium strategies shows that even if rally limits are higher than the number of rallies held without any rally limits, they still suppress the number of rallies throughout all periods. This reduction occurs because candidates, anticipating potential future contingencies that may require intensive rallying, choose to conserve their limited rally opportunities early on, even when these limits seem generous initially.[39] Panels (i)–(l) of Figure 10 present the results of this exercise.

8 Conclusion

This paper shows that political rallies can be persuasive, electorally pivotal, and challenging to regulate, even in a consolidated democracy. I show this in two steps. In the first step,

[37]Furthermore, the influence on campaign strategy is minimal until the onset of the blackout period, due to which I have omitted the figure.

[38]Cost estimates reported in Table 3 include the monetary and opportunity costs of holding rallies, which may differ among candidates.

[39]This behavior is absent in static models and only plays a role when one considers dynamic strategies.

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.

The analysis of electoral effects reveals that Trump's rallies significantly increased his probability of winning by 33%, whereas the rallies of Romney, Clinton, and Obama did not increase their probability of winning. Additionally, the persuasion rates of rallies compared to television ads (DellaVigna and Kaplan, 2007; Spenkuch and Toniatti, 2018) demonstrate that a single rally is more effective than a TV ad. However, due to the limited number of rallies a candidate can realistically hold, their cumulative impact falls short of that of TV ads. The study also evaluates policy-relevant counterfactual experiments, finding that short campaign silences—commonly lasting only 1–2 days in many countries—are ineffective due to their brief duration. Conversely, ostensibly nonbinding spending limits effectively reduce the influence of extensive campaigning on electoral outcomes.

A Proofs

A.1 Proof for Proposition 2.1

Proof: First note that under assumption 2.1, 2.2, and 2.3 the relevant state variables for first mover is cost shocks and current popularity and for second mover in addition to these two first mover action is also a relevant state variable. Then consider period T, since popularity shocks are normally distributed we can write:

$$\mathbb{E}\left[V_{R,T+1}(p_{T+1})|a_{RT},a_{DT},p_T\right] = \sum_{k=1}^{K} e_k E\Phi\left(\frac{\alpha_R a_{RkT} + \alpha_D a_{DkT} + \rho p_T + \delta_k}{\sigma_\nu}\right), \tag{A.1}$$

where $\Phi()$ is the Standard Normal pdf. Note that $\sum_{k=1}^{K} e_k E\Phi\left(\frac{\alpha_R a_{RkT}+\alpha_D a_{DkT}+\rho p_T+\delta_k}{\sigma_\nu}\right)$ is bounded by E. For D, $\mathbb{E}\left[V_{D,T+1}(p_{T+1})|\dots\right] = E - \mathbb{E}\left[V_{R,T+1}(p_{T+1})|\dots\right]$ and it is also bounded. The second mover's option specific value function u_{isT} is given as:

$$u_{isT}(k,p_T,a_{jfT}) = -c_i\mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[V_{i,T+1}(p_{T+1})|a_{jfT},k,p_T\right], \tag{A.2}$$

which is also finite. Note that $a_{isT} = \arg\max_k\left\{u_{isT}(k,p_T,a_{jfT}) + \epsilon_{i,k,s,T}\right\}$ which exists as each of the numbers are finite. Moreover, it is unique with probability 1 otherwise we will have two independent T1EV random variables satisfying an equality with positive probability. Moreover, from McFadden (1989) we know that $Prob(a_{isT} = k)$ will take the standard multinomial logistic form given by $\sigma_{isT}(k;p_T,a_{jfT}) = \frac{\exp(u_{isT}(k,p_T,a_{jfT}))}{\sum_l \exp(u_{isT}(l,p_T,a_{jfT}))}$. Similarly, for the first mover the option specific value function is given by:

$$u_{ifT}(k,p_T) = -c_i\mathbb{1}\{k \neq 0\} + \beta\sum_{l=0}^{K}\sigma_{jsT}(l,p_T,k)\mathbb{E}\left[V_{i,T+1}(p_{T+1})|l,k,p_T\right]. \tag{A.3}$$

Then with the same arguments one can show that $\sigma_{ifT}(k;p_T) = \frac{\exp(u_{ifT}(k,p_T))}{\sum_l \exp(u_{ifT}(l,p_T))}$. Then the expected value for the second mover and first mover is given by:

$$\mathbb{E}_{\epsilon_{is,T}}\left[V_{isT}(l,p_T,\epsilon_{is,T})\right] = \log\left(\sum_{k=0}^{K} u_{isT}(k,p_T,l)\right) + \gamma, \tag{A.4}$$

$$\mathbb{E}_{\epsilon_{if,T}}\left[V_{ifT}(p_T,\epsilon_{is,T})\right] = \log\left(\sum_{k=0}^{K} u_{ifT}(k,p_T)\right) + \gamma. \tag{A.5}$$

This true because the cost shocks are T1EV.[40] Moreover, note that since u_{isT} and u_{ifT} are bounded their respected expected value functions are also bounded because $\log(\sum_j exp(x_k)) \leq \log((K+1)exp(\max_j(x_k))) = \max_j(x_k)+\log(K+1)$. Replacing x_k with $u_{isT}(k)$ or $u_{ifT}(k)$ proves the claim. The value function in period T is given by:

$$\begin{aligned} V_{i,T}(p_T) =& f_i \times \ln\left(\sum_{k=0}^{K} \exp\left\{u_{if,t}(k;p_T)\right\}\right) \\ &+ (1-f_i) \times \sum_{k=0}^{K}\left[\sigma_{jf,T}(k;p_T)\ln\left(\sum_{l=0}^{K}\exp\left\{u_{is,T}(l;k,p_T))\right\}\right)\right] + \gamma. \end{aligned} \tag{A.6}$$

Note this value function is also bounded because each individual term is bounded.

Suppose V_{t+1} is bounded then with similar arguments $u_{ist}(k,p_t,a_{jft}) = -c_i \mathbb{1}\{k \neq 0\} + \beta\mathbb{E}\left[V_{i,t+1}(p_{t+1})|a_{jft},k,p_t\right]$ is also bounded and so is $u_{ift} = -c_i\mathbb{1}\{k\neq 0\} + \beta\sum_{l=0}^{K}\sigma_{jst}(l,p_t,k)$ $\mathbb{E}\left[V_{i,t+1}(p_{t+1})|l,k,p_t\right]$. Then the respective policy functions a_{ist} and a_{ift} also exist and are unique almost surely. Moreover, the conditional choice probabilities take the form $\sigma_{ist}(k;p_t,a_{jft})$ $= \frac{\exp(u_{ist}(k,p_t,a_{jft}))}{\sum_l \exp(u_{ist}(l,p_t,a_{jft}))}$ and $\sigma_{ift}(k;p_t) = \frac{\exp(u_{ift}(k,p_t))}{\sum_l \exp(u_{ift}(l,p_t))}$. Given this we can show that the following holds:

$$\mathbb{E}_{\epsilon_{is,t}}\left[V_{ist}(l,p_t,\epsilon_{is,t})\right] = \log\left(\sum_{k=0}^{K} u_{ist}(k,p_t,l)\right) + \gamma, \tag{A.7}$$

$$\mathbb{E}_{\epsilon_{if,t}}\left[V_{ift}(p_t,\epsilon_{is,t})\right] = \log\left(\sum_{k=0}^{K} u_{ifT}(k,p_t)\right) + \gamma, \tag{A.8}$$

$$\begin{aligned} V_{i,t}(p_t) =& f_i \times \ln\left(\sum_{k=0}^{K} \exp\left\{u_{if,t}(k;p_t)\right\}\right) \\ &+ (1-f_i) \times \sum_{k=0}^{K}\left[\sigma_{jf,t}(k;p_t)\ln\left(\sum_{l=0}^{K}\exp\left\{u_{is,t}(l;k,p_t))\right\}\right)\right] + \gamma. \end{aligned} \tag{A.9}$$

Given the above equation, V_{it} is also bounded using similar arguments that were used to argue $V_{i,T}$ is bounded. Then by induction argument the above holds for all $t = 1,2,\ldots,T$.

[40]Note if $-\epsilon_k \sim T1EV$ and independent then $\max_{k\in\{0,1,\ldots,K\}}\{\tilde{\delta}_k - \epsilon_{iftk}\} \sim$ Gumbel($\mu = \ln\sum_k \exp(\tilde{\delta}_k)$, $\beta = 1$). Where μ denotes the location parameter of a Gumbel Distribution and β denotes the scale parameter and the mean is given by $\mu + \beta\gamma$ where γ is Euler's constant.

A.1.1 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:

$$\mathbb{P}\left[\tilde{X}_t \in B|\tilde{X}_{t-1}\right] = \sum_{a \in \{0,1,\ldots,K\}^2} \mathbb{P}\left[A_t = a, P_{t+1} \in B_a|(A_{t-1}, P_t)\right] \tag{A.10}$$

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 $\{\{0, 1, \ldots, K\}^2 \times \mathbb{R}^K\}$ and $B_a = \{p \in \mathbb{R}^K : (a, p) \in B\}$. 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:

$$\mathbb{P}\left[P_{t+1} \in B_a|(A_t = a, P_t)\right] = \int_{p \in B_a} f\left(p|A_t = a, P_t\right) dp \tag{A.11}$$

Where $f(.|.)$ is the density of popularity. Also note that the equilibrium defines $\mathbb{P}\left[A_t = a|P_t\right] = \sigma_t\left(A_t = a; P_t\right)$ Therefore, we can express $\mathbb{P}\left[\tilde{X}_t \in B|(A_{t-1}, P_t)\right]$ as followed:

$$\begin{aligned}
\mathbb{P}\left[\tilde{X}_t \in B|\tilde{X}_{t-1}\right] &= \sum_{a \in \{0,1,\ldots,K\}^2} \mathbb{P}\left[A_t = a, P_{t+1} \in B_a|(A_{t-1}, P_t)\right] \\
&= \sum_{a \in \{0,1,\ldots,K\}^2} \mathbb{P}\left[P_{t+1} \in B_a|(A_t = a, P_t, A_{t-1})\right] \mathbb{P}\left[A_t = a|P_t, A_{t-1}\right] \\
&= \sum_{a \in \{0,1,\ldots,K\}^2} \mathbb{P}\left[P_{t+1} \in B_a|(A_t = a, P_t)\right] \mathbb{P}\left[A_t = a|P_t\right] \\
&= \sum_{a \in \{0,1,\ldots,K\}^2} \left(\int_{p \in B_a} f\left(p|A_t = a, P_t\right) dp\right) \sigma_t\left(A_t = a; P_t\right) \\
&= \sum_{a \in \{0,1,\ldots,K\}^2} \int_{p \in B_a} f\left(p|A_t = a, P_t\right) \sigma_t\left(A_t = a; P_t\right) dp \\
\Rightarrow \int_{x \in B} \psi_t(x|X_{t-1}) dx &= \int_{(a,p) \in B} f\left(p|A_t = a, P_t\right) \sigma_t\left(A_t = a; P_t\right) d(a, p)
\end{aligned} \tag{A.12}$$

The second equality holds by law of total probability. The third equality holds by modeling assumption 2.1. The fourth equality is substituting $\mathbb{P}\left[P_{t+1} \in B_a|(A_t = a, P_t)\right]$ using assumption 2.1. The fifth and sixth inequality is a re-writing of the integral.

Q.E.D.

A.1.2 Proof for Proposition 4.1

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

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

Finally we have

$$
\begin{aligned}
&\Rightarrow \mathbb{P}\left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3}\right) \in B \middle| \tilde{X}_{4d-4}\right] \\
&= \int_{(a_l,p_{l+1})_{l=1,\ldots,4} \in B} \left(\prod_{l=1}^{4} \sigma_{4(d-1)+l}(a_l;p_l)\right) \times \left(\prod_{l=1}^{4} f\left(p_{l+1}|a_l,p_l\right)\right) d\left((a_l,p_{l+1})_{l=1,\ldots,4}\right)
\end{aligned} \tag{A.13}
$$

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 $\left\{\{0,1,\ldots,K\}^8 \times \mathbb{R}^{4K}\right\}$ under the respective product sigma-algebra and the following holds

$$
\begin{aligned}
\mathbb{P}\left[X_d \in C | X_{d-1}\right] &= \mathbb{P}\left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3}\right) \in B^C \middle| X_{d-1}\right] \\
&= \mathbb{P}\left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3}\right) \in B^C | (P_{4d-3}, A_{4d-4})\right] \\
&= \mathbb{P}\left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3}\right) \in B^C | \tilde{X}_{4d-4}\right] \\
&= \int_{(a_l,p_{l+1})_{l=1,\ldots,4} \in B^C} \left(\prod_{l=1}^{4} \sigma_{4(d-1)+l}(a_l;p_l)\right) \times \left(\prod_{l=1}^{4} f\left(p_{l+1}|a_l,p_l\right)\right) d\left((a_l,p_{l+1})_{l=1,\ldots,4}\right) \\
&= \int_{x \in C} \left(\int_{p_2,p_3,p_4 \in \mathbb{R}^{3K}} \left(\prod_{l=1}^{4} \sigma_{4(d-1)+l}(a_l;p_l)\right) \times \left(\prod_{l=1}^{4} f\left(p_{l+1}|a_l,p_l\right)\right) d(p_1,p_2,p_3) \right) d(x)
\end{aligned} \tag{A.14}
$$

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 $\tilde{X}_{4d-4}$ as its component and the corresponding probability is well defined by equation A.13. The following equality is substitution of the expression found in A.13. The last equality is a 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}.

B 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[41]. I also removed rallies from states where less than two rallies were held at

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

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. 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 · *'days before election'* + *'period within the day'*.

Table 8 shows the total number of activities that were available for 120 days before the election. For the main analysis and model estimation, I focus on activities in the last 100 days. 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.

C Simulated Likelihood Procedure

C.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 $\lambda_d^\theta(X_d \big| X_{d-1})$. Lets denote the set of probability integral transforms of $3K$ dimensional Sobol sequence, till M, by $\zeta = \left\{\zeta^m = (\zeta_{1,1}^m, \ldots, \zeta_{1,K}^m, \ldots, \zeta_{3,1}^m, \ldots, \zeta_{3,K}^m)\right\}_{m=1}^M$ [42]. Based on ζ we can define the following set of plausible popularity values:

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

[42]Here $\zeta_{l,k}^m$ is $\Phi^{-1}\left(u_{\text{Sobol},l,k}^m\right)$, where $u_{\text{Sobol},l,k}^m$ 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.

Table 8: 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.

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

$$\hat{p}^{m,d}_{l+1,k} = \alpha_R \mathbb{1}\{A_{R,d,l} == k\} + \alpha_D \mathbb{1}\{A_{D,d,l} == k\} + \tilde{\alpha}\mathbb{1}\{A_{R,d,l} == k, A_{D,d,l} == k\} + \rho \hat{p}^{m,d}_{l,k} + \delta_k + \sigma_v \zeta^m_{l,k} \quad \text{(C.2)}$$

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:

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

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

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

Where $\hat{p}^{m,d}_{l} = (\hat{p}^{m,d}_{l,1}, \ldots, \hat{p}^{m,d}_{l,K})$, the function $\hat{\sigma}_{4(d-1)+l}(A_{d,l};\hat{p}^{m,d}_{l})$ is approximate probability of observing action profile $A_{d,l}$ in period $4*(d-1)+l$. This term is evaluated by using equations D.10 and D.8 in the Online Appendix. Moreover, $\phi(.)$ 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. Finally, the approximate log-likelihood is given by:

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

References

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

Aguirregabiria, V. and M. Marcoux (2021). Imposing equilibrium restrictions in the estimation of dynamic discrete games. *Quantitative Economics 12*(4), 1223–1271.

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

Bajari, P., C. L. Benkard, and J. Levin (2007). Estimating dynamic models of imperfect competition. *Econometrica 75*(5), 1331–1370.

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.

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

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

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.

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

Egesdal, M., Z. Lai, and C.-L. Su (2015). Estimating dynamic discrete-choice games of incomplete information. *Quantitative Economics 6*(3), 567–597.

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.

Grosjean, P., F. Masera, and H. Yousaf (2022). Inflammatory political campaigns and racial bias in policing.

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.

Iaryczower, M., G. Lopez-Moctezuma, and A. Meirowitz (2022). Career concerns and the dynamics of electoral accountability. *American Journal of Political Science*.

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

Igami, M. (2017). Estimating the innovator's dilemma: Structural analysis of creative destruction in the hard disk drive industry, 1981–1998. *Journal of Political Economy 125*(3), 798–847.

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.

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

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*, 995.

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

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.

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.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, 999–1033.

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.

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.

Online Appendix: Not For Publication

D Numerical Approximation

D.1 Primitives

Let $\mathbf{P} = \{(\mathbf{p}_1^r, \ldots, \mathbf{p}_K^r)\}_{r=1}^R$ be the state variable grid. Let $\mathbf{T}(\mathbf{p}) = \Big(\mathbf{T}_1(\mathbf{p}), \mathbf{T}_2(\mathbf{p}), \ldots, \mathbf{T}_R(\mathbf{p})\Big)$ 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. Let the coefficients of the polynomial terms approximating $V_{i,t}(.)$ be denoted by $\gamma_{i,t}^V$, $u_{i,f,t}(k;.)$ 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.

D.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.

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

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

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

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

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

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

D.3 Period 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 $\mathbf{p} \in \mathbf{P}$ such that $T_{ij} = T_j(p^i)$ where p^i is the i^{th} point in $\mathbf{P}$ and T_j is the $j-th$ order Chebyshev polynomial. 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 $\tilde{V}_{i,t}$, $\tilde{u}_{i,f,k,t}$, and $\tilde{u}_{i,s,k,l,t}$ and call them $\gamma^V_{i,t}$, $\gamma^f_{i,k,t}$, and $\gamma^s_{i,k,l,t}$. Once, we have obtained these coefficients, it allows us to interpolate the value functions and conditional choice probabilities at any given popularity standing p. Then it is straightforward to compute the approximated values of the value functions. Moreover, the conditional choice probabilities are also straightforward to calculate. This property will be used extensively in the next subsection.

D.4 Period 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:

$$I^{r'}_{r,k,l} = \sum_{s=1}^{S}\mathbb{T}_{r'}\left(\begin{bmatrix}\alpha_R\mathbb{1}\{k==1\} + \alpha_D\mathbb{1}\{l==1\} + \tilde{\alpha}\mathbb{1}\{k==1, l==1\} + \rho\mathbf{p}^r_1 + \delta_1 + \sigma_\nu\nu^s_1 \\ \vdots \\ \alpha_R\mathbb{1}\{k==K\} + \alpha_D\mathbb{1}\{l==K\} + \tilde{\alpha}\mathbb{1}\{k==K, l==K\} + \rho\mathbf{p}^r_K + \delta_K + \sigma_\nu\nu^s_K\end{bmatrix}\right)\omega^s \tag{D.6}$$

Here, $\nu^s_1, \ldots, \nu^s_K$ are Gaussian shocks and w^s is the weight of these shocks. The choice of the Gaussian quadrature is discussed later. Consequently define $I_{r,k,l} = (I^1_{r,k,l}, I^2_{r,k,l}, \ldots, I^R_{r,k,l})$.

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:

$$\tilde{u}_{i,s,t}(a_i;a_j,\mathbf{p}^r) = -c_i(1-a_{i,0}) + \beta\sum_{r_2=1}^{R}\gamma^V_{i,t+1;r_2}I^{r_2}_{r,a_i,a_j} \tag{D.7}$$

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:

$$\tilde{\sigma}_{i,s,t}(a_i;a_j,\mathbf{p}^r) = \frac{\exp\left(-c_i(1-a_{i,0}) + \beta\sum_{r_2=1}^{R}\gamma^V_{i,t+1,r_2}\left(I^{r_2}_{r,a_i,a_j} - I^{r_2}_{r,0,a_j}\right)\right)}{1+\sum_{k=1}^{K}\exp\left(-c_i + \beta\sum_{r_2=1}^{R}\gamma^V_{i,t+1,r_2}\left(I^{r_2}_{r,k,a_j} - I^{r_2}_{r,0,a_j}\right)\right)} \tag{D.8}$$

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

$$\tilde{u}_{i,f,t}(a_i;a_j,\mathbf{p}^r) = -c_i(1-a_{i,0}) + \beta\sum_{a_j=0}^{K}\left(\sum_{r_2=1}^{R}\gamma^V_{i,t+1;r_2}I^{r_2}_{r,a_i,a_j}\right)\times\tilde{\sigma}_{j,s,t}(a_j;a_i,\mathbf{p}^r) \tag{D.9}$$

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 $\tilde{\sigma}_{i,f,t}(a_i;\mathbf{p}^r)$.

$$\tilde{\sigma}_{i,f,t}(a_i;\mathbf{p}^r) = \frac{\exp\left(-c_i(1-a_{i,0}) + \beta\sum_{r_2=1}^{R}\gamma^V_{i,t+1,r_2}\left(\sum_{a_j=0}^{K}(I^{r_2}_{r,a_i,a_j}\tilde{\sigma}_{j,s,t}(a_j;a_i,\mathbf{p}^r) - I^{r_2}_{r,0,a_j}\tilde{\sigma}_{j,s,t}(a_j;0,\mathbf{p}^r)\right)\right)}{1+\sum_{k=1}^{K}\exp\left(-c_i + \beta\sum_{r_2=1}^{R}\gamma^V_{i,t+1,r_2}\left(\sum_{a_j=0}^{K}(I^{r_2}_{r,k,a_j}\tilde{\sigma}_{j,s,t}(a_j;k,\mathbf{p}^r) - I^{r_2}_{r,0,a_j}\tilde{\sigma}_{j,s,t}(a_j;0,\mathbf{p}^r)\right)\right)} \tag{D.10}$$

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

$$\tilde{V}_{i,t}(\mathbf{p}^r) = f_i\log\left(\sum_{a_i=0}^{K}\exp\left(\tilde{u}_{i,f,t}(a_i;\mathbf{p}^r)\right)\right) + (1-f_i)\sum_{a_j=0}^{K}\log\left(\sum_{a_i=0}^{K}\tilde{u}_{i,s,t}(a_i;a_j,\mathbf{p}^r)\right)\times\tilde{\sigma}_{j,s,t}(a_j;a_i,\mathbf{p}^r) \tag{D.11}$$

D.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, $\mathbf{U} = \{(\mathbf{u}^r_1,\ldots,\mathbf{u}^r_K)\}_{r=1}^{R}$ over

$[-1,1]^K$ and its corresponding Chebyshev polynomial $\mathbf{\Psi}(u)$. Then the grid $\mathbf{P} = \{(\mathbf{p}_1^r, \ldots, \mathbf{p}_K^r)\}_{r=1}^R$ for a given set of parameters $\rho, \sigma_\nu, \delta_1, \ldots, \delta_K$ is constructed as followed:

$$\begin{aligned} \mathbf{p}^r &= \underline{p}_k + (\bar{p}_k - \underline{p}_k)\frac{\mathbf{u}_k^r + 1}{2} \\ \text{where} \quad \bar{p}_k &= \frac{\delta_k}{1-\rho} + \left(\frac{\alpha_R}{1-\rho} + \frac{3\sigma_\nu}{\sqrt{1-\rho^2}}\right) \\ \underline{p}_k &= \frac{\delta_k + \alpha_D}{1-\rho} + \left(\frac{\alpha_D}{1-\rho} - \frac{3\sigma_\nu}{\sqrt{1-\rho^2}}\right) \end{aligned} \tag{D.12}$$

The Chebyshev polynomial $\mathbf{T}(p) = (\mathbf{T}_1(p), \mathbf{T}_2(p), \ldots, \mathbf{T}_R(p))$ is defined as followed:

$$\mathbf{T}_r(p) = \mathbf{\Psi}_r\left(2\Big(\frac{p_1 - \underline{p}_1}{\bar{p}_1 - \underline{p}_1}\Big) - 1, 2\Big(\frac{p_2 - \underline{p}_2}{\bar{p}_2 + \underline{p}_2}\Big) - 1, \ldots, 2\Big(\frac{p_K - \underline{p}_K}{\bar{p}_K + \underline{p}_K}\Big) - 1\right) \tag{D.13}$$

Where $\mathbf{\Psi}_r(.)$ is the r^{th} Chebyshev polynomial term. The Gaussian quadrature, denoted by $\nu = \{(\nu_1^s, \ldots, \nu_K^s, \omega^s)\}_{s=1}^S$, 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 $2^\mu + 1$.

D.6 Algorithm

Here I will describe the algorithm that is used to solve the game using the equations discussed above. The algorithm will be defined for a given parameter values, $\theta_{\text{Popularity}} = \Big\{ \alpha_R, \alpha_D, \tilde{\alpha}, \rho, \sigma_\nu, \delta_1, \delta_2, \ldots, \delta_K \Big\}$ and $\theta_{\text{Cost}} = \Big\{ c_R, c_D, c_1, c_2, \ldots, c_K \Big\}$ 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 $\nu = \{(\nu_1^s, \ldots, \nu_K^s, w^s)\}_{s=1}^S$ from http://www.sparse-grids.de/ for K dimensions and approximation level $2^\mu +$ 1.

Step 1 Compute the parameter-specific $\mathbf{P}, \mathbf{T}$ using equations D.12 and D.13.

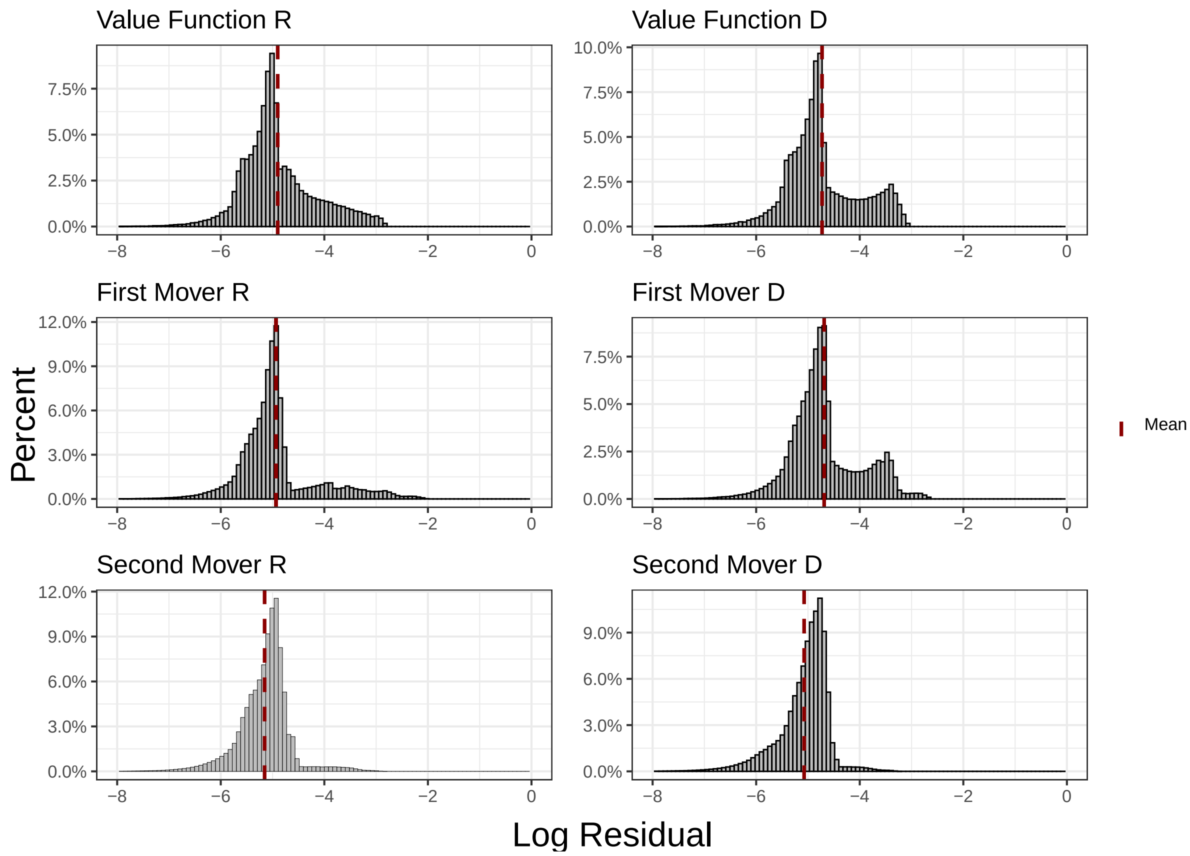
Step 2 Pre-Compute integrals of Chebyshev terms contingent on current popularity and candidate decisions by equation D.6.

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 D.5, D.3, D.1,D.4, D.2 respectively.

Figure 11: 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.

- 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 $k, l = 0, \ldots, K$ 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 D.11, D.9, D.7,D.10, D.8 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 $k, l = 0, \ldots, K$ for period $t = T$ respectively.

D.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 $p_{t,m} = (p_{1,t,m}, \ldots, p_{K,t,m})$, let $Gp_{k,l}(p) = E[p_{t+1}|a_{R,t} =$

$k, a_{D,t} = l, p]$ and define

$$u_{i,s,t}^{t+1}(k; l, p_{t,m}) = -c_i\left(1 - a_{i,0}\right) + \beta \sum_{r=1}^{R} \gamma_{i,t+1;r}^{V} \left(\sum_{s=1}^{S} \mathbf{T}_r\left(Gp_{k,l}(p_{t,m} + \sigma_v v^s)\right) \omega^s \right) \tag{D.14}$$

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:

$$\mathcal{R}_{i,s,k,l,t}(\gamma; p_{t,m}) = 1 - \frac{u_{i,s,t}^{t+1}(k; l, p_{t,m})}{\hat{u}_{i,s,t}(k; l, p_{t,m})} \quad \text{for all } i, k, l, t \tag{D.15}$$

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

$$u_{i,f,t}^{t+1}(k; p_{t,m}) = \sum_{l=1}^{k} u_{i,s,t}^{t+1}(k; l, p_{m,t}) \left(\frac{\exp\left(u_{j,s,t}^{t+1}(l; k, p_{m,t}) - u_{j,s,t}^{t+1}(0; k, p_{m,t})\right)}{1 + \sum_{l'=1}^{K} \exp\left(u_{j,s,t}^{t+1}(l'; k, p_{m,t}) - u_{j,s,t}^{t+1}(0; k, p_{m,t})\right)} \right) \tag{D.16}$$

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

$$\mathcal{R}_{i,f,k,t}(\gamma; p_{t,m}) = 1 - \frac{u_{i,f,t}^{t+1}(k; p_{t,m})}{\hat{u}_{i,f,t}(k; p_{t,m})} \quad \text{for all } i, k, l, t \tag{D.17}$$

In similar fashion one can define $V_{i,t}^{t+1}(p_{t,m})$ 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 11 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.

E Algorithm for Evaluating the Approximate Likelihood

E.1 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 in sections D and C. The algorithm will be defined for a given parameter values, $\theta = \{\alpha_R, \alpha_D, \tilde{\alpha}, \rho, \sigma_v, \delta_1 \ldots, \delta_K, c_R, c_D, c_1, \ldots, c_K\}$ and the approximation level μ. The steps of the algorithm are below:

Step 0-3 Execute steps 0-3 from the Algorithm described in Online Appendix Subsection D.6.

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 C.1, C.2 and C.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 D.10 and D.8.
- Calculate $\hat{\lambda}_d^{\theta}\left(X_d \middle| X_{d-1}\right)$ using equation C.4.

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

F Addition Figures

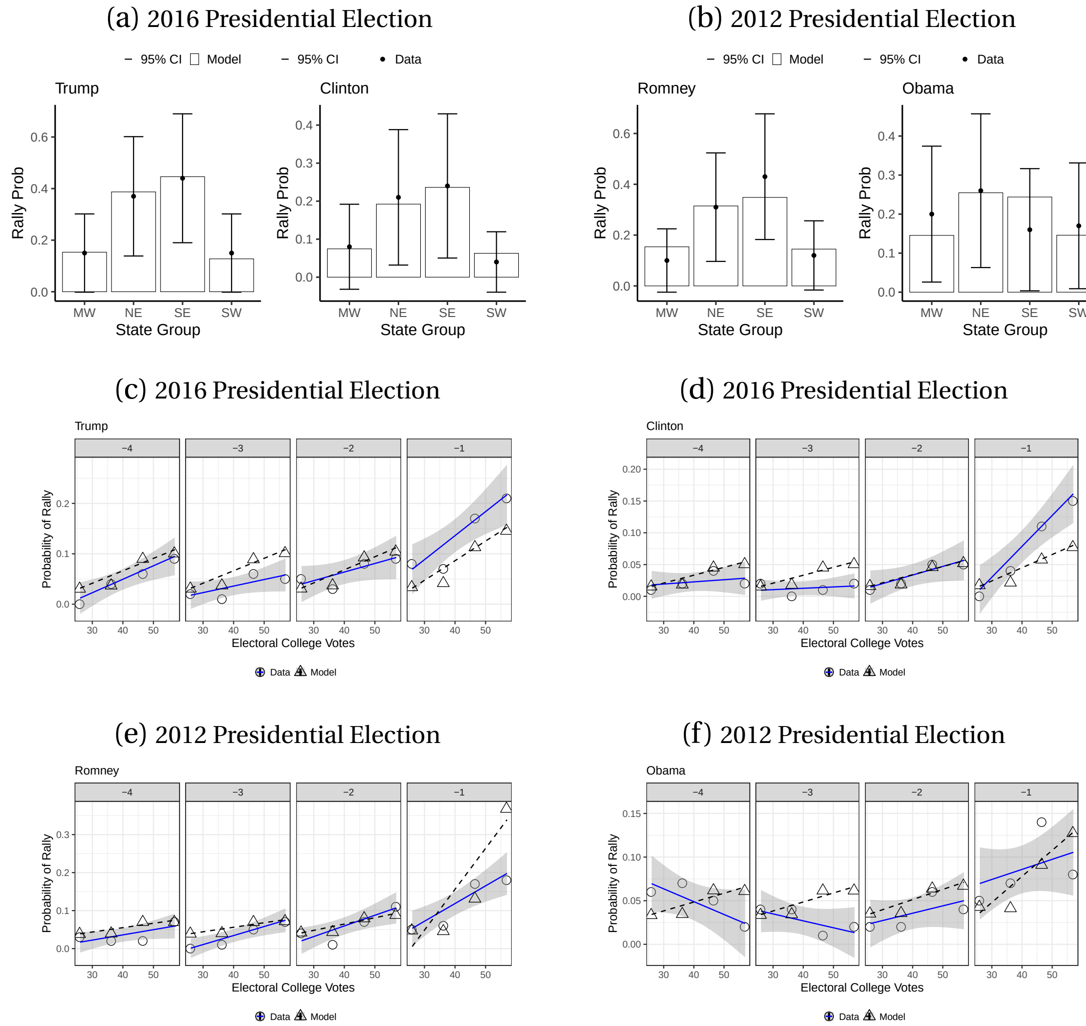


Figure 12: *Panels (a) and (b) plots the model performance within the sample. The columns represent the model prediction, the solid points show observed probability of rallies for these groups and error-bars represent the 95% confidence intervals. Panels (c)-(f) show that the model supports the increasing 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. 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.*

G Additional Tables

Table 9: Correlation between Rallies and Electoral College Votes

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*

Table 10: In Sample Model Fit

Panel (A): Comparison of Means

	Romney		Obama		Trump		Clinton	
	Model	Data	Model	Data	Model	Data	Model	Data
Southwest	0.145	0.12	0.146	0.17	0.128	0.15	0.0628	0.04
		0.069		0.081		0.076		0.04
Midwest	0.154	0.1	0.146	0.2	0.153	0.15	0.0745	0.08
		0.063		0.088		0.076		0.056
Northeast	0.315	0.31	0.255	0.26	0.387	0.37	0.192	0.21
		0.11		0.099		0.12		0.09
Southeast	0.348	0.43	0.244	0.16	0.446	0.44	0.236	0.24
		0.12		0.079		0.13		0.096

Panel (B): Measures of Fit

	Romney	Obama	Trump	Clinton
Correlation	0.7415	0.7668	0.6947	0.8378
Mean Squared Error	0.3602	0.3296	0.4140	0.2384
Correct Predictions	0.7650	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%.