

Rally the Vote: Electoral Competition with Direct Campaign Communication

Anubhav Jha^{*}

June 2025

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

JEL Codes: C73, D72, P0, M37

^{*}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.

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 (Bugge 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 advertising (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 spend-

¹See 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.

²This 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.

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

ing (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 advances the study of political rallies on four fronts. First, it introduces a finite-horizon dynamic game in which presidential candidates decide when and where to hold rallies while state-level popularity decays geometrically over time. Second, estimating the model with 2012 and 2016 U.S. data shows that a single rally shifts a state’s poll margin by about 0.08 percentage points and that the effect decays at roughly 28% per week. Third, the estimates imply that rallies raised Donald Trump’s Electoral-College win probability by 33% in 2016 and led Obama and Clinton to schedule about a dozen extra rallies in response to their opponents. Fourth, counterfactual simulations reveal that campaign-silence periods of four days or fewer have negligible impact, whereas seemingly non-binding caps on rally counts still curb campaigning because candidates optimally reserve rallies for future contingencies.

The central element underlying these contributions is a finite-horizon dynamic game of perfect information. 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 aims to remain popular in as many states as possible on election day, earning a payoff proportional to each state’s electoral college votes. Rallies boost

⁴For 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.

local popularity, which evolves as an AR(1) process to capture the decay of campaign effects. The autocorrelation parameter, referred to as persistence in popularity, governs the decay of campaign effects and captures the urgency with which candidates increase their campaign activity over time.⁵ Rallies are costly, and these costs vary across candidates and states. Candidates are assumed to move in a stochastic order, with equal probabilities of being first or second movers, eliminating any ex-ante advantage. This imposes a perfect information structure on the game, enabling a unique solution for rally decision probabilities via backward induction.

One key element of the model is that candidates can schedule rallies immediately, rather than committing to them in advance. This distinction hinges on whether campaign schedules allow last-minute revisions or are rigid. I provide two pieces of supporting evidence for the assumption that candidates retain such flexibility. First, I document a non-exhaustive list of last-minute cancellations, additions, and rescheduling of rallies, suggesting that rally schedules are not binding. Second, a model selection test compares the main model to an alternative in which candidates make rally decisions one week in advance. While the parameter estimates are similar, the main model performs significantly better in 2012 and is preferred overall when combining results from both years.⁶ These findings demonstrate that assuming rally decisions are made shortly before they occur is both credible and does not produce misleading conclusions.

The model sharpens our understanding of how rally effectiveness can be identified in the data. 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 combine two data sources. Rally dates and locations

⁵See [Acharya et al. \(2022\)](#) for a similar pattern in campaign spending.

⁶For 2016, the difference in average log-likelihood between the two models is not statistically significant.

come from the candidate calendars compiled by [Appleman \(2012, 2016\)](#), and daily state-level poll margins are taken from FiveThirtyEight’s daily polling averages. These data reveal three systematic patterns that shape the model. First, the average number of events per day rises sharply as election day draws near. Second, rallies are *poll-contingent*: candidates disproportionately visit states where the margin is close or they are trailing—that is, where competition is “neck and neck.” Third, conditional on competitiveness, rallies concentrate in high-electoral-vote states; large but safe states draw few events.

I estimate the model using a likelihood approach that leverages the Markov structure implied by equilibrium behavior. To ensure computational tractability, I estimate the model by grouping 12 swing states into four groups. Using Monte Carlo experiments, I show that excluding stronghold states introduces negligible bias, and that grouping swing states yields conservative estimates of rally effectiveness. This attenuation arises as rally impacts are averaged over a larger population in which a smaller share is directly affected, leading to smaller but directionally consistent counterfactual effects and reducing the risk of overstating rally effectiveness. I also test robustness across three alternative groupings.

The resulting estimates reveal sizable and consistent rally effects across candidates and years. Trump’s rallies lifted his state poll margin by 0.084 pp, while Clinton’s added 0.075 pp; in 2012, Romney gained 0.10 pp and Obama 0.04 pp.⁷ In persuasion terms, a single rally dwarfs a TV advertisement: a Trump rally persuades 0.168% of persuadable voters versus 0.01% for a Republican television spot,⁸ so roughly 17 ads would be needed to match one *MAGA* event.

The estimated persistence parameter implies a weekly decay rate of about 28%, which lies between the estimates of [Gerber et al. \(2011\)](#) and [Acharya et al. \(2022\)](#). This is broadly consistent with the literature and unlikely to reflect artifacts from polling frequency or FiveThirtyEight’s weighting scheme. To further validate this, I re-estimate the model using data from the final 50 days of the campaign—when polling frequency is more than three times higher than in the preceding 50 days—and also using raw poll averages. In both cases, I find no significant change in the estimated decay rate.

Given the sizable persuasive effect of a single rally, I turn to counterfactual experiments

⁷Poll margins are measured in percentage points (pp).

⁸Persuasion rates denote the share of persuadable voters converted; ad figure from [Spenkuch and Toniatti \(2018\)](#).

that quantify rallies’ overall electoral impact and strategic use. The first estimates the cumulative effect of rallies, noting that while candidates hold multiple rallies per state, each decays over time—creating a tension between frequency and diminishing effects. Comparing scenarios with (i) no rallies and (ii) only one candidate rallies, I find that Trump’s rallies raised his win probability by 33%, while other candidates saw little change. The second experiment isolates strategic behavior by comparing scenarios where (i) both candidates rally—strategic interaction is present—and (ii) only one does—no interaction. I find that Obama and Clinton strategically responded to their opponents by holding 12 and 13 more rallies, respectively; Romney showed no response, while Trump held 3.5 fewer.

Many countries enforce campaign silence periods of varying lengths—e.g., one day in France (Pickles, 1960) and two or more days in Cyprus, Indonesia, and Brazil (Knews, 2022; IFES, 2012; Globo, 2020). However, it remains unclear whether such laws effectively reduce the influence of campaigning on election outcomes. I find that short silences (≤ 4 days) have little impact, as they do not provide sufficient time for rally effects to decay. I also examine spending limits by capping the total number of rallies, assuming equal monetary costs across candidates and years.⁹ Even limits that exceed the observed number of rallies (e.g., 110) can reduce overall campaigning and significantly affect outcomes.¹⁰ This is because candidates optimally reserve rallies for potential future contingencies, a behavior that does not arise in static models.

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 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

⁹Cost parameters in the model reflect the opportunity cost of holding a rally.

¹⁰“Seemingly nonbinding limits” refer to caps that are higher than the number of rallies actually held.

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,¹¹ 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 concludes.

¹¹The assumption of SUTVA is more challenging to justify at this geographic level due to spontaneous news coverage of rallies in geographically closer media markets.

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.

A key feature of the model is the joint treatment of both candidates' rally decisions. Rally choices are inherently interdependent: each candidate's action reflects not only their own popularity and costs but also their opponent's current behavior and anticipated response. Ignoring these interactions would introduce endogeneity, as opponent actions are correlated with a candidate's decision to rally. Structural models of political competition—particularly those studying dynamic campaign strategies and advertising—routinely incorporate such strategic dependencies to ensure credible estimation of behavioral primitives.

2.1 Preliminary

The set of states is given by $K = \{1, 2, \dots, K\}$. Let $i \in \{R, D\}$ and $t \in \{1, 2, \dots, T\}$ 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):

$$p_{k,t+1} = \alpha_R a_{Rk,t} + \alpha_D a_{Dk,t} + \rho p_{kt} + \delta_k + v_{k,t+1}, \quad (2.1)$$

where α_i is i 's effectiveness in influencing popularity, the parameter ρ is the persistence in popularity, $v_{k,t}$ is a popularity shock, and δ_k is a state-specific drift.¹² I assume the following.

¹²Readers may wonder whether OLS applied to Equation (2.1) could consistently estimate rally effects. This concern is addressed in Section 4.2. Briefly, the structural model operates at a higher frequency than the ob-

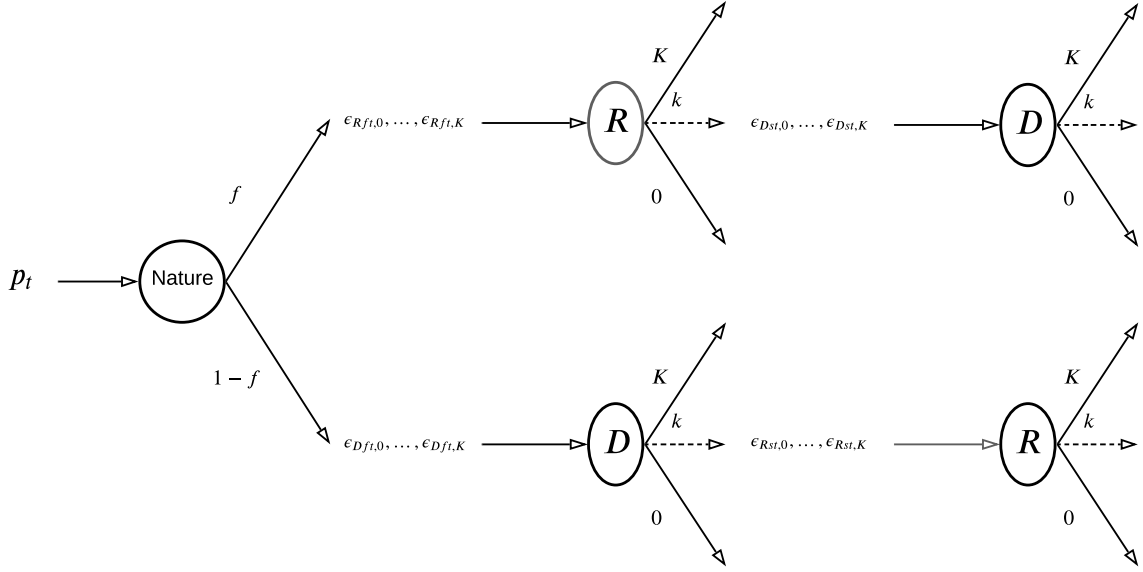


Figure 1: Stage Game

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 $t + 1$.

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

$$(v_{1,t}, v_{2,t}, \dots, v_{K,t}) \sim N(0, \sigma_v^2 I_K). \quad (2.2)$$

Where $\sigma_v^2 I_K$ is a positive definite matrix and σ_v is the volatility in popularity.¹³ Let the density of popularity in period $t + 1$ given period t primitives be denoted by $f(p_{t+1} | a_{R,t}, a_{D,t}, p_t)$.

served polling data, which induces correlation between daily rally counts and unobserved within-day popularity shocks. OLS is inconsistent—even though rally choices in each sub-period are made before shocks are realized—because at the daily level, the total number of rallies is mechanically correlated with the sum of unobserved shocks over the day. As a result, OLS regressors correlate with the composite error term. In contrast, the likelihood approach accommodates this structure by integrating over latent sub-daily popularity states using a Kalman-filter-based setup. See Section 4.2, Section E.1, and Table A4 for details.

¹³In 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.

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).¹⁴

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:

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

2.2 Timing of Decisions and Information

Although campaigns do publish week-ahead rally calendars, *those calendars are not binding contracts*. Table A2 documents numerous cases—across parties and offices—where rallies were cancelled, moved, or added within a 24-hours' notice in response to protests (popularity shocks), controversy (popularity shocks), weather (cost shocks), security threats (cost shocks), or health shocks (cost shocks). Hence candidates retain *substantial real-time discretion* over the location and timing of events.

To capture that institutional reality, I treat rally location choices as a *game without commitment*: at the start of each period, candidates observe the current state of the race and then decide where (or whether) to rally in that period. Candidates in this model do not commit to future rallies, as these can be revised in the next period if circumstances change. Modeling rally decisions as being scheduled in advance would artificially constrain behavior and contradict the evidence provided in Table A2. In Appendix D.1, I discuss a version of the model in which candidates schedule rallies one week in advance. The estimates from that model are also reported in Section 6, and they show that the results do not change significantly. Moreover, Vuong's model selection test rejects the scheduling model in favor of the one presented in this section.

Stage-game structure. In each period $t \in \{1, 2, \dots, T\}$ the following six sub-periods, illustrated in Figure 1, unfold:

¹⁴The 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.

- τ_1 The popularity-level vector $p_t = (p_{1t}, p_{2t}, \dots, p_{Kt})$ 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, $\epsilon_{if,t} = (\epsilon_{if,t,0}, \epsilon_{if,t,1}, \dots, \epsilon_{if,t,K})$, are realized. These cost shocks capture the effect of unforeseen events on rally decisions.¹⁵
- τ_4 The first mover, i , makes a rally decision by solving the following bellman equation:

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

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, $\epsilon_{js,t} = (\epsilon_{js,t,0}, \epsilon_{js,t,1}, \dots, \epsilon_{js,t,K})$, 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:

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

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.

¹⁵The 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.

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

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

With probability f_i , i is the first mover and the term $\mathbb{E}_{\epsilon_{ift}} \left(V_{ift}(p_t, \epsilon_{ift}) \right)$ 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.¹⁶ Moreover, we can also show that best responses, in probability space, are functions of current popularity, cost shocks, 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*

¹⁶The 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 2.3, the following equations hold for all $t = 1, 2, \dots, T$

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

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

$$\sigma_{is,t}(k; l, p_t) = P(k = a_{is,t}^*(a_{jf,t} = l, p_t, \epsilon_{is,t})) = \frac{\exp(u_{is,t}(k; l, p_t) - u_{is,t}(0; l, p_t))}{1 + \sum_{q=1}^K \exp(u_{is,t}(q; l, p_t) - u_{is,t}(0; l, p_t))}, \quad (2.9)$$

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.¹⁷

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 functions¹⁸ makes applying monotone comparative static results infeasible. Moreover, I analyze a candidate i 's probability of choosing k before nature chooses the order of play in a given period, call this quantity $\sigma_{it}(k; p_t)$.¹⁹

¹⁷Or deterministic part of period t payoffs.

¹⁸Such 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.

¹⁹To calculate this, first I calculate the joint probability of candidates i and j choosing k and l respectively, call this $\sigma_t(k, l; p_t) = \frac{\sigma_{if,t}(k; p_t) \sigma_{js,t}(l; k, p_t) + \sigma_{jf,t}(l; p_t) \sigma_{is,t}(k; l, p_t)}{2}$. Then $\sigma_{it}(k; p_t) = \sum_l \sigma_t(k, l; p_t)$. I follow Judd et al. (2014), Judd et al. (2017), and Heiss and Winschel (2008) for simulating $\sigma_{is,t}$ and $\sigma_{if,t}$ for $i \in \{R, D\}$. Online Appendix F provides the details.

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 periods 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.

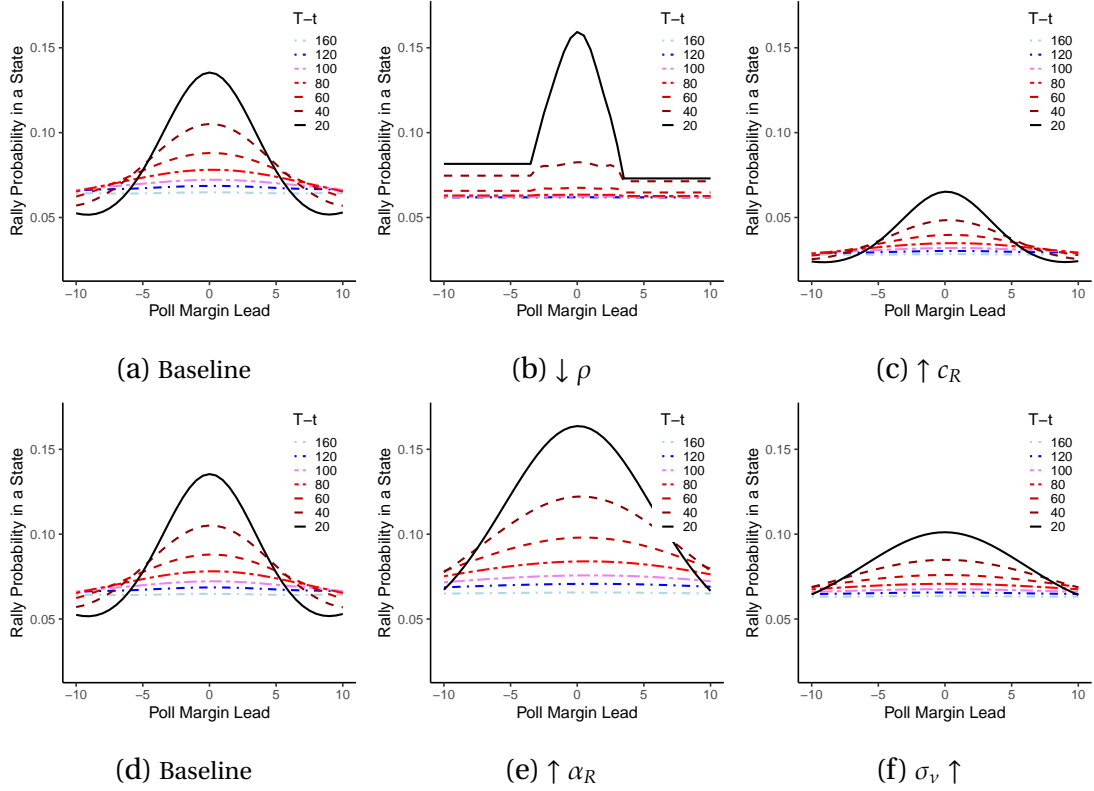


Figure 2: This figure illustrates how the probability of rallying in a state changes when the key parameters ρ , α_i , c_i , and σ_v change. Here $T - t$ denotes the periods before the election. For the baseline I fix parameter values $\alpha_R = -\alpha_D = 0.07$, $\rho = 0.99$, $c_R = c_D = 2.5$, and $\sigma_v = 0.2$. 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.

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 were obtained from Democracy in Action, a website created by Eric M. Appleman. For the 2012 and 2016 presidential elections, I used Appleman (2012) and Appleman (2016), respectively. The website provides daily records of campaign activities undertaken by each candidate, including rallies, fundraisers, interviews, church visits, and other appearances (see Table A1 for the classification). To isolate political rallies, I retained only those entries whose description field mentions the terms “rally,” “speech,” or “special event.” Other activities—such as fundraisers, interviews, or church visits—were excluded because they do not consistently involve persuasive public communication. Fundraisers are typically private events aimed at resource mobilization rather than voter persuasion. Church visits are heterogeneous: some involve silent attendance or private meetings, others include short remarks, and many are not intended as public addresses. In instances where a church visit involved a speech, the entry was labeled accordingly based on the presence of “speech” in the description. “Stops” refer to impromptu visits where candidates might greet crowds, dine, or briefly interact with the public, but rarely involve formal persuasion. Arbitrarily including such varied activities would introduce measurement error and undermine the identification of persuasive effects. Rallies, by contrast, are deliberate, public-facing, and consistently structured to persuade voters—making them the most appropriate unit for modeling dynamic campaign behavior.

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

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.²⁰

While FiveThirtyEight’s polling averages provide a high-quality and widely used data source, they could potentially influence model estimates—particularly the persistence parameter—due to polling inaccuracies, the infrequent release of underlying polls, or the modeling procedure itself. To address these concerns, I conduct a series of robustness checks in Section 6 that test the sensitivity of my results to each of these issues. Across all exercises, I find that the key parameter estimates remain stable, supporting the reliability of the main results.

The aggregate number of activities in the raw data obtained after classifying the set of all activities is provided in Table A1.²¹ The table also shows how many rallies were removed specifically after the cleaning process for each candidate. Detailed summary statistics for the swing states are provided in Table 1. Strong-hold states are excluded for reasons detailed in Section 4.3. The swing states provide a sufficient cross-sectional variation in a Republican candidate’s poll margin lead, ranging from −5.26 percentage points to 8.59 percentage points in 2012 and from −7.1 percentage points to 14.3 percentage points. However, within-state 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

²⁰The FiveThirtyEight polling series is constructed using a Bayesian model with data-driven weights, which could in principle raise concerns about mechanical smoothing or artificially inflated persistence. However, the weekly decay rate implied by my model—approximately 28%—falls between the slower-moving dynamics reported in Acharya et al. (2022) (10–15% weekly decay) and the more rapidly fading experimental effects in Gerber et al. (2011), where influence typically dissipates within 1–2 days. These comparisons suggest that the results are consistent with external estimates and not an artefact of the specific poll aggregation method. For details on the poll aggregation method used visit <https://fivethirtyeight.com/features/a-users-guide-to-fivethirtyeights-2016-general-election-forecast/>

²¹These numbers correspond to last 125 days. For the analysis I focus on last 100 days.

Table 1: Summary Statistics

State	2012 Election				2016 Election				
	Poll Lead		Rallies		Poll Lead		Rallies		EC Votes
	Mean	Std. Dev.	Obama	Romney	Mean	Std. Dev.	Clinton	Trump	
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.

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 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

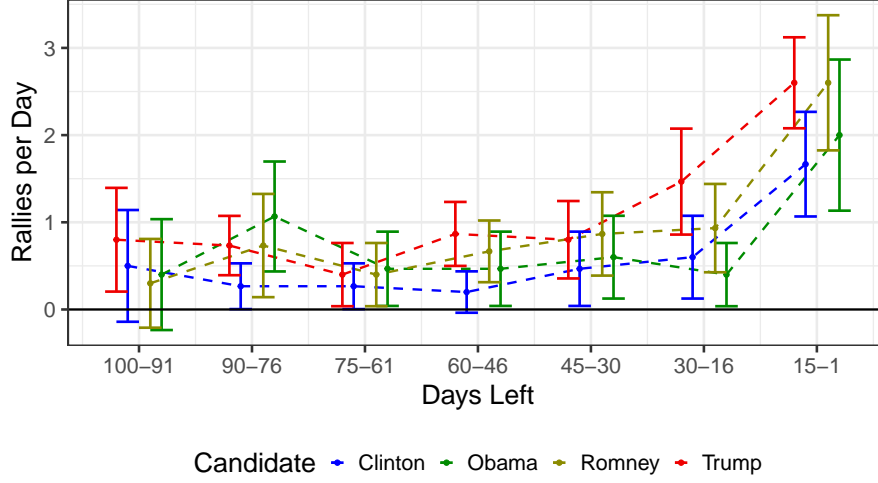


Figure 3: This figure shows average rallies per day for 15-day bins (and a 10-day bin for 100–91 days before the election). For each of these bins the corresponding 95% confidence interval for average rallies per day is also provided.

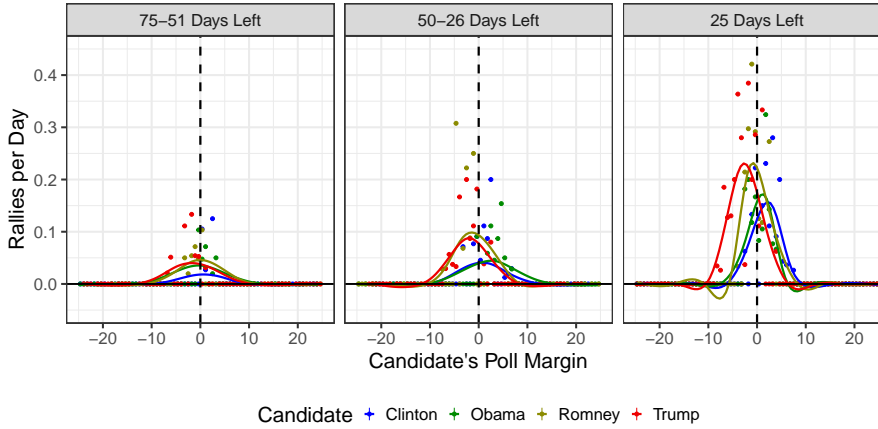


Figure 4: This figure shows a bin scatter of a candidate's number of rallies in a day and poll margin lead along with a generalized additive model fit line.

my model can support due to candidate-specific cost and rally effectiveness.²²

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,

²²In 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.

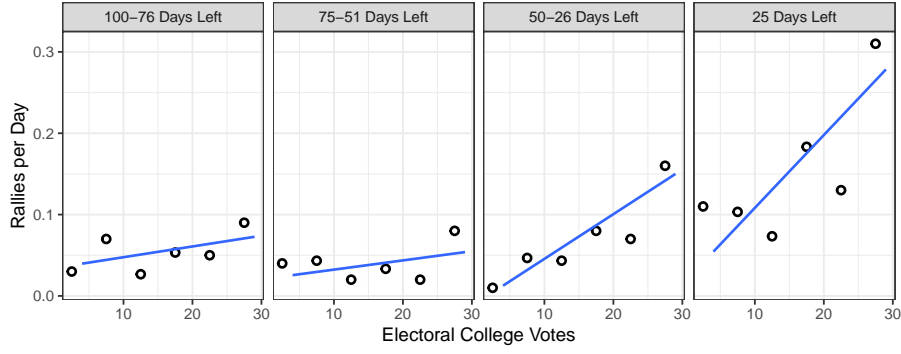


Figure 5: This figure shows a bin scatter of a candidate's number of rallies and electoral college votes at the state \times day level for swing states.

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 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.²³ Within this set of states, candidates prioritize states with higher electoral college votes over those with lower electoral college votes. From the Table A8 and Figure 5 show that for all candidates, the correlation between rallies and electoral college votes increases as the election day approaches.²⁴

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 $c_R, c_D, c_1, \dots, c_K$. Controlling for all state fixed effects along with the candidate specific cost of rallying leads to a multicollinearity problem in this setting. For any

²³These 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 \(2020\)](#).

²⁴The 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.

given set of values $c_R, c_D, c_1, \dots, c_K$, I can reparameterize the costs by setting $\tilde{c}_R = c_R + c_K$, $\tilde{c}_D = c_D + c_K$, $\tilde{c}_k = c_k - c_K$ for $k = 1, \dots, K-1$, and $\tilde{c}_K = 0$, without changing the effective cost of rallying in any state, which is given by $\tilde{c}_i + \tilde{c}_k = c_i + c_k$ for $i \in \{R, D\}$ and $k \in \{1, 2, \dots, K\}$. This transformation leaves the total cost of rallying in a state k unchanged. 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 $\alpha_R, \alpha_D, \rho, \sigma_v, \delta_1, \dots, \delta_K$. 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_v . 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). To identify f , one would need observations on who moved first, which is unavailable. I calibrate f to a value of 0.5, as this value eliminates any ex-ante first mover or second mover advantage in the game. I also check robustness of the estimates if f is set to 0.33 or 0.67 and do not find significant changes in the estimates. Finally, I calibrate E to 538, which is the total number of electoral college votes. As a result, payoffs are measured in units of electoral college votes.

4.2 Likelihood

Campaigns can—and frequently do—schedule several rallies on the same calendar day, while reliable polling data are available only at the daily frequency. I therefore model popularity as a latent AR(1) process evolving at the *quarter-day* level (four sub-periods per day), with the FiveThirtyEight state margin serving as a single daily observation of that latent state.

An alternative approach is to assume that candidates can allocate at most four rallies across K states in each period, where a period corresponds to a day. This allows for multiple rallies in the same state or no rallies at all in some (or all) states. The choice space in this setting contains $\binom{K+r}{r}$ potential alternatives, where r is the maximum number of rallies and

K is the number of states. With $K = 4$ and $r = 4$, this yields 70 combinations. For the first mover, this implies 70 option-specific value functions to solve for.²⁵ For the second mover, we have 4,900 option-specific value functions, since there are 70 options for each possible choice of the first mover. Since there are two candidates, we must track 9,940 value functions per period. Given the absence of stationarity, the total number of value functions over time approaches *one million* (100 daily observations). While modeling the process at the quarter-day level may initially appear capricious and confusing, this approach results in substantial dimensionality reduction. For a given value of K and r , we have a total of $2 \cdot (K + 1 + (K + 1)^2) \cdot r \cdot 100$ option-specific value functions. For $K = r = 4$, we get only 24,000 option-specific value functions as opposed to a million.

More importantly, because an AR(1) process is closed under temporal aggregation (see [Hamilton 1994](#), Ch. 6), the higher-frequency specification is *statistically equivalent* to an AR(1) process defined directly at the daily level. The persistence parameters are linked by a deterministic mapping, and no identification assumptions are violated. The likelihood constructed in this section integrates over the three unobserved within-day states, exactly as in a standard state–space (Kalman filter) framework.

The estimated persistence, $\hat{\rho} \approx 0.99$ (equivalent to a daily persistence of 0.95), lies well within the range reported by [Hill et al. \(2013\)](#), [Gerber et al. \(2011\)](#), and [Acharya et al. \(2022\)](#), confirming that the sub-daily treatment does *not* spuriously inflate serial correlation. In weekly terms, the implied decay rate of approximately 28% falls between the highly persistent estimates in [Acharya et al. \(2022\)](#), who report typical weekly decay rates of 10–15%, and the more rapidly fading experimental effects in [Gerber et al. \(2011\)](#). I also analyze the sensitivity of the estimated parameter to polling accuracy, FiveThirtyEight’s modeling procedure, and the infrequency of polls in Section 6, and find no substantial deviations.

Given the latent structure governing the evolution of popularity over time, I estimate the model by maximizing the full-information likelihood, solving the dynamic game at each candidate parameter value. 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

²⁵Each option corresponds to a specific allocation of rallies across the states, such that the total number of rallies does not exceed 4. Formally, an allocation is given by $n_{ift} = (n_{ift,1}, n_{ift,2}, \dots, n_{ift,K})$, where $n_{iftk} \geq 0$ for all k , and $\sum_k n_{iftk} \leq 4$. The expected payoff that the first mover receives from each possible rally allocation, denoted by $u_{ift,n_{ift}}(p_t)$, depends on the current popularity across all K states. This defines an option-specific value function that must be approximated using a sparse polynomial.

(Rust, 1987).²⁶ 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 $t = 1, 2, \dots, T$. Let P_t denote the realized poll margins in a given period t , for $t = 1, 2, \dots, T + 1$. 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 $t = 1, 2, \dots, T$. 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, $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_T$, 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 $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_T$ obey the Markov property. Moreover, the transition density $\psi_t(X_t|X_{t-1})$ for $t \geq 1$ is given by:*

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

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 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,

²⁶There 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).

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 $P_{4(d-1)+l} = (P_{4(d-1)+l,1}, P_{4(d-1)+l,2}, \dots, P_{4(d-1)+l,K})$ where $P_{4(d-1)+l,k} \in \mathbb{R}$. For day d observation I consider $d + 1$ observed popularity. Hence for day d the observation is given by $X_d = \{A_{4d-3}, A_{4d-2}, A_{4d-1}, A_{4d}, P_{4d+1}\}$.

Proposition 4.1 shows that the random vectors $\{X_1, X_2, \dots, X_D\}$ 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 $\{X_1, X_2, \dots, X_D\}$ obeys the Markov property and its governed by the transition density $\lambda_d^\theta(X_d|X_{d-1})$, which is given by:*

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

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, \dots, X_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 $\ell\ell(\theta; X_1, X_2, \dots, X_D) = \sum_{d=1}^D \log(\lambda_d^\theta(X_d|X_{d-1}))$. The integration in Proposition 4.1 is not feasible analytically and therefore I rely on Monte Carlo methods which are discussed in Appendix C.

4.3 State Groups

Estimating the model for all U.S. states is infeasible, as it introduces 50 state variables into the dynamic programming problem that each candidate must solve. Even when considering only the swing states, the resulting dimensionality reduction is insufficient for reliable estimation. To address this, I construct groups of swing states based on their U.S. regions, allowing the model to be estimated with sufficient accuracy. I also examine the robustness of the estimates across three alternative definitions of state groups, discussed in Section 6. Additionally, I also analyze the implications of grouping state on parameter estimates and counterfactual results, reported in Tables A5, A6, and A7.

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 (0.68)	0.17 (0.81)	0.15 (0.76)	0.04 (0.40)	3.93 (0.56)	2.27 (0.96)	26
Midwest	0.10 (0.62)	0.20 (0.87)	0.15 (0.76)	0.08 (0.56)	-2.73 (1.49)	-2.08 (1.36)	32
Northeast	0.31 (1.07)	0.26 (0.99)	0.37 (1.16)	0.22 (0.91)	-2.72 (1.07)	-0.39 (1.28)	42
Southwest	0.43 (1.24)	0.16 (0.78)	0.44 (1.25)	0.24 (0.95)	1.52 (1.11)	0.21 (0.96)	57

**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.

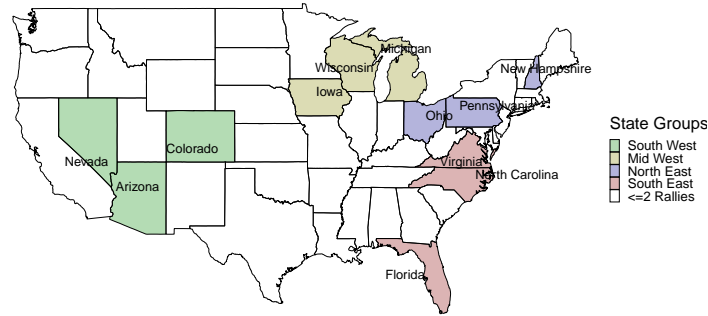


Figure 6: The figure displays states considered in the analysis. The unlabeled states, which are omitted, had less than 2 rallies by each candidates. The states considered in the analysis are all swing states. To keep the dimensionality of the model tractable I group these swing states into 4 groups.

Removing stronghold states from the analysis is not problematic. I demonstrate this through Monte Carlo experiments in Table A5. In the Monte Carlo experiments, the data are generated from a four-state model and then estimated using a two-state specification (by dropping the stronghold states). As the table shows, the mean parameter estimates are consistently close to the true values across all parameters, regardless of the electoral college distribution or state-specific drift configurations. Excluding stronghold states does not materially alter the estimates. From the candidate's perspective, the net benefit of rallying in a stronghold state—where the drift term δ_k is large relative to the effectiveness parameter α_i —is close to zero, since rallies are not sufficiently persuasive to overcome large

pre-existing support. From the popularity side, stronghold states typically receive very few rallies, and thus lack the within-state variation needed to identify the effect of rallies on popularity. Moreover, this practice of removing stronghold states from analysis is common in the literature studying persuasive effects of campaigning, including [Spenkuch and Toniatti \(2018\)](#); [Snyder and Yousaf \(2020\)](#).

Grouping swing states into clusters is motivated by both computational tractability and interpretability. Solving the model at the individual state level is infeasible due to the high dimensionality of the state space. To evaluate the implications of this grouping, I conduct Monte Carlo simulations (see Tables [A6](#) and [A7](#)) using a data-generating process with four states: two large and two small. The drift parameters are selected to ensure heterogeneity in the competitiveness of the election across states. The model is then estimated using grouped data, in which popularity is averaged across states within each group. I consider multiple grouping methods to assess the sensitivity of the estimates to how the groups are constructed.²⁷

In the grouped specification, the estimated group-level effectiveness parameters are smaller in magnitude than the state-level effectiveness values. This is natural as the estimated effects in the grouped version correspond to the impact of rallies on a larger population unit—i.e., a group of states—rather than on individual states. The attenuation arises because averaging across heterogeneous states dampens the within-group variation relevant for identifying the impact of rallies. Grouping therefore operates conservatively—recovering the effect of rallies at the level of pooled populations rather than exaggerating individual state responsiveness. This conservativeness carries through to counterfactual analyses: the simulated counterfactuals remain directionally consistent and do not overstate the total effect of rallies, further supporting the robustness of the grouped-state approach (see Panel (D) in Tables [A6](#) and [A7](#)).

I create four groups of states, where each group is the intersection of a U.S. region and

²⁷For instance, pairing one large and one small state. In this grouping, the first group combines an R-leaning and a slightly D-leaning state, whereas the other group combines a D-leaning and a slightly R-leaning state. Another grouping structure may combine large states together and small states together, where both large states are R-leaning and both small states are D-leaning. I also examine additional grouping combinations in Tables [A6](#) and [A7](#). Furthermore, by varying the overall drift magnitude—such as larger drifts in Table [A6](#) and smaller drifts in Table [A7](#)—I assess how state-level competitiveness influences the estimates.

the swing states²⁸ 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. Note that grouping also absorbs spillovers or interdependence across nearby states, which are otherwise omitted from the model.

I also consider three alternative definitions of state groups for robustness tests. One groups only those states that are studied in [Snyder and Yousaf \(2020\)](#), another 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. Finally, the third one treats Ohio as part of Midwest state group and not part of Northeast state group. I do not find substantial deviation of estimates from the preferred specification. The results from this exercise are discussed in Section 6.

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 percentage points (s.e. = 0.0155), while a Clinton rally increased her lead over Trump by 0.0745 percentage points (s.e. = 0.0152). For 2012, I find a Romney rally increased his lead over Obama by 0.100 percentage points (s.e. = 0.0227) while an Obama rally increased his

²⁸Here, the swing states, for the lack of a better name, are defined as states with at least two rallies in either election.

lead over Romney by 0.042 percentage points (s.e. = 0.0135). These estimates are statistically significant at the 1% level of significance. The net effect of both candidates rallying in a state for 2016 is a gain of 0.0094 percentage points (s.e. = 0.012) in favor of Trump, which is statistically insignificant, while for 2012 the net effect is a gain of 0.058 percentage points (s.e. = 0.014) in favor of Romney which is statistically significant.

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 0.990 (s.e. = 0.002) for 2012 and 0.991 (s.e. = 0.001) for 2016. The implied weekly decay rate is 28%.²⁹ This decay rate lies in the right tail of the perceived decay rate distribution estimated in Acharya et al. (2022), but it is lower than the estimate reported in Hill et al. (2013). The third-degree lagged polynomial specification considered in Gerber et al. (2011) yields a decay rate of 25%, which is close to the estimate in this paper.

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 in 2012, estimated in Spenkuch and Toniatti (2018), are also provided. I adopt the defini-

²⁹To calculate the decay rate, λ , over Δ periods, I use the relation $\rho = e^{-\frac{\lambda}{\Delta}}$. For the weekly decay rate, $\Delta = 7 \times 4$, and therefore $\lambda = -28 \times \log(\rho)$, where \log denotes the natural logarithm.

Table 3: Parameter Estimates

Main Parameters			Fixed Effects		
Parameters	2012	2016	Parameters	2012	2016
<i>Popularity</i>					
α_R	0.1 (0.0227)	0.0839 (0.0155)	δ_1	0.04 (0.011)	0.023 (0.008)
α_D	-0.042 (0.0135)	-0.0745 (0.0152)	δ_2	-0.027 (0.0082)	-0.02 (0.009)
ρ	0.99 (0.002)	0.991 (0.001)	δ_3	-0.027 (0.0082)	-0.0069 (0.0073)
σ	0.147 (0.0143)	0.16 (0.0148)	δ_4	0.00019 (0.0079)	-0.0074 (0.0073)
c_R	2.9 (0.287)	2.36 (0.208)	c_1	0.367 (0.324)	0.943 (0.411)
c_D	2.83 (0.206)	3.26 (0.259)	c_2	0.421 (0.277)	0.788 (0.308)
			c_3	-0.08640 (0.281)	-0.0443 (0.22)
Observations		2012 100	2016 100		
Log Likelihood		-656.26	-654.60		

Standard errors are reported in parentheses.

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.

tion 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 :

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

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 .³⁰ 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 . Then I take average across all states and periods and calculate its expectation

³⁰Note 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_R^{Rally} and vice-versa.

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			-

Standard errors are reported in parentheses.

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 \(2018\)](#).

using simulations.

I find that the persuasion rates of rallies in 2012 and 2016 were higher than those of TV advertising in 2012. For instance, 17 TV ad spots are as persuasive as one Trump rally. These ratios for Romney, Obama, 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.³¹ 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 A9 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 sub-samples, training and validation. I randomly select (without replacement) 20% of the observations for the validation sample. I estimate the model on the remaining 80% of the

³¹However, 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 minimum of correct predictions across all candidates is 76%.

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 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

6.1 Accounting for TV ads

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 Wesleyan

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 ad 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.

6.2 Candidates Schedule Rallies one week in advance

One key concern readers may have is the potential disconnect between how rallies are scheduled in practice and the modeling choice made in this paper, where candidates are assumed to have the flexibility to schedule rallies immediately before they are held. To address this, I construct a version of the model with commitment, where candidates schedule rallies one week (i.e., 28 periods) in advance. The key departures from the baseline specification are detailed in Section D.1. In the commitment model, a candidate's information set includes current popularity across all states as well as the rally schedules already committed to by both themselves and their opponent. Each period, candidates choose where to hold a rally one week later. They formulate a rational forecast of what their popularity will be a week ahead, based on their currently committed schedule, the opponent's schedule, and the popularity shock observed in the current period.

I show that rather than tracking the full schedule of future rallies as the state variable, it is suffice to track week-ahead popularity forecasts, which capture the net effect of all scheduled rallies on expected popularity and substantially reduces the dimension of the state variables. Decision-making in this model ends a week before the election.

The estimates for this model are reported in columns (13) and (14) of Table 6. The parameter estimates do not change significantly relative to the main specification. However, when examining the average log-likelihood, we find that for 2016 it increases insignificantly by 0.047 units (s.e. = 0.07), while for 2012 it decreases significantly by 0.221 units (s.e. =

0.10). The combined change in average log-likelihood is -0.173 units (s.e. = 0.06), which is statistically significant. Therefore, Vuong’s model selection test overall favors the main specification.

Notably, since the parameter estimates remain largely unchanged, the substantial gap in log-likelihood is attributable to model fit: rally decisions are better explained as functions of current popularity shocks rather than as functions of week-lagged popularity. This provides evidence that candidate behavior is better captured by a model allowing greater flexibility in adjusting rally schedules, rather than by one in which schedules are fixed and cannot be revised.

6.3 Alternate state-group definitions

I also consider three alternative state groupings. First, I reclassify North Carolina and Virginia into the Northeastern group. This is entirely plausible, given that both states are geographically closer to the Northeastern group than to Florida. As a result, the Southeastern group includes only Florida. Estimates under this grouping are reported in Columns (3) and (4) of Table 6. For 2016, I do not find significant differences in the rally effectiveness estimates. For 2012, the effectiveness estimates decrease in magnitude but remain significantly greater than zero, and are statistically indistinguishable in standard deviation units of popularity. That is, the effect of a Romney rally, measured in standard deviation units of popularity (α_R/σ_v), is statistically indistinguishable between the main specification ($\alpha_R/\sigma_v = 0.681$, s.e. = 0.13) and this alternative grouping ($\alpha/\sigma_v = 0.44$, s.e. = 0.07). For Obama, the decrease in effectiveness relative to the main specification is statistically insignificant, both in absolute terms and in standard deviation units (p -value = 0.57).

Second, I adopt the swing state classification used by [Snyder and Yousaf \(2020\)](#) in their event study of political rallies. Results are reported in Columns (5) and (6) of Table 6. For 2016, I again find no significant changes in effectiveness estimates relative to the baseline—both in absolute terms and in standard deviation units. For 2012, the estimates decrease in magnitude but remain significantly greater than zero. As before, for Obama, the change in effectiveness is statistically insignificant in both absolute and standardized terms. For Romney, the decrease is statistically significant, though the estimate remain significantly above zero.

Table 6: Robustness Tests

Parameters	Control for T.V. Ads		Alt. State Groups:		Swing States in		Aggregate Shocks		Spatial Autocorrelation		Polling Error		Schedule Rallies	
	Spenkuch and Toniatti (2018)		Isolate Florida		Snyder and Yousaf (2020)		σ_{agg}		ρ_{cov}		State-Wise		a week in advance	
	2012 (1)	2016 (2)	2012 (3)	2016 (4)	2012 (5)	2016 (6)	2012 (7)	2016 (8)	2012 (9)	2016 (10)	2012 (11)	2016 (12)	2012 (13)	2016 (14)
α_R	0.0733 (0.0127)	0.0862 (0.0182)	0.0271 (0.00458)	0.0796 (0.0149)	0.037 (0.00628)	0.0839 (0.0155)	0.0724 (0.0108)	0.0682 (0.0146)	0.0817 (0.0136)	0.065 (0.0136)	0.0913 (0.0224)	0.0745 (0.0265)	0.109 (0.0275)	0.0883 (0.0159)
α_D	-0.0534 (0.00923)	-0.0828 (0.019)	-0.021 (0.00427)	-0.0593 (0.0113)	-0.0322 (0.00639)	-0.0745 (0.0152)	-0.0496 (0.00784)	-0.0582 (0.0127)	-0.00436 (0.00824)	-0.0575 (0.0129)	-0.0569 (0.0242)	-0.0789 (0.0156)	-0.064 (0.0209)	-0.0646 (0.0124)
ρ	0.987 (0.002)	1.01 (0.001)	0.991 (0.001)	0.988 (0.002)	0.987 (0.003)	0.991 (0.001)	0.986 (0.002)	0.989 (0.001)	0.994 (0.001)	0.99 (0.001)	0.994 (0.001)	0.992 (0.002)	0.993 (0.002)	0.991 (0.001)
σ_v	0.157 (0.0133)	0.166 (0.0128)	0.0611 (0.00729)	0.157 (0.0118)	0.0901 (0.0062)	0.16 (0.0148)	0.128 (0.0112)	0.121 (0.00939)	0.143 (0.0132)	0.154 (0.0127)	0.147 (0.0146)	0.16 (0.0149)	0.135 (0.0141)	0.145 (0.013)
c_R	2.71 (0.268)	2.68 (0.241)	3.33 (0.271)	2.69 (0.235)	2.78 (0.291)	2.36 (0.208)	2.74 (0.275)	2.38 (0.198)	2.86 (0.277)	2.38 (0.201)	2.75 (0.275)	2.39 (0.208)	2.73 (0.266)	2.42 (0.218)
c_D	2.83 (0.186)	3.55 (0.305)	3.44 (0.218)	3.49 (0.292)	2.92 (0.22)	3.26 (0.259)	2.85 (0.197)	3.19 (0.257)	2.63 (0.195)	3.21 (0.255)	2.84 (0.193)	3.33 (0.291)	2.82 (0.199)	3.17 (0.26)
ρ_{corr}	- -	- -	- -	- -	- -	- -	- -	- -	0.005 (0.001)	0.01 (0.002)	- -	- -	- -	- -
σ_{agg}	- -	- -	- -	- -	- -	- -	0.069 (0.0144)	0.103 (0.0138)	- -	- -	- -	- -	- -	- -
Fixed Effects:														
Cost	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Poll Margins	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
$\bar{L}\bar{L}$	-6.9405	-6.7907	-2.8564	-6.3453	-4.6854	-6.546	-6.424	-618.58	-6.5334	-6.2152	-6.7018	-6.5844	-6.7235	-6.447
Observations	100	100	100	100	100	100	100	100	100	100	100	100	93	93

Standard errors are reported in parentheses.

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 allows for 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. Columns (13) and (14) report estimates for a model where candidates schedule rallies one-week in advance, their information set consists of opponent schedules and poll-margins realized at the day of scheduling. Here the standard errors have been computed by using observation wise gradient and likelihood hessian. I use HAC estimation for this purpose.

Finally, I test a third grouping in which Ohio is included in the Midwest group (see Columns (9) and (10) of Table A3). Under this specification, I do not find significant differences in the standardized effectiveness estimates for either Romney (difference = -0.20 , s.e. = 0.15) or Obama (difference = -0.02 , s.e. = 0.10). For 2016, I do not find significant differences in the rally effectiveness estimates, both in absolute terms and in standard deviation units.

6.4 Polling

I conduct three robustness checks to assess the reliability of my estimates. First, I account for ex-post polling errors by comparing each state's election-day poll margin from FiveThirtyEight with the actual vote margin, and adjusting daily poll data accordingly. Estimating the model using these corrected margins yields no significant changes (see columns (11) and (12) of Table 6). Second, I address concerns about the infrequent release of individual polls, which may bias the persistence parameter upward. Note that FiveThirtyEight incorporates national polls into their averages, with at least one new national poll released each day during the final 100 days of the campaign. While state-level polls within the 12 swing states are less frequent, their frequency increases sharply closer to the election: the number of state polls more than triples in the last 50 days (77.5 polls per day) compared to the preceding 50 days (25 polls per day). Re-estimating the model using data from this later period again produces stable estimates. Third, I test whether FiveThirtyEight's weighting procedure—based on pollster ratings, sample size, and ideological lean—affects the results. Using raw state-level polls from 2016,³² I construct a daily series by averaging all polls based on the dates when respondents were surveyed. Since at least one poll is in the field each day for every state group, constructing a balanced panel of poll averages is feasible. This alternative specification also yields parameter estimates that remain largely unchanged (see column (11) in Table A3).

6.5 Correlated popularity shocks

I also test if estimates change significantly 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

³²Comparable raw polling data for 2012, including field dates, is limited.

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 $\Omega_{ii} = \sigma_v^2$ 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 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 (0.010123)	0.000359 (0.040219)	0.000477 (0.0207)
Baseline Model vs Max Win Prob	0.0545 (0.0613)	0.0932 (0.0875)	0.0739 (0.0533)
Observations	100	100	200

Standard errors are reported in parentheses.

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.

6.6 Model selection tests

I test three competing models: (a) candidates are myopic and unable to perform backward induction perfectly; (b) candidates are non-strategic, neither observing nor anticipating their opponents' actions; and (c) candidates seek to maximize their probability of winning rather than the total number of electoral college votes.

One key assumption I make is that candidates can execute backward induction flawlessly.

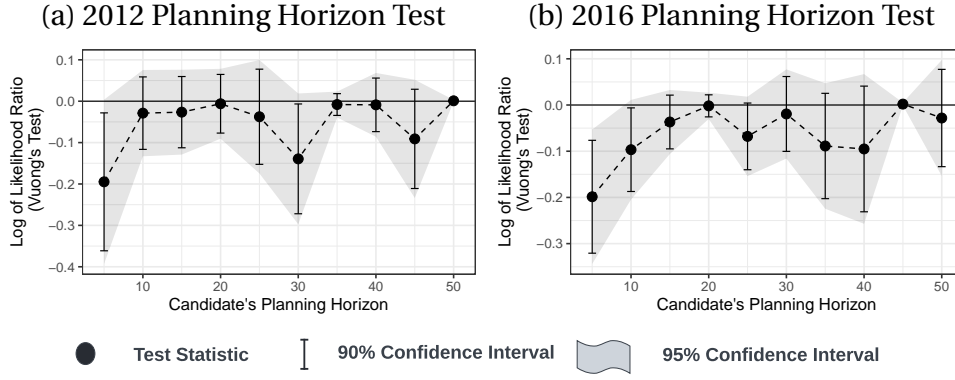


Figure 7: This figure shows results from conducting Vuong's closeness test where the "Full Planning Horizon" model (baseline model) is compared with models that have shorter planning horizons. A negative statistic indicates that the baseline model performs better than the competing models. The x-axis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

This might be untrue if candidates are myopic.³³ The second assumption is that candidates make rally decisions 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.³⁴

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

³³In this context (i) myopic candidates, (ii) a voter base that is more attentive closer to election, or (iii) a voter 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.

³⁴While 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.

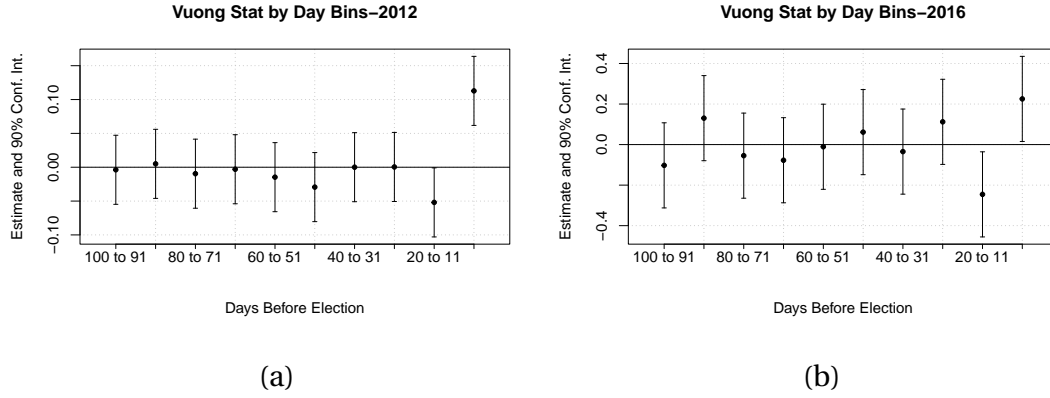


Figure 8: *Vuong Statistic by Day Bins. The figure shows that candidates are strategic especially 10 days prior to the election, while being non-strategic between 20 and 10 days before the election.*

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.

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. 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. 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 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

³⁵To 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 $\{\Delta y^m : \Delta y^m = y_{i \text{ Rallies}}^m - y_{\text{None rally}}^m\}_{m=1}^M$.

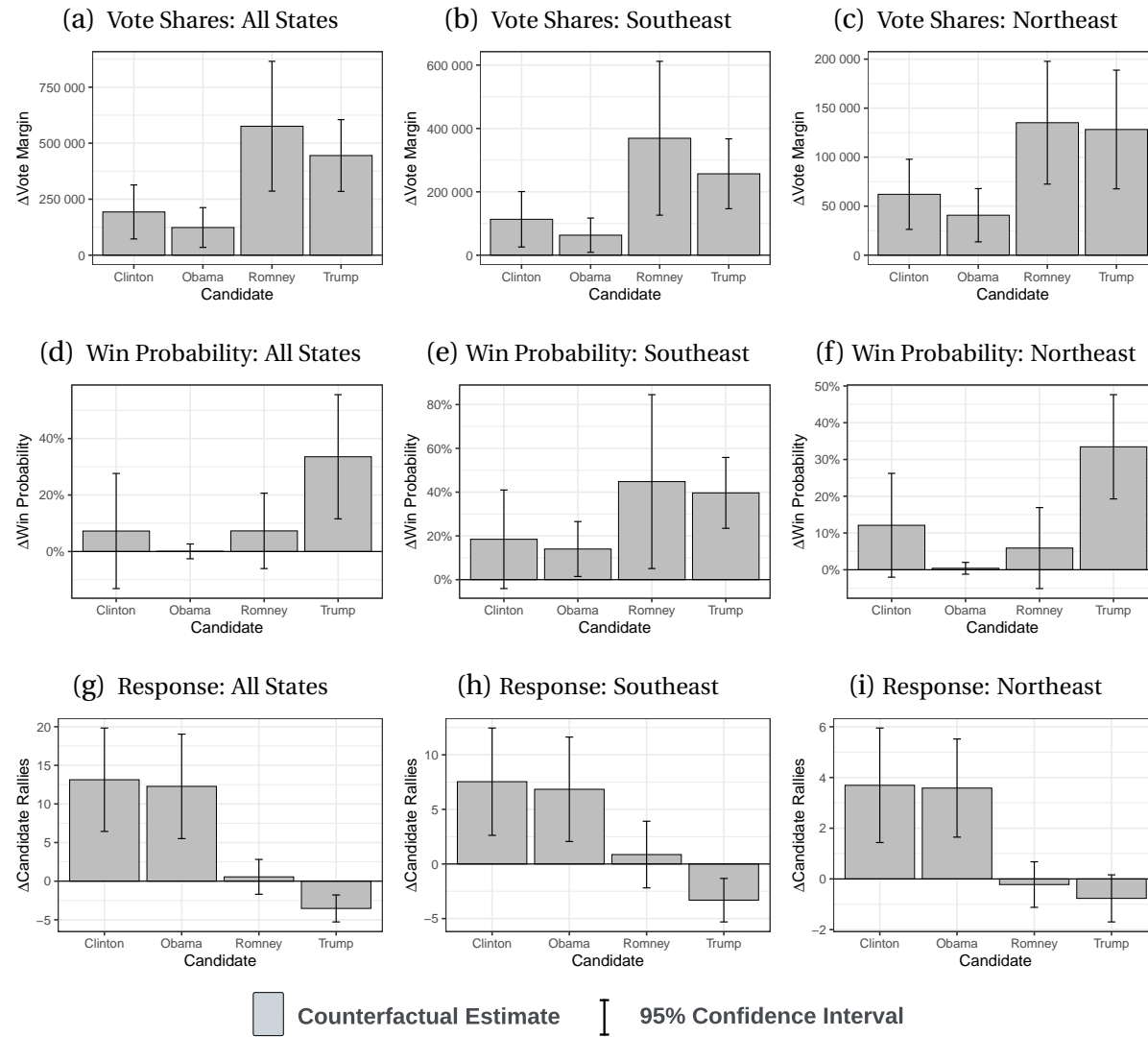


Figure 9: This figure shows the cumulative effect of a candidate's rallies on their vote margin lead in Panels (a)–(c) and their probability of winning in Panels (d)–(f). Estimates of number of rallies held or not held as a campaign response to the opponent's rallies are shown in Panels (g)–(i).

and Clinton conducted an additional 12.3 (s.e. = 3.45) and 13.14 (s.e. = 3.42) rallies, respectively, in response to rallies held by Romney and Trump. Conversely, Trump held 3.5 (s.e. = 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 outcomes.

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. 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,

³⁶For 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.

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

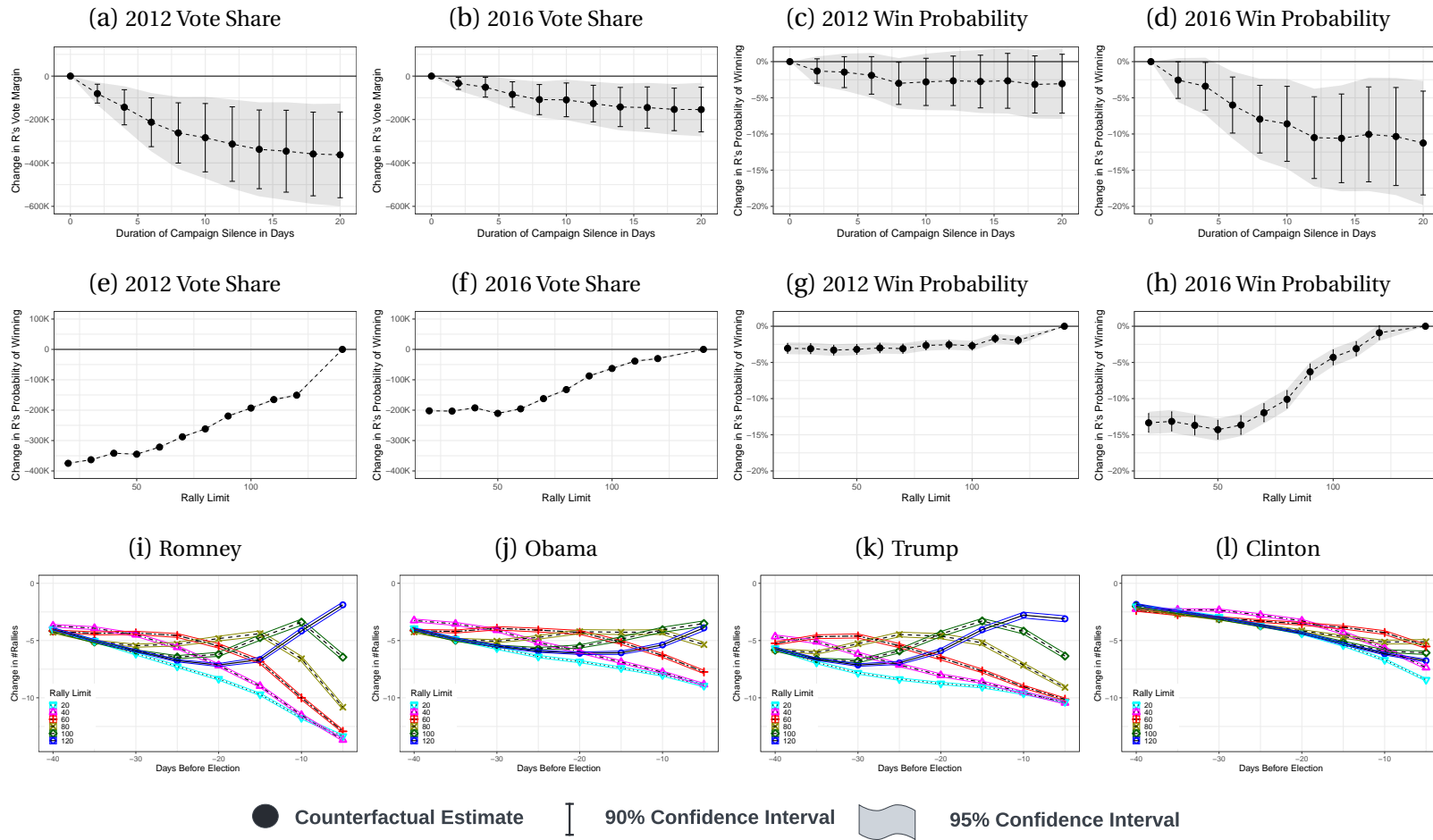


Figure 10: *Campaign Silence Duration and Election Day Poll Margin for 2012 Presidential Elections. This figure provides estimates for changes in electoral outcomes when campaign silences of varying duration are imposed. For each campaign silence duration, I calculate R's probability of winning along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

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.³⁸ 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, 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.

³⁸This behavior is absent in static models and only plays a role when one considers dynamic strategies.

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:

$$\mathbb{E} [V_{R,T+1}(p_{T+1})|a_{RT}, a_{DT}, p_T] = \sum_{k=1}^K e_k E\Phi\left(\frac{\alpha_R a_{RkT} + \alpha_D a_{DkT} + \rho p_T + \delta_k}{\sigma_v}\right), \quad (\text{A1})$$

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

$$u_{isT}(k, p_T, a_{jT}) = -c_i \mathbb{1}\{k \neq 0\} + \beta \mathbb{E} [V_{i,T+1}(p_{T+1})|a_{jT}, k, p_T], \quad (\text{A2})$$

which is also finite. Note that $a_{isT} = \arg \max_k \{u_{isT}(k, p_T, a_{jT}) + \epsilon_{i,k,s,T}\}$ which exists as each of the numbers are finite. Moreover, it is unique with probability 1 otherwise we will have two independent TIEV 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 $\sigma_{isT}(k; p_T, a_{jT}) = \frac{\exp(u_{isT}(k, p_T, a_{jT}))}{\sum_l \exp(u_{isT}(l, p_T, a_{jT}))}$. Similarly, for the first mover the option specific value function is given by:

$$u_{ifT}(k, p_T) = -c_i \mathbb{1}\{k \neq 0\} + \beta \sum_{l=0}^K \sigma_{jsT}(l, p_T, k) \mathbb{E} [V_{i,T+1}(p_{T+1})|l, k, p_T]. \quad (\text{A3})$$

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:

$$\mathbb{E}_{\epsilon_{is,T}} [V_{isT}(l, p_T, \epsilon_{is,T})] = \log \left(\sum_{k=0}^K u_{isT}(k, p_T, l) \right) + \gamma, \quad (\text{A4})$$

$$\mathbb{E}_{\epsilon_{if,T}} [V_{ifT}(p_T, \epsilon_{if,T})] = \log \left(\sum_{k=0}^K u_{ifT}(k, p_T) \right) + \gamma. \quad (\text{A5})$$

This is true because the cost shocks are T1EV.³⁹ Moreover, note that since u_{isT} and u_{ifT} are bounded their respective 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:

$$\begin{aligned} V_{i,T}(p_T) = & f_i \times \ln \left(\sum_{k=0}^K \exp \left\{ u_{if,t}(k; p_T) \right\} \right) \\ & + (1 - f_i) \times \sum_{k=0}^K \left[\sigma_{jf,T}(k; p_T) \ln \left(\sum_{l=0}^K \exp \left\{ u_{is,T}(l; k, p_T) \right\} \right) \right] + \gamma. \end{aligned} \quad (A6)$$

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

Suppose V_{t+1} is bounded then with similar arguments $u_{ist}(k, p_t, a_{jft}) = -c_i \mathbb{1}\{k \neq 0\} + \beta \mathbb{E} [V_{i,t+1}(p_{t+1}) | a_{jft}, k, p_t]$ is also bounded and so is $u_{ift} = -c_i \mathbb{1}\{k \neq 0\} + \beta \sum_{l=0}^K \sigma_{jst}(l, p_t, k) \mathbb{E} [V_{i,t+1}(p_{t+1}) | l, k, p_t]$. 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:

$$\mathbb{E}_{\epsilon_{is,t}} [V_{ist}(l, p_t, \epsilon_{is,t})] = \log \left(\sum_{k=0}^K u_{ist}(k, p_t, l) \right) + \gamma, \quad (A7)$$

$$\mathbb{E}_{\epsilon_{if,t}} [V_{ift}(p_t, \epsilon_{is,t})] = \log \left(\sum_{k=0}^K u_{ift}(k, p_t) \right) + \gamma, \quad (A8)$$

$$\begin{aligned} V_{i,t}(p_t) = & f_i \times \ln \left(\sum_{k=0}^K \exp \left\{ u_{if,t}(k; p_t) \right\} \right) \\ & + (1 - f_i) \times \sum_{k=0}^K \left[\sigma_{jf,t}(k; p_t) \ln \left(\sum_{l=0}^K \exp \left\{ u_{is,t}(l; k, p_t) \right\} \right) \right] + \gamma. \end{aligned} \quad (A9)$$

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 $t = 1, 2, \dots, T$.

³⁹Note if $-\epsilon_k \sim T1EV$ and independent then $\max_{k \in \{0,1,\dots,K\}} \{\tilde{\delta}_k - \epsilon_{ifk}\} \sim \text{Gumbel}(\mu = \ln \sum_k \exp(\tilde{\delta}_k), \beta = 1)$. 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 $A_t \in \{0, 1, \dots, K\}^2$ and $P_{t+1} \in \mathbb{R}^K$. In order to derive this density we consider the following probability:

$$\mathbb{P}[\tilde{X}_t \in B | \tilde{X}_{t-1}] = \sum_{a \in \{0, 1, \dots, K\}^2} \mathbb{P}[A_t = a, P_{t+1} \in B_a | (A_{t-1}, P_t)] \quad (\text{A10})$$

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 $\{\{0, 1, \dots, K\}^2 \times \mathbb{R}^K\}$ 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 $\{0, 1, \dots, K\}^2$ is finite. Note by model assumption on the popularity process the following holds:

$$\mathbb{P}[P_{t+1} \in B_a | (A_t = a, P_t)] = \int_{p \in B_a} f(p | A_t = a, P_t) dp \quad (\text{A11})$$

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

$$\begin{aligned} \mathbb{P}[\tilde{X}_t \in B | \tilde{X}_{t-1}] &= \sum_{a \in \{0, 1, \dots, K\}^2} \mathbb{P}[A_t = a, P_{t+1} \in B_a | (A_{t-1}, P_t)] \\ &= \sum_{a \in \{0, 1, \dots, K\}^2} \mathbb{P}[P_{t+1} \in B_a | (A_t = a, P_t, A_{t-1})] \mathbb{P}[A_t = a | P_t, A_{t-1}] \\ &= \sum_{a \in \{0, 1, \dots, K\}^2} \mathbb{P}[P_{t+1} \in B_a | (A_t = a, P_t)] \mathbb{P}[A_t = a | P_t] \\ &= \sum_{a \in \{0, 1, \dots, K\}^2} \left(\int_{p \in B_a} f(p | A_t = a, P_t) dp \right) \sigma_t(A_t = a; P_t) \\ &= \sum_{a \in \{0, 1, \dots, K\}^2} \int_{p \in B_a} f(p | A_t = a, P_t) \sigma_t(A_t = a; P_t) dp \\ &\Rightarrow \int_{x \in B} \psi_t(x | X_{t-1}) dx = \int_{(a, p) \in B} f(p | A_t = a, P_t) \sigma_t(A_t = a; P_t) d(a, p) \end{aligned} \quad (\text{A12})$$

The second equality holds by law of total probability. The third equality holds by modeling assumption 2.1. The fourth equality is substituting $\mathbb{P}[P_{t+1} \in B_a | (A_t = a, P_t)]$ 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 $B \subset \{0, 1, \dots, K\}^8 \times \mathbb{R}^{4K}$:

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

Finally we have

$$\begin{aligned}
& \Rightarrow \mathbb{P} \left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3} \right) \in B \middle| \tilde{X}_{4d-4} \right] \\
&= \int_{(a_l, p_{l+1})_{l=1, \dots, 4} \in B} \left(\prod_{l=1}^4 \sigma_{4(d-1)+l}(a_l; p_l) \right) \times \left(\prod_{l=1}^4 f(p_{l+1} | a_l, p_l) \right) d((a_l, p_{l+1})_{l=1, \dots, 4}) \quad (\text{A13})
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{P}[X_d \in C | X_{d-1}] &= \mathbb{P} \left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3} \right) \in B^C \middle| X_{d-1} \right] \\
&= \mathbb{P} \left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3} \right) \in B^C | (P_{4d-3}, A_{4d-4}) \right] \\
&= \mathbb{P} \left[\left(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3} \right) \in B^C | \tilde{X}_{4d-4} \right] \\
&= \int_{(a_l, p_{l+1})_{l=1, \dots, 4} \in B^C} \left(\prod_{l=1}^4 \sigma_{4(d-1)+l}(a_l; p_l) \right) \times \left(\prod_{l=1}^4 f(p_{l+1} | a_l, p_l) \right) d((a_l, p_{l+1})_{l=1, \dots, 4}) \\
&= \int_{x \in C} \left(\int_{p_2, p_3, p_4 \in \mathbb{R}^{3K}} \left(\prod_{l=1}^4 \sigma_{4(d-1)+l}(a_l; p_l) \right) \times \left(\prod_{l=1}^4 f(p_{l+1} | a_l, p_l) \right) d(p_1, p_2, p_3) \right) d(x) \quad (\text{A14})
\end{aligned}$$

Table A1: 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.

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 A13. The following equality is substitution of the expression found in A13. 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 $(\tilde{X}_{4d}, \tilde{X}_{4d-1}, \tilde{X}_{4d-2}, \tilde{X}_{4d-3})$ along the dimensions of P_{4d-2} , P_{4d-1} and P_{4d} .

B Data Appendix

B.1 Types of Events Held

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, 2004, 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 hangar 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 candidates sometimes 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.⁴⁰ I also removed rallies from states where at most two rallies were held. States with infrequent rallies are typically strongholds, where campaign events have negligible influence on electoral outcomes. Excluding these states from the structural estimation does not introduce substantial bias, as demonstrated clearly in Table A5.

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}

⁴⁰Not all such rallies were removed only four such rallies were removed to ensure at most four rallies in a day.

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 A1 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.

B.2 Last minute changes to rally schedules by U.S. Politicians

Table A2 documents several instances in which U.S. political figures—including presidents, governors, senators, and presidential candidates—cancelled, rescheduled, or added rallies at the last minute. These changes were often made within 24 hours of the scheduled event and were driven by evolving conditions such as protests, weather, public health, or strategic shifts in campaign focus. This institutional flexibility highlights that rally schedules are not binding commitments, but rather contingent plans subject to real-time revision. Accordingly, the modeling framework in the main text treats the strategic environment as a game without commitment, where candidates retain the option to revise their rally choices period-by-period based on current conditions. The table provides empirical support for this assumption.

Table A2: Notable Cancellations, Rescheduling, or Additions of U.S. Political Rallies

Name	Event Date	Change Date	Reason for Change	Location/Event
Donald Trump	Mar 11, 2016	Mar 11, 2016	Cancelled due to large protests	University of Illinois, Chicago
Donald Trump	Oct 2, 2020	Oct 2, 2020	Cancelled after COVID-19 diagnosis	Sanford, Florida
Donald Trump	Apr 20, 2024	Apr 20, 2024	Cancelled due to storm weather	Wilmington, North Carolina
Donald Trump	Jun 20, 2020	Jun 12, 2020	Rescheduled rally to avoid June-teenth controversy	Tulsa, Oklahoma
Joe Biden	Jul 17, 2024	Jul 17, 2024	Cancelled after testing positive for COVID-19	Las Vegas, Nevada
Hillary Clinton	Sep 12, 2016	Sep 11, 2016	Cancelled multi-day campaign trip after pneumonia diagnosis and health incident at 9/11 memorial	California (multiple cities)
Hillary Clinton	Jul 8, 2016	Jul 8, 2016	Cancelled rally with Biden out of respect for victims of Dallas police shooting	Scranton, Pennsylvania
Hillary Clinton	Nov 2, 2016	Nov 2, 2016	Cancelled rally after fatal shootings of two police officers	Des Moines, Iowa
Barack Obama	Oct 29–30, 2012	Oct 29, 2012	Cancelled due to Hurricane Sandy response	Multiple East Coast campaign stops
Mitt Romney	Sep 16, 2012	Sep 16, 2012	Cancelled due to fatal small plane crash at event site	Pueblo Weisbrod Aircraft Museum, Colorado
Mitt Romney	Oct 28, 2012	Oct 28, 2012	Cancelled campaign events due to Hurricane Sandy	Virginia Beach, Virginia
George W. Bush	Sep 11, 2001	Sep 11, 2001	All events cancelled after 9/11 attacks	Nationwide
Tim Walz (Gov.)	Sep 9 2024	Sep 9, 2024	Cancelled rally to respond to wildfires	Reno, Nevada
Phil Murphy (Gov.)	Jan 16 2019	Jan 16 2019	Phil Murphy cancels Paterson town hall at mosque amid protests over Jameek Lowery death	New Jersey
Doug Burgum (Gov.)	Jun 7, 2023	Jun 5, 2023	Last-minute rally to launch presidential campaign	Fargo, North Dakota
Amy Klobuchar (Sen.)	Mar 1, 2020	Mar 1, 2020	Cancelled rally due to protests over past prosecution	St. Louis Park, Minnesota

Notes: This table lists last-minute rally cancellations, additions, and rescheduling, compiled through an extensive search of news reports referencing changes to scheduled campaign events. The entries are not exhaustive; additional changes may have occurred for which information is not readily available. Coverage of such changes is particularly limited for earlier years, such as 2012, and likely omits some rally adjustments from that period.

C Simulated Likelihood Procedure

C.1 Simulated Likelihood

In order to calculate $\lambda_d^\theta(X_d|X_{d-1})$ we need to execute an integration over \mathbb{R}^{3K} . This exercise is not feasible analytically and therefore we use a Quasi Monte-Carlo scheme that relies on Sobol sequence. We use $M = 2^{10} \times K$ points to evaluate $\lambda_d^\theta(X_d|X_{d-1})$. Let's denote the set of probability integral transforms of $3K$ dimensional Sobol sequence, till M , by $\zeta = \{\zeta^m = (\zeta_{1,1}^m, \dots, \zeta_{1,K}^m, \dots, \zeta_{3,1}^m, \dots, \zeta_{3,K}^m)\}_{m=1}^M$ ⁴¹. Based on ζ we can define the following set of plausible popularity values:

$$\hat{p}_{1,k}^{m,d} = P_{d,1,k} \quad (C1)$$

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

$$\hat{p}_{l+1,k}^{m,d} = \alpha_R \mathbb{1}\{A_{R,d,l} == k\} + \alpha_D \mathbb{1}\{A_{D,d,l} == k\} + \rho \hat{p}_{l,k}^{m,d} + \delta_k + \sigma_v \zeta_{l,k}^m \quad (C2)$$

Lastly call $\hat{p}_{5,k}^{m,d}$ 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 $P_{d,1}, A_{d,1}, \dots, A_{d,4}$. This gives us a plausible mean for observed popularity on day $d + 1$ sub-period 1. This predicted popularity is given by:

$$\hat{p}_{5,k}^{m,d} = \alpha_R \mathbb{1}\{A_{R,d,4} == k\} + \alpha_D \mathbb{1}\{A_{D,d,4} == k\} + \rho \hat{p}_{4,k}^{m,d} + \delta_k \quad (C3)$$

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

$$\begin{aligned} \lambda_d^\theta(X_d|X_{d-1}) &\approx \hat{\lambda}_d^\theta(X_d|X_{d-1}) \\ &\approx \frac{1}{M} \sum_{m=1}^M \left\{ \left(\prod_{l=1}^4 \hat{\sigma}_{4(d-1)+l}(A_{d,l}; \hat{p}_l^{m,d}) \right) \times \frac{1}{\sigma_v^K} \left(\prod_{k=1}^K \phi \left(\frac{P_{d+1,1,k} - \hat{p}_{5,k}^{m,d}}{\sigma_v} \right) \right) \right\} \end{aligned} \quad (C4)$$

Where $\hat{p}_l^{m,d} = (\hat{p}_{l,1}^{m,d}, \dots, \hat{p}_{l,K}^{m,d})$, the function $\hat{\sigma}_{4(d-1)+l}(A_{d,l}; \hat{p}_l^{m,d})$ is approximate probability of observing action profile $A_{d,l}$ in period $4*(d-1)+l$. This term is evaluated by using equations F10 and F8 in the Online Appendix. Moreover, $\phi(\cdot)$ is the p.d.f. of standard normal distribution. The density $\hat{\lambda}^\theta(X_d|X_{d-1})$ provides a close approximation of $\lambda^\theta(X_d|X_{d-1})$. If ζ were drawn from a standard normal distribution instead, call this density $\tilde{\lambda}^\theta(X_d|X_{d-1})$ then it is not hard to see that $\tilde{\lambda}^\theta(X_d|X_{d-1}) \rightarrow \lambda^\theta(X_d|X_{d-1})$ as $M \rightarrow \infty$. The error of this integral

⁴¹Here $\zeta_{l,k}^m$ is $\Phi^{-1}(u_{\text{Sobol},l,k}^m)$, 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.

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:

$$\begin{aligned} \ell(\theta; X_0, X_1, \dots, X_{\bar{D}}) &\approx \hat{\ell}(\theta; X_0, X_1, \dots, X_{\bar{D}}) \\ &\approx \frac{1}{\bar{D}} \sum_{d=1}^{\bar{D}} \log \left[\frac{1}{M} \sum_{m=1}^M \left\{ \left(\prod_{l=1}^4 \hat{\sigma}_{4(d-1)+l} (A_{d,l}; \hat{p}_l^{m,d}) \right) \times \frac{1}{\sigma_v^K} \left(\prod_{k=1}^K \phi \left(\frac{P_{d+1,1,k} - \hat{p}_{5,k}^{m,d}}{\sigma_v} \right) \right) \right\} \right] \end{aligned} \quad (C5)$$

D Additional Robustness Test

D.1 Model with Commitment: Scheduling Rallies a Week in Advance

To test a version of the model in which candidates must commit to scheduling rallies a week in advance, we modify the popularity transition equations accordingly. In this setting, candidates form forecasts of candidate R 's popularity 28 periods (one week) into the future based off opponent's committed schedules and current popularity shocks. After every period candidates update their forecast of week-ahead popularity based off their information set: their rally schedule, opponent rally schedule, and current popularity shocks for all states. Formally, in the beginning of period t , the information set of both players is given by $\{\{a_{R,s}, a_{D,s}\}_{s=t}^{t+27}, p_t\}$. Here $\{a_{R,s}, a_{D,s}\}_{s=t}^{t+27}$ is the rally schedule of candidates from period t to period $t + 27$. Note, that the candidates do not require to track rallies that took place before period t since, that information of its net-outcome is contained within the current observed popularity. Candidates choose $a_{i,t+28}$ in period t given the information set.

Let the current period be t . Then the transition equation for the forecasted popularity in period $t + 29$ for a forecast taken at period $t + 1$, given the information set in period $t + 1$ is given by:

$$\begin{aligned} \mathbb{E}_{t+1} p_{k,t+29} &= \rho^{29} \cdot p_{k,t} + \sum_{s=0}^{28} \rho^{28-s} \left(\delta_k + \sum_{i \in \{R,D\}} \alpha_i a_{i,k,t+s} \right) + \rho^{28} \sigma_v v_{k,t+1} \\ \Rightarrow \mathbb{E}_{t+1} p_{k,t+29} &= \underbrace{\sum_{i \in \{R,D\}} \alpha_i a_{i,k,t+28}}_{\equiv \mathbb{E}_t p_{k,t+29}} + \rho \mathbb{E}_t p_{k,t+28} + \delta_k + \underbrace{\rho^{28} \sigma_v v_{k,t+1}}_{\equiv \text{shock in } t} \end{aligned} \quad (D1)$$

where $\mathbb{E}_t p_{k,s}$ denotes the period s popularity forecast taken in period t based on period t information set. This information set includes all scheduled rallies up to period s and any shocks realized by period t .

The recursive form above helps reduce the dimensionality of the state space. Rather than tracking the full schedule of rallies directly, we summarize their influence through the forecasted popularity path, which internalizes the effects of scheduled rallies on future popularity.

Note that under this commitment model, decision-making concludes in period T , while the election occurs in period $T + 29$. The payoff that candidates expect to receive at time $T + 1$ is therefore based on this forecasted popularity trajectory and is given by:

$$\begin{aligned}
V_{T+1}(p_T, \{a_{R,T+s}, a_{D,T+s}\}_{s=0}^{28}, v_{T+1}) &= \\
&\mathbb{E}_{T+1} \left[\sum_{k=1}^K e_k E \cdot \mathbb{1} \left\{ \rho^{29} p_{k,T} + \sum_{s=0}^{28} \rho^{28-s} \left(\sum_{i \in \{R,D\}} \alpha_i a_{i,k,T+s} + \delta_k + \sigma_v v_{k,T+s+1} \right) > 0 \right\} \right] \\
&= \sum_{k=1}^K e_k \cdot E \cdot \Phi \left(\frac{\mathbb{E}_{T+1} p_{k,T+29}}{\sigma \sqrt{\sum_{s=0}^{27} \rho^{2s}}} \right) \\
&\approx \sum_{k=1}^K e_k \cdot E \cdot \Phi \left(\frac{\sqrt{1 - \rho^2} \cdot \mathbb{E}_{T+1} p_{k,T+29}}{\sigma} \right) \equiv V_{T+1}(\mathbb{E}_{T+1} p_{k,T+29}).
\end{aligned} \tag{D2}$$

The first equality defines V_{T+1} as the expected sum of electoral college votes that candidate R can win in period $T+1$, conditional on the information set at $T+1$, incorporating the scheduled actions over the next 28 periods and the popularity level at period T .⁴²

The second equality follows from the definition of the forecasted popularity $\mathbb{E}_{T+1} p_{k,T+29}$. It also uses the fact that the random component follows a multivariate normal distribution $\mathcal{N}(0, I_{28})$, which is affine invariant under the transformation matrix Ω , where

$$\Omega_{ss} = \rho^{28-s} \sigma_v, \quad \Omega_{ss'} = 0 \text{ for } s \neq s'.$$

This allows the sum of shocks to be treated as a normal variable with variance $\sigma^2 \sum_{s=0}^{27} \rho^{2s}$, leading to the cumulative normal distribution function $\Phi(\cdot)$. The final approximation uses a closed-form simplification under the assumption of geometric decay in the autoregressive structure.

The timing of decisions and order of information revelation remains the same, however the state variables change in value functions. Instead of tracking current popularity candidates track forecasted popularity defined in Equation D1. Instead of tracking the current

⁴²These payoffs are not discounted using a factor of β^{28} , since no decision is made after this period. I also estimate the case where these payoffs are discounted, as shown in columns (7) and (8) of Table A3, and find that the estimates do not change significantly but the model fit worsens. As a result I report the non-discounted payoff version in the main text of the paper.

rally of the first mover, the second mover candidate tracks rallies scheduled for period $t + 28$ ($a_{jf,t+28}$). Finally, in period t candidates do not make decisions for rallies to be held in period t but rather in period $t + 28$. The second mover and first mover option-specific value functions are expressed in the model as

$$u_{is,t}(k; l, \mathbb{E}_t p_{k,t+28}) = -c_j \times \mathbb{1}\{k \neq 0\} - \epsilon_{js,t,k} + \beta \mathbb{E}_t \left[V_{j,t+1}(\mathbb{E}_{t+1} p_{t+29}) \middle| a_{jt+28} = k, a_{it+28} = l, \mathbb{E}_t p_{k,t+28} \right], \quad (\text{D3})$$

$$u_{if,t}(k; \mathbb{E}_t p_{k,t+28}) = \sum_{l=0}^K u_{is,t}(k; l, \mathbb{E}_t p_{k,t+28}) \cdot \sigma_{js,t}(l; k, \mathbb{E}_t p_{k,t+28}), \quad (\text{D4})$$

where $\sigma_{js,t}$ is equilibrium choice probability of candidate j when they are the second mover. The equilibrium conditional choice probabilities and value functions then can be expressed as,

$$V_{i,t}(\mathbb{E}_t p_{k,t+28}) = f_i \times \ln \left(\sum_{k=0}^K \exp \left\{ u_{if,t}(k; \mathbb{E}_t p_{k,t+28}) \right\} \right) + (1 - f_i) \times \sum_{k=0}^K \left[\sigma_{jf,t}(k; p_t) \ln \left(\sum_{l=0}^K \exp \left\{ u_{is,t}(l; k, \mathbb{E}_t p_{k,t+28}) \right\} \right) \right] + \gamma, \quad (\text{D5})$$

$$\sigma_{if,t}(k; \mathbb{E}_t p_{k,t+28}) = \frac{\exp \left(u_{if,t}(k; \mathbb{E}_t p_{k,t+28}) - u_{if,t}(0; \mathbb{E}_t p_{k,t+28}) \right)}{1 + \sum_{l=1}^K \exp \left(u_{if,t}(l; \mathbb{E}_t p_{k,t+28}) - u_{if,t}(0; \mathbb{E}_t p_{k,t+28}) \right)}, \quad (\text{D6})$$

$$\sigma_{is,t}(k; l, \mathbb{E}_t p_{k,t+28}) = \frac{\exp \left(u_{is,t}(k; l, \mathbb{E}_t p_{k,t+28}) - u_{is,t}(0; l, \mathbb{E}_t p_{k,t+28}) \right)}{1 + \sum_{q=1}^K \exp \left(u_{is,t}(q; l, \mathbb{E}_t p_{k,t+28}) - u_{is,t}(0; l, \mathbb{E}_t p_{k,t+28}) \right)}. \quad (\text{D7})$$

The likelihood is constructed following similar steps as in the main model. However, the transition densities now incorporate both day $d-1$ and day $d-7$ to compute the probabilities of observing rallies and poll margins on day d . Because rally decisions are made one week in advance, candidates observe shocks on day $d-7$ and form expectations about the polling outcomes they will face on day d . As a result, poll margins on day d are influenced not only by their autoregressive dependence on day $d-1$, but also by rally decisions made on day $d-7$ in response to polls observed on day $d-7$.

D.2 Other Robustness Tests

Table A3: Additional Robustness Tests

Parameters	First Mover Prob $f = 0.33$		First Mover Prob $f = 0.67$		Total EC Votes $E = 157$		Schedule Rallies discounted elec. payoff		Alt. State Groups Ohio in MW		Raw Polls	Last 50 Days	
	2012	2016	2012	2016	2012	2016	2012	2016	2012	2016	2016	2012	2016
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
α_R	0.101 (0.0227)	0.0836 (0.0154)	0.101 (0.0249)	0.077 (0.0142)	0.105 (0.0278)	0.0867 (0.0229)	0.108 (0.0275)	0.0946 (0.0176)	0.0208 (0.00343)	0.0917 (0.0211)	0.0367 (0.011)	0.0915 (0.023)	0.0692 (0.0161)
α_D	-0.0422 (0.0135)	-0.0742 (0.0151)	-0.0562 (0.0179)	-0.062 (0.0132)	-0.0379 (0.0135)	-0.0767 (0.0217)	-0.0604 (0.0199)	-0.0658 (0.0137)	-0.0117 (0.00254)	-0.0592 (0.0117)	-0.11 (0.028)	-0.0454 (0.0159)	-0.0756 (0.0187)
ρ	0.99 (0.002)	0.991 (0.001)	0.992 (0.002)	0.99 (0.001)	0.99 (0.002)	0.988 (0.002)	0.993 (0.002)	0.991 (0.002)	0.992 (0.001)	0.99 (0.001)	0.993 (0.001)	0.991 (0.002)	0.992 (0.001)
σ	0.147 (0.0143)	0.16 (0.0148)	0.147 (0.0145)	0.16 (0.0148)	0.146 (0.0143)	0.161 (0.0151)	0.135 (0.0142)	0.145 (0.0131)	0.0433 (0.00348)	0.17 (0.0208)	0.283 (0.0197)	0.135 (0.0119)	0.145 (0.01)
c_R	2.9 (0.287)	2.36 (0.208)	2.85 (0.278)	2.43 (0.201)	2.44 (0.238)	2.01 (0.212)	2.6 (0.25)	2.32 (0.208)	2.77 (0.263)	2.47 (0.204)	2.16 (0.212)	2.73 (0.347)	2.02 (0.315)
c_D	2.83 (0.206)	3.25 (0.259)	2.88 (0.202)	3.22 (0.256)	2.59 (0.175)	2.85 (0.251)	2.73 (0.189)	3.06 (0.251)	2.8 (0.178)	3.15 (0.258)	3.47 (0.269)	2.83 (0.279)	3.03 (0.33)
Fixed Effects:													
Cost	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Poll Margins	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
\bar{L}	-6.562	-6.5462	-6.5805	-6.5723	-6.8004	-6.7846	-6.7832	-6.4989	-1.4889	-6.8787	-10.051	-7.4172	-7.8602
Observations	100	100	100	100	100	100	93	93	100	100	100	50	50

^a Note: The table shows estimates for model parameters under 3 modifications. Columns (1) and (2) set first mover probability $f = 0.33$ for the republican candidate to be the first mover. Columns (3) and (4) set first mover probability $f = 0.67$ for the republican candidate to be the first mover. Columns (5) and (6) calibrate $E = 157$. Columns (7) and (8) estimate the model where candidates schedule rallies one week in advance and the electoral payoffs are discounted by β^{28} to account for the fact that elections take place a week after (28 periods later) the period rallies are scheduled. Columns (9) and (10) estimate the model the specification where Ohio is made the part of Midwest state group rather than Northeast state group. Column (11) reports estimates for 2016 when simple raw averages of individual poll results are used to construct popularity. Columns (12) and (13) report estimates when the model is estimated over the last 50 days to test whether poll-frequency influences parameter estimates. The standard errors have been computed by using observation wise gradient and likelihood hessian. I use HAC estimation for this purpose.

References

- Acharya, A., E. Grillo, T. Sugaya, and E. Turkel (2022). Electoral campaigns as dynamic con-
tests.
- Aguirregabiria, V. and M. Marcoux (2021). Imposing equilibrium restrictions in the estima-
tion 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.
- 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.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press. See Chapter 6, Section 6.3 on temporal aggregation of AR(1) processes.
- 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.
- 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.

Supplemental Appendix: Not For Publication

E Monte-Carlo Experiments

E.1 OLS Estimation and Bias

A natural benchmark to consider is a reduced-form OLS regression that relates next-day polling margins to current-day popularity and rally activity. Specifically, one might estimate:

$$p_{k,d+1} = \alpha_R^{OLS} N_{R,kd} + \alpha_D^{OLS} N_{D,kt} + \rho^{OLS} p_{kd} + \delta_k + \tilde{v}_{k,d+1}, \quad (\text{A.2})$$

where $p_{k,d}$ is the Republican candidate's polling margin in state k on day d , and $N_{i,kd}$ denotes the total number of rallies by candidate $i \in \{R, D\}$ in state k on day d . This regression mirrors the structure of the popularity evolution equation in the model (Equation 2.1), and—at first glance—may seem like a valid approach for estimating the structural parameters α_R and α_D .

However, Equation A.2 does not yield consistent estimates of the rally effects. In the structural model, rallies and popularity evolve at a sub-daily frequency (four sub-periods per day), and candidates make rally decisions at each sub-period. While each decision precedes that sub-period's popularity shock, the total number of rallies on day d aggregates over all sub-periods. Likewise, the observed poll margin $p_{k,d+1}$ reflects cumulative effects of all rallies and all within-day shocks. Consequently, even if each rally decision is uncorrelated with the shock in its own sub-period, the daily sum of rallies is correlated with the total shock realized over the day.

This correlation violates the OLS exogeneity assumption. The composite error term $v_{k,d+1}$ in Equation A.2 accumulates unobserved innovations that are not orthogonal to the aggregated rally indicators $N_{i,kd}$. Intuitively, on days when realized shocks are more favorable (e.g., random boosts in candidate support), more rallies are likely to occur or appear more effective—creating a spurious upward correlation. Conversely, campaigns may reduce rally activity on bad-shock days, leading to downward bias. Either way, the independence assumption required for OLS consistency fails.

To illustrate the implications of this bias, I simulate data using the structural model with known parameter values and estimate Equation A.2 via OLS. Table A4 reports the results for varying persistence levels in the underlying AR(1) process. For example, when $\rho = 0.99$,

the true values of α_R and α_D are both 0.08, but OLS estimates are consistently biased downward—typically closer to 0.05–0.07. The table also shows that the bias becomes more severe when persistence is lower (e.g., $\rho = 0.85$), reflecting stronger attenuation due to less predictable popularity dynamics.

These results confirm that OLS underestimates the true effects of rallies and demonstrate the importance of using the structural likelihood, which correctly integrates over unobserved within-day popularity states. The transition density in Lemma 4.1 and the likelihood function derived in Section 4 explicitly resolve the endogeneity.

Table A4: OLS v. Model

Panel A: Monte-Carlo for $\rho = 0.85$						
Parameter Name	N: Sample Size	True Value	OLS Bias	Model Bias	OLS MSE	Model MSE
α_R	40	0.08	-0.014	0.015	0.019	0.005
α_R	80	0.08	-0.015	0.010	0.008	0.002
α_R	120	0.08	-0.025	0.004	0.006	0.003
α_R	160	0.08	-0.023	0.004	0.005	0.002
α_D	40	-0.08	0.017	-0.004	0.018	0.005
α_D	80	-0.08	0.007	-0.009	0.009	0.003
α_D	120	-0.08	0.017	-0.010	0.008	0.003
α_D	160	-0.08	0.012	-0.003	0.005	0.002
ρ	40	0.85	-0.030	-0.027	0.003	0.003
ρ	80	0.85	-0.015	-0.014	0.001	0.001
ρ	120	0.85	-0.010	-0.010	0.001	0.001
ρ	160	0.85	-0.007	-0.007	0.000	0.000
Panel B: Monte-Carlo for $\rho = 0.99$						
Parameter Name	N: Sample Size	True Value	OLS Bias	Model Bias	OLS MSE	Model MSE
α_R	40	0.08	0.003	-0.002	0.027	0.016
α_R	80	0.08	-0.006	-0.002	0.012	0.007
α_R	120	0.08	-0.010	0.000	0.008	0.004
α_R	160	0.08	-0.012	-0.002	0.006	0.003
α_D	40	-0.08	0.008	-0.003	0.025	0.012
α_D	80	-0.08	-0.010	-0.004	0.012	0.006
α_D	120	-0.08	0.001	-0.003	0.011	0.005
α_D	160	-0.08	-0.005	-0.002	0.007	0.003
ρ	40	0.99	-0.027	-0.017	0.001	0.001
ρ	80	0.99	-0.012	-0.009	0.000	0.000
ρ	120	0.99	-0.007	-0.006	0.000	0.000
ρ	160	0.99	-0.005	-0.005	0.000	0.000

Notes: This table reports the average OLS estimates of Equation (A.2) across 200 simulations for each value of the persistence parameter ρ and sample size N . Simulated data are generated using the structural model with true parameter values $\alpha_R = \alpha_D = 0.07$ and $\rho \in \{0.85, 0.95, 0.99\}$, the other values are given in text. Despite rally decisions being made before sub-period shocks are realized, OLS estimates are biased due to the correlation between the total number of rallies and accumulated within-day popularity shocks—an artifact of polling data being observed only once per day. The results illustrate that OLS fails to recover structural parameters even under ideal simulated conditions.

E.2 Assessing Effects of Removing Stronghold States

Table A5: Assessing impact of removing stronghold states

	Case 1		Case 2		Case 3		Case 4	
Parameter	DGP	Est	DGP	Est	DGP	Est	DGP	Est
α_R	0.084	0.085	0.084	0.087	0.084	0.089	0.084	0.088
α_D	-0.074	-0.074	-0.074	-0.077	-0.074	-0.08	-0.074	-0.077
ρ	0.98	0.976	0.98	0.977	0.98	0.977	0.98	0.979
c_R	1.36	1.386	1.36	1.385	1.36	1.378	1.36	1.378
c_D	1.26	1.274	1.26	1.269	1.26	1.272	1.26	1.264
δ_1	-0.007	-0.009	-0.007	-0.008	-0.03	-0.034	-0.03	-0.032
δ_2	0.007	0.008	0.007	0.008	0.03	0.033	0.03	0.029
σ	0.16	0.159	0.16	0.159	0.16	0.159	0.16	0.159
c_1	0.05	0.04	0.05	0.043	0.05	0.042	0.05	0.055

Notes: This table presents results from a Monte Carlo simulation designed to evaluate the robustness of parameter recovery when stronghold states are excluded from estimation. Data are generated from a four-state model where two states are strongholds (characterized by large drift magnitudes) and two are swing states. The estimation is then conducted on a reduced two-state model using only the swing states, mirroring the empirical strategy in the paper. Case 1 has EC configuration as $e_1 = 2/6$, $e_2 = 2/6$, $e_3 = 1/6$, and e_4 while drift configuration is $\delta_1 = -0.2$, $\delta_2 = 0.2$, $\delta_3 = -0.007$, and $\delta_4 = 0.007$, that is two large stronghold states and two small swing states. Case 2 deviates by modifying $e_2 = 1/6$ and $e_3 = 2/6$, that is one large and one small swing states and same for stronghold state. Case 3 deviates from Case 1 by setting $\delta_3 = -0.03$ and $\delta_4 = 0.03$, that is the competitiveness of swing states is lowered. Finally, case 4 deviates by setting $e_2 = 1/6$, $e_3 = 2/6$, and $\delta_3 = -0.03$ and $\delta_4 = 0.03$. That is combining the features of Case 2 and Case 4. The estimates for rally effectiveness, costs, persistence, and drifts remain close to their true values in all settings, demonstrating that excluding stronghold states does not materially bias parameter estimates and that identification is driven by strategic variation in competitive environments.

To assess whether excluding stronghold states biases the estimation of key model parameters, I conduct a set of Monte Carlo experiments based on a four-state data-generating process (DGP). In each simulation, two of the states are designed to be strongholds—characterized

by large drift magnitudes—which render the probability of electoral flipping due to rallies effectively zero. The remaining two states are swing states with comparatively smaller drifts, where campaign efforts are potentially more consequential.

Data are generated from this full four-state model, and in line with patterns observed in the actual data, one or two rallies are allowed to occur in the stronghold states. To achieve this empirically realistic rally distribution, state-specific costs were calibrated accordingly. Estimation is then performed on a two-state version of the model, discarding the stronghold states entirely. This mirrors the main empirical strategy in the paper.

Across 50 Monte Carlo replications, the mean parameter estimates from the two-state estimation are remarkably close to the true DGP values. Table A5 summarizes the results for four different configurations. The first two cases simulate low-drift (more competitive) swing states, with drift magnitudes of $\delta = \pm 0.007$ (implying long-run averages of roughly 0.7 percentage points). The latter two simulate high-drift (less competitive) swing states, with drifts of $\delta = \pm 0.03$ (implying long-run averages of roughly around 3 percentage points). For each drift setting, I consider two electoral college distribution: (i) both stronghold states are large and both swing states are small, and (ii) one large and one small stronghold state along with with one large and one small swing state.

The key takeaway is that removing stronghold states—defined by one candidate have strong lead over the other and at most two rallies—does not compromise parameter recovery. In all configurations, the estimates for effectiveness, costs, persistence, and drift remain stable and close to their true counterparts. This supports the validity of excluding stronghold states in the main analysis and confirms that the model is primarily identified off strategic variation in competitive states.

E.3 Assessing Effects of Grouping States

To evaluate whether coarsening states into regional groups biases the structural estimates, I conduct Monte Carlo simulations based on grouping a four-state version of the model into a two-state version of the model. The data-generating process (DGP) specifications for each simulation are provided in Tables A6 and A7. I examine two main environments: (i) a high-drift case where elections are more lopsided within states, and (ii) a low-drift case where states are more competitive.

For each environment, I consider four grouping strategies for the purpose of estimation:

- (a) Combine one large and one small state (based on electoral college votes), and mix Republican-leaning and Democratic-leaning states.
- (b) Combine large states with large states, and small states with small states, mixing party lean.
- (c) Combine Republican-leaning with Republican-leaning and Democratic-leaning with Democratic-leaning, mixing state sizes.
- (d) Combine Republican-leaning with Republican-leaning and Democratic-leaning with Democratic-leaning, matching large with large.

In each case, data are simulated from the full four-state DGP, but the model is estimated using grouped data (two aggregated states). I consider 50 Monte-carlo simulation for each of eight cases. I compare the estimated parameters to the known DGP values.

Across all grouping strategies and both drift environments, the estimated effectiveness parameters α_i reflect a more conservative pattern relative to the data-generating process (DGP) values. This is expected: the estimated effects correspond to the impact of rallies on a larger population unit—i.e., a group of states—rather than on individual states. Aggregating popularity across heterogeneous states naturally dampens within-group variation, and the model recovers the influence of rallies on the pooled group-level population rather than state level population. As such, the estimates reflect how rallies shift average popularity across regional blocs, smoothing over localized state-level dynamics while preserving the broader implications of campaign behavior.

Other parameters adjust accordingly: drift estimates δ_k adapt nonlinearly and do not simply converge to group averages, while cost estimates c_i tend to be lower in the grouped versions, consistent with the higher frequency of rallies per group.

The results also validate the model's policy implications. In the high-drift environment, where one party dominates in each state, the marginal effect of rallies on the total probability of winning is close to zero both in the DGP and the estimated model. In the low-drift environment, where states are more competitive, the estimated effect of rallies on the total probability of victory is smaller than—but directionally consistent with—the true effect, confirming that the grouped model remains conservative. These findings, presented in Tables A6 and A7, show that the counterfactual experiment results would hold even more strongly in a full 12-state specification, reinforcing the robustness of the grouped model's

insights.

Table A6: Assessing impact of grouping states: larger drift

Panel A: Estimates								
Parameter	Case 1		Case 2		Case 3		Case 4	
	DGP	Est	DGP	Est	DGP	Est	DGP	Est
α_R	0.084	0.046	0.084	0.03	0.084	0.042	0.084	0.043
α_D	-0.074	-0.038	-0.074	-0.024	-0.074	-0.036	-0.074	-0.037
ρ	0.98	0.984	0.98	0.993	0.98	0.974	0.98	0.974
c_R	1.36	0.692	1.36	0.736	1.36	0.677	1.36	0.652
c_D	1.26	0.567	1.26	0.565	1.26	0.558	1.26	0.578
δ_1	0.11	0.088	0.065	0.023	0.157	0.202	0.135	0.176
δ_2	-0.11	-0.087	-0.065	-0.025	-0.157	-0.201	-0.135	-0.176
σ	0.119	0.119	0.119	0.113	0.119	0.12	0.119	0.113
c_1	0	-0.011	0	-0.025	0	-0.021	0	-0.044
Panel B: EC Votes DGP v. Group Total								
Parameter	Case 1		Case 2		Case 3		Case 4	
	DGP	Total	DGP	Total	DGP	Total	DGP	Total
e_1	0.33	0.5	0.33	0.67	0.33	0.5	0.33	0.67
e_2	0.17	-	0.33	-	0.17	-	0.33	-
e_3	0.33	0.5	0.17	0.33	0.33	0.5	0.17	0.33
e_4	0.17	-	0.17	-	0.17	-	0.17	-
Panel C: Drift DGP v. Group Average								
Parameter	Case 1		Case 2		Case 3		Case 4	
	DGP	Avg	DGP	Avg	DGP	Avg	DGP	Avg
δ_1	0.2	0.11	0.2	0.065	0.2	0.16	0.2	0.135
δ_2	-0.07	-	-0.07	-	0.07	-	0.07	-
δ_3	-0.2	-0.11	-0.2	-0.065	-0.2	-0.16	-0.2	-0.135
δ_4	0.07	-	0.07	-	-0.07	-	-0.07	-
Panel D: CF Total Rally Effect DGP v. Grouped								
Parameter	Case 1		Case 2		Case 3		Case 4	
	DGP	Est	DGP	Est	DGP	Est	DGP	Est
R's effect on R's win probability	0	0	0.01	0.06	0	0	0	0.01
D's effect on R's win probability	0	0	0	-0.07	0	0	-0.01	0

Notes: This table reports results from Monte Carlo simulations designed to evaluate the robustness of parameter recovery when states are aggregated into regional groups. Data are generated from a four-state model with high drift magnitudes, corresponding to electorally lopsided states and therefore less competitive elections. The model is estimated on a two-group version where states are grouped using one of four strategies: (a) Case 1– combining large and small states with mixed partisan lean; (b) Case 2– grouping by size while mixing partisan lean; (c) Case 3– grouping by partisan lean while mixing size; and (d) Case 4– grouping by both size and partisan lean. For each configuration, 50 replications are conducted. Grouped estimates of rally effectiveness α_i are consistently smaller in magnitude than the underlying DGP values, reflecting attenuation due to aggregation. Other parameters (costs and drifts) adjust accordingly. These results show that grouping introduces conservativeness without invalidating policy-relevant conclusions.

Table A7: Assessing impact of grouping states: smaller drift

Panel A: Estimates								
Parameter	Case 1		Case 2		Case 3		Case 4	
	DGP	Est	DGP	Est	DGP	Est	DGP	Est
α_R	0.084	0.035	0.084	0.028	0.084	0.041	0.084	0.036
α_D	-0.074	-0.032	-0.074	-0.027	-0.074	-0.031	-0.074	-0.029
ρ	0.98	0.987	0.98	0.985	0.98	0.986	0.98	0.985
c_R	1.36	0.759	1.36	0.707	1.36	0.757	1.36	0.737
c_D	1.26	0.627	1.26	0.599	1.26	0.595	1.26	0.584
δ_1	0.011	0.007	0.007	0.002	0.016	0.009	0.013	0.007
δ_2	-0.011	-0.009	-0.007	-0.01	-0.016	-0.01	-0.013	-0.01
σ	0.119	0.12	0.119	0.114	0.119	0.12	0.119	0.114
c_1	0	-0.02	0	0.009	0	0.007	0	0.016
Panel B: EC Votes DGP v. Group Total								
Parameter	Case 1		Case 2		Case 3		Case 4	
	DGP	Total	DGP	Total	DGP	Total	DGP	Total
e_1	0.33	0.5	0.33	0.67	0.33	0.5	0.33	0.67
e_2	0.17	-	0.33	-	0.17	-	0.33	-
e_3	0.33	0.5	0.17	0.33	0.33	0.5	0.17	0.33
e_4	0.17	-	0.17	-	0.17	-	0.17	-
Panel C: Drift DGP v. Group Average								
Parameter	Case 1		Case 2		Case 3		Case 4	
	DGP	Avg	DGP	Avg	DGP	Avg	DGP	Avg
δ_1	0.02	0.011	0.02	0.007	0.02	0.016	0.02	0.0135
δ_2	-0.007	-	-0.007	-	0.007	-	0.007	-
δ_3	-0.02	-0.011	-0.02	-0.007	-0.02	-0.016	-0.02	-0.0135
δ_4	0.007	-	0.007	-	-0.007	-	-0.007	-
Panel D: CF Total Rally Effect DGP v. Grouped								
Parameter	Case 1		Case 2		Case 3		Case 4	
	DGP	Est	DGP	Est	DGP	Est	DGP	Est
R's effect on R's win probability	0.44	0.32	0.41	0.32	0.44	0.36	0.32	0.27
D's effect on R's win probability	-0.42	-0.29	-0.45	-0.34	-0.43	-0.27	-0.57	-0.37

Notes: This table reports results from Monte Carlo simulations designed to evaluate the robustness of parameter recovery when states are aggregated into regional groups. Data are generated from a four-state model with low drift magnitudes, corresponding to competitive elections. The model is estimated on a two-group version where states are grouped using one of four strategies: (a) Case 1—combining large and small states with mixed partisan lean; (b) Case 2—grouping by size while mixing partisan lean; (c) Case 3—grouping by partisan lean while mixing size; and (d) Case 4—grouping by both size and partisan lean. For each configuration, 50 replications are conducted. Grouped estimates of rally effectiveness α_i are consistently smaller in magnitude than the underlying DGP values, reflecting attenuation due to aggregation. Other parameters (costs and drifts) adjust accordingly. These results show that grouping introduces conservativeness without invalidating policy-relevant conclusions.

F Numerical Approximation

F.1 Primitives

Let $\mathbf{P} = \{(\mathbf{p}_1^r, \dots, \mathbf{p}_K^r)\}_{r=1}^R$ be the state variable grid. Let $\mathbf{T}(\mathbf{p}) = (\mathbf{T}_1(\mathbf{p}), \mathbf{T}_2(\mathbf{p}), \dots, \mathbf{T}_R(\mathbf{p}))$ 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:

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

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

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

I approximate the value functions by Chebyshev polynomials. Let the coefficients of the polynomial terms approximating $V_{i,t}(\cdot)$ be denoted by $\gamma_{i,t}^V$, $u_{i,f,t}(k; \cdot)$ by $\gamma_{i,t,k}^f$ and $u_{i,s,t}(k; l, \cdot)$ by $\gamma_{i,t,k,l}^s$. Apart from these, I also need a Gaussian quadrature for calculating conditional expectation. Let $v = \{(v_1^s, \dots, v_K^s, \omega^s)\}_{s=1}^S$ be a Gaussian quadrature.

F.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.

$$\tilde{u}_{i,s,T}(a_i; a_j, \mathbf{p}^r) = -c_i(1 - a_{i,0}) + \beta \sum_{k=1}^K E_k \Phi \left(\frac{\alpha_i a_{i,k} + \alpha_j a_{j,k} + \rho \mathbf{p}_k^r + \delta_k}{\sigma_v} \right) \quad (\text{F1})$$

$$\tilde{\sigma}_{i,s,T}(a_j; a_i, \mathbf{p}^r) = \frac{\exp(\tilde{u}_{i,s,T}(a_i; a_j, \mathbf{p}^r) - \tilde{u}_{i,s,T}(0; a_j, \mathbf{p}^r))}{1 + \sum_{l=1}^K \exp(\tilde{u}_{i,s,T}(l; a_j, \mathbf{p}^r) - \tilde{u}_{i,s,T}(0; a_j, \mathbf{p}^r))} \quad (\text{F2})$$

$$\tilde{u}_{i,f,T}(a_i; a_j, \mathbf{p}^r) = -c_i(1 - a_{i,0}) + \beta \sum_{a_j=0}^K \sum_{k=1}^K E_k \Phi \left(\frac{\alpha_i a_{i,k} + \alpha_j a_{j,k} + \rho \mathbf{p}_k^r + \delta_k}{\sigma_v} \right) \times \tilde{\sigma}_{i,f,T}(a_j; a_i, \mathbf{p}^r) \quad (\text{F3})$$

$$\tilde{\sigma}_{i,f,T}(a_j; \mathbf{p}^r) = \frac{\exp(\tilde{u}_{i,f,T}(a_i; \mathbf{p}^r) - \tilde{u}_{i,f,T}(0; \mathbf{p}^r))}{1 + \sum_{l=1}^K \exp(\tilde{u}_{i,f,T}(l; \mathbf{p}^r) - \tilde{u}_{i,f,T}(0; \mathbf{p}^r))} \quad (\text{F4})$$

$$\tilde{V}_{i,T}(\mathbf{p}^r) = f_i \log \left(\sum_{a_i=0}^K \exp \left(\tilde{u}_{i,f,T}(a_i; \mathbf{p}^r) \right) \right) + (1 - f_i) \sum_{a_j=0}^K \log \left(\sum_{a_i=0}^K \tilde{u}_{i,s,T}(a_i; a_j, \mathbf{p}^r) \right) \times \tilde{\sigma}_{j,s,T}(a_j; a_i, \mathbf{p}^r) \quad (\text{F5})$$

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

F.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 $\gamma_{i,t}^V$, $\gamma_{i,k,t}^f$, and $\gamma_{i,k,l,t}^s$. 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.

F.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:

$$I'_{r,k,l} = \sum_{s=1}^S \mathbb{T}_{r'} \left(\begin{bmatrix} \alpha_R \mathbb{1}\{k == 1\} + \alpha_D \mathbb{1}\{l == 1\} + \rho \mathbf{p}_1^r + \delta_1 + \sigma_v \nu_1^s \\ \vdots \\ \alpha_R \mathbb{1}\{k == K\} + \alpha_D \mathbb{1}\{l == K\} + \rho \mathbf{p}_K^r + \delta_K + \sigma_v \nu_K^s \end{bmatrix} \right) \omega^s \quad (\text{F6})$$

Here, ν_1^s, \dots, ν_K^s are Gaussian shocks and w^s is the weight of these shocks. The choice of the Gaussian quadrature is discussed later. Consequently define $I_{r,k,l} = (I_{r,k,l}^1, I_{r,k,l}^2, \dots, I_{r,k,l}^R)$. 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:

$$\tilde{u}_{i,s,t}(a_i; a_j, \mathbf{p}^r) = -c_i(1 - a_{i,0}) + \beta \sum_{r_2=1}^R \gamma_{i,t+1;r_2}^V I_{r,a_i,a_j}^{r_2} \quad (\text{F7})$$

Here $\gamma_{i,t+1;r_2}^V$ is the coefficient of the r_2^{th} term of the polynomial interpolating $V_{i,t+1}()$. The sum “ $\sum_{r=1}^R \gamma_{i,t+1;r_2}^V I_{r,a_i,a_j}^{r_2}$ ” 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:

$$\tilde{\sigma}_{i,s,t}(a_i; a_j, \mathbf{p}^r) = \frac{\exp\left(-c_i(1 - a_{i,0}) + \beta \sum_{r_2=1}^R \gamma_{i,t+1;r_2}^V (I_{r,a_i,a_j}^{r_2} - I_{r,0,a_j}^{r_2})\right)}{1 + \sum_{k=1}^K \exp\left(-c_i + \beta \sum_{r_2=1}^R \gamma_{i,t+1;r_2}^V (I_{r,k,a_j}^{r_2} - I_{r,0,a_j}^{r_2})\right)} \quad (\text{F8})$$

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

$$\tilde{u}_{i,f,t}(a_i; a_j, \mathbf{p}^r) = -c_i(1 - a_{i,0}) + \beta \sum_{a_j=0}^K \left(\sum_{r_2=1}^R \gamma_{i,t+1;r_2}^V I_{r,a_i,a_j}^{r_2} \right) \times \tilde{\sigma}_{j,s,t}(a_j; a_i, \mathbf{p}^r) \quad (\text{F9})$$

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

$$\tilde{\sigma}_{i,f,t}(a_i; \mathbf{p}^r) = \frac{\exp\left(-c_i(1 - a_{i,0}) + \beta \sum_{r_2=1}^R \gamma_{i,t+1;r_2}^V \left(\sum_{a_j=0}^K (I_{r,a_i,a_j}^{r_2} \tilde{\sigma}_{j,s,t}(a_j; a_i, \mathbf{p}^r) - I_{r,0,a_j}^{r_2} \tilde{\sigma}_{j,s,t}(a_j; 0, \mathbf{p}^r) \right)\right)}{1 + \sum_{k=1}^K \exp\left(-c_i + \beta \sum_{r_2=1}^R \gamma_{i,t+1;r_2}^V \left(\sum_{a_j=0}^K (I_{r,k,a_j}^{r_2} \tilde{\sigma}_{j,s,t}(a_j; k, \mathbf{p}^r) - I_{r,0,a_j}^{r_2} \tilde{\sigma}_{j,s,t}(a_j; 0, \mathbf{p}^r) \right)\right)} \quad (\text{F10})$$

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

$$\tilde{V}_{i,t}(\mathbf{p}^r) = f_i \log\left(\sum_{a_i=0}^K \exp(\tilde{u}_{i,f,t}(a_i; \mathbf{p}^r))\right) + (1 - f_i) \sum_{a_j=0}^K \log\left(\sum_{a_i=0}^K \tilde{u}_{i,s,t}(a_i; a_j, \mathbf{p}^r)\right) \times \tilde{\sigma}_{j,s,t}(a_j; a_i, \mathbf{p}^r) \quad (\text{F11})$$

F.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, $\mathbf{U} = \{(\mathbf{u}_1^r, \dots, \mathbf{u}_K^r)\}_{r=1}^R$ over $[-1, 1]^K$ and its corresponding Chebyshev polynomial $\Psi(u)$. Then the grid $\mathbf{P} = \{(\mathbf{p}_1^r, \dots, \mathbf{p}_K^r)\}_{r=1}^R$

for a given set of parameters $\rho, \sigma_v, \delta_1, \dots, \delta_K$ is constructed as followed:

$$\begin{aligned} \mathbf{p}^r &= \mathbf{p}_k + (\bar{p}_k - \mathbf{p}_k) \frac{\mathbf{u}_k^r + 1}{2} \\ \text{where } \bar{p}_k &= \frac{\delta_k}{1 - \rho} + \left(\frac{\alpha_R}{1 - \rho} + \frac{3\sigma_v}{\sqrt{1 - \rho^2}} \right) \\ \mathbf{p}_k &= \frac{\delta_k + \alpha_D}{1 - \rho} + \left(\frac{\alpha_D}{1 - \rho} - \frac{3\sigma_v}{\sqrt{1 - \rho^2}} \right) \end{aligned} \quad (\text{F12})$$

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

$$\mathbf{T}_r(p) = \Psi_r \left(2 \left(\frac{p_1 - \mathbf{p}_1}{\bar{p}_1 - \mathbf{p}_1} \right) - 1, 2 \left(\frac{p_2 - \mathbf{p}_2}{\bar{p}_2 + \mathbf{p}_2} \right) - 1, \dots, 2 \left(\frac{p_K - \mathbf{p}_K}{\bar{p}_K + \mathbf{p}_K} \right) - 1 \right) \quad (\text{F13})$$

Where $\Psi_r(\cdot)$ is the r^{th} Chebyshev polynomial term. The Gaussian quadrature, denoted by $\nu = \{(v_1^s, \dots, v_K^s, \omega^s)\}_{s=1}^S$, 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 $2^\mu + 1$.

F.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, $\theta_{\text{Popularity}} = \{\alpha_R, \alpha_D, \rho, \sigma_v, \delta_1, \delta_2, \dots, \delta_K\}$ and $\theta_{\text{Cost}} = \{c_R, c_D, c_1, c_2, \dots, c_K\}$ and the approximation level μ .

Step 0 Generate the Smolyak pair, \mathbf{U}, Ψ for K dimensions and approximation level μ by following [Judd et al. \(2014\)](#). Obtain KPN Gaussian quadrature $\nu = \{(v_1^s, \dots, v_K^s, w^s)\}_{s=1}^S$ 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 [F12](#) and [F13](#).

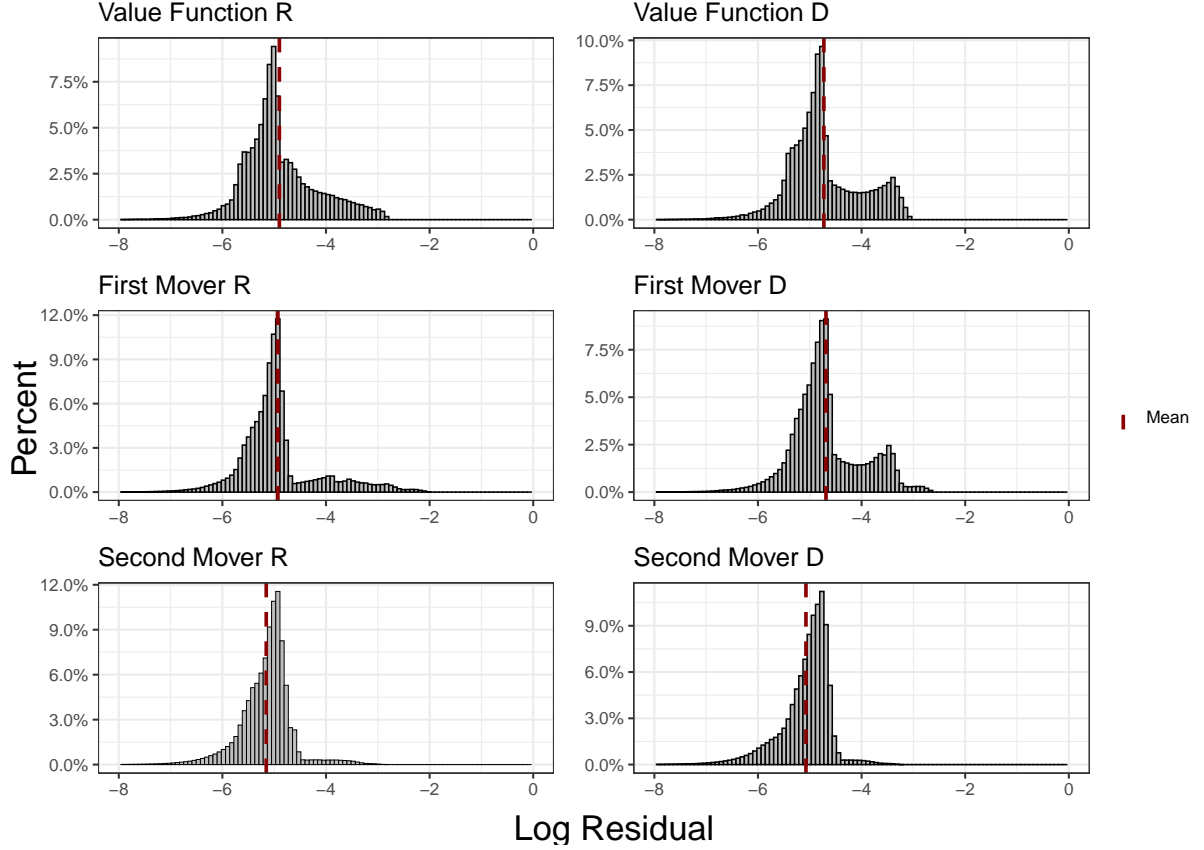
Step 2 Pre-Compute integrals of Chebyshev terms contingent on current popularity and candidate decisions by equation [F6](#).

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 [F5](#), [F3](#), [F1](#), [F4](#), [F2](#) respectively.
- Obtain coefficients of the interpolating polynomials, $\{\gamma_{i,T}^V, \gamma_{i,k,T}^f, \gamma_{i,k,l,T}^s\}$ for candidate $i = R, D$ and $k, l = 0, \dots, K$ for period $t = T$ respectively.

Figure A1: Residual Equation Errors



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

2 For $t = 1, 2, \dots, T - 1$ 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 F11, F9, F7, F10, F8 respectively.
- Obtain coefficients of the interpolating polynomials, $\{\gamma_{i,T-t}^V, \gamma_{i,k,T-t}^f, \gamma_{i,k,l,T-t}^s\}$ for candidate $i = R, D$ and $k, l = 0, \dots, K$ for period $t = T$ respectively.

E.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, $\{(p_{1,t,i}, \dots, p_{K,t,m})\}_{t=1}^T\}_{m=1}^{400}$. For each $p_{t,m} = (p_{1,t,m}, \dots, p_{K,t,m})$, let $Gp_{k,l}(p) = E[p_{t+1}|a_{R,t} =$

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

$$u_{i,s,t}^{t+1}(k; l, p_{t,m}) = -c_i(1 - a_{i,0}) + \beta \sum_{r=1}^R \gamma_{i,t+1,r}^V \left(\sum_{s=1}^S \mathbf{T}_r(Gp_{k,l}(p_{t,m} + \sigma_v v^s)) \omega^s \right) \quad (\text{F14})$$

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:

$$\mathcal{R}_{i,s,k,l,t}(\gamma; p_{t,m}) = 1 - \frac{u_{i,s,t}^{t+1}(k; l, p_{t,m})}{\hat{u}_{i,s,t}(k; l, p_{t,m})} \quad \text{for all } i, k, l, t \quad (\text{F15})$$

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

$$u_{i,f,t}^{t+1}(k; p_{t,m}) = \sum_{l=1}^k u_{i,s,t}^{t+1}(k; l, p_{m,t}) \left(\frac{\exp(u_{j,s,t}^{t+1}(l; k, p_{m,t}) - u_{j,s,t}^{t+1}(0; k, p_{m,t}))}{1 + \sum_{l'=1}^K \exp(u_{j,s,t}^{t+1}(l'; k, p_{m,t}) - u_{j,s,t}^{t+1}(0; k, p_{m,t}))} \right) \quad (\text{F16})$$

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

$$\mathcal{R}_{i,f,k,t}(\gamma; p_{t,m}) = 1 - \frac{u_{i,f,t}^{t+1}(k; p_{t,m})}{\hat{u}_{i,f,t}(k; p_{t,m})} \quad \text{for all } i, k, l, t \quad (\text{F17})$$

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 A1 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.

G Algorithm for Evaluating the Approximate Likelihood

G.1 Algorithm

Here I will describe the algorithm used for computing $\hat{\ell}(\theta; X_0, X_1, \dots, X_{\bar{D}})$ step-by-step. We will use the equations discussed in sections F and C. The algorithm will be defined for a given parameter values, $\theta = \{\alpha_R, \alpha_D, \rho, \sigma_v, \delta_1, \dots, \delta_K, c_R, c_D, c_1, \dots, c_K\}$ 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 F6.

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

- Calculate $\hat{p}_{l,k}^{m,d}$ for $k = 1, \dots, K, l = 1, \dots, 5$ and $m = 1, \dots, M$ using equations C1, C2 and C2.
- For each $m \in \{1, 2, \dots, M\}$ and $l \in \{1, 2, \dots, 5\}$ calculate $\hat{\sigma}_{4(d-1)+1}(A_{d,l}; \hat{p}_l^{m,d})$, where $\hat{p}_l^{m,d} = (\hat{p}_{l,1}^{m,d}, \dots, \hat{p}_{l,K}^{m,d})$, using equation F10 and F8.
- Calculate $\hat{\lambda}_d^\theta(X_d|X_{d-1})$ using equation C4.

Step 5 Calculate $\hat{\ell}(\theta; X_0, X_1, X_2, \dots, X_{\bar{D}})$ using equation C5.

H Addition Figures

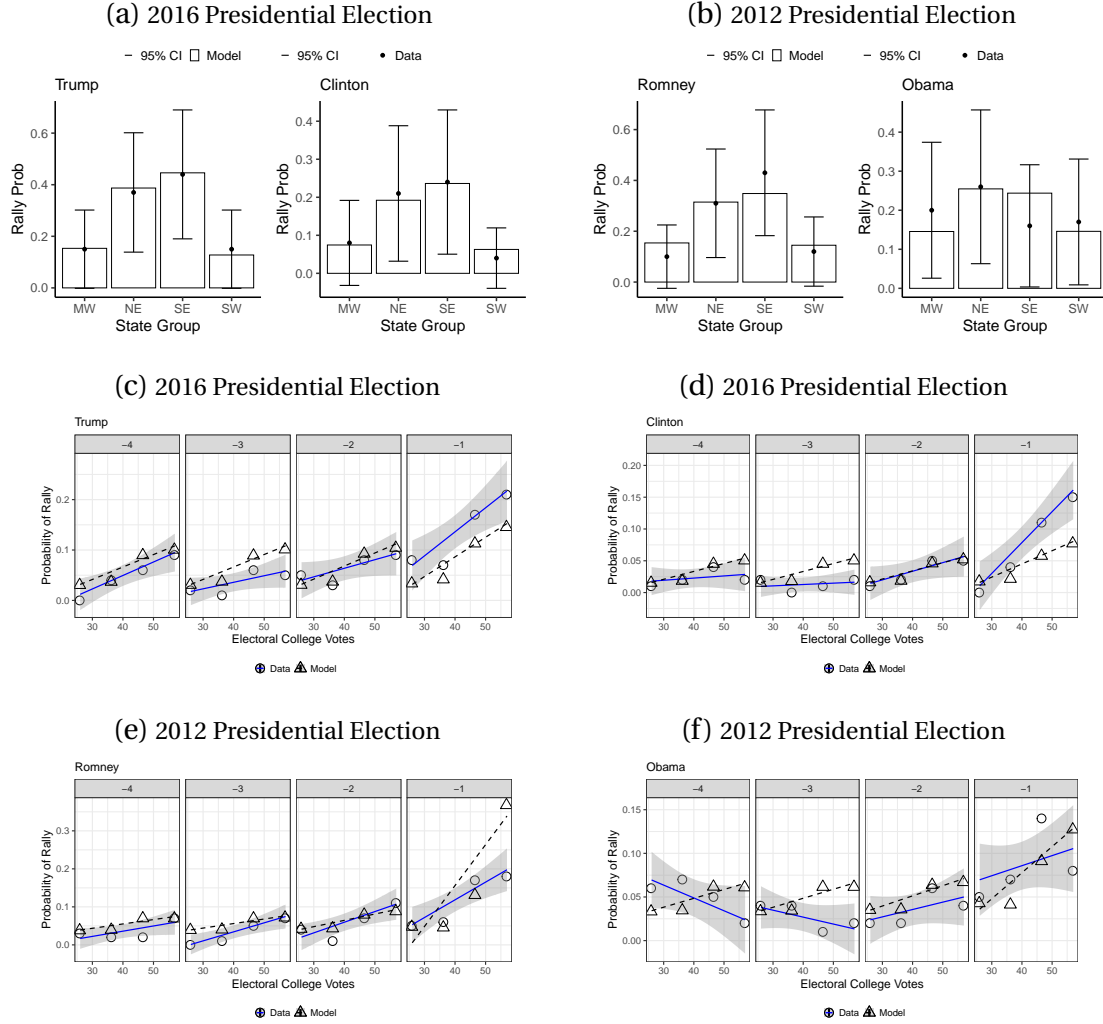


Figure A2: Panels (a) and (b) plots the model performance within the sample. The columns represent the model prediction, the solid points show observed probability of rallies for these groups and error-bars represent the 95% confidence intervals. Panels (c)-(f) show that the model supports the increasing 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.

I Additional Tables

Table A8: Correlation between Rallies and Electoral College Votes

Dependent Variable: Model:	Rally Count ($A_{i,d,k,y}$)				
	Full Sample (1)	Obama'12 (2)	Romney'12 (3)	Clinton'16 (4)	Trump'16 (5)
<i>Variables</i>					
$\mathbb{1} \{-100 \leq d \leq -76\} \times E_k$	0.001 (0.002)	-0.005** (0.002)	0.001 (0.003)	0.003 (0.002)	0.006** (0.003)
$\mathbb{1} \{-75 \leq d \leq -51\} \times E_k$	0.001 (0.002)	-0.004* (0.002)	0.006** (0.003)	0.0006 (0.002)	0.002 (0.003)
$\mathbb{1} \{-50 \leq d \leq -26\} \times E_k$	0.005** (0.001)	0.002 (0.002)	0.007** (0.003)	0.006*** (0.002)	0.007** (0.003)
$\mathbb{1} \{-25 \leq d \leq -1\} \times E_k$	0.009** (0.003)	0.002 (0.002)	0.008*** (0.003)	0.012*** (0.002)	0.013*** (0.003)
<i>Fixed-effects</i>					
$i \times y$	Yes	-	-	-	-
Day-Bin	Yes	Yes	Yes	Yes	Yes
<i>Fit statistics</i>					
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 A9: 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%.