Social Interactions in Voting Behavior: Evidence from India

Umair Khalil*  Sulagna Mookerjee†  Ryan Tierney‡

Abstract

We study how voters’ turnout decisions affect the turnout decisions of other voters by exploiting the unique staggered nature of India’s elections. Voting takes places in several phases, with constituencies quasi-randomly assigned to phases. At the time that later phases vote, turnout in earlier phases is public knowledge but results are not. Using an instrumental variables strategy, we find that a 1 percentage point (pp) increase in turnout in the previous phase depresses turnout in the current phase by around 0.38 pp. Falsification tests examining the effect on turnout, of either constituencies in the same phase or in future phases in the same election, produce no effect. Our results are broadly consistent with pivotal voter and ethical voter models, and suggest that partial equilibrium evaluations of turnout interventions may overstate general equilibrium effects.

Keywords: Voting Behavior; Staggered Elections; Election Spillovers

JEL Codes: C31; D72

*Corresponding author: Centre for Health Economics, Building H, 900 Dandenong Road, Monash Business School, Monash University, Caulfield East, 3145 VIC, Australia. Email: umair.khalil@monash.edu
†Binghamton University. Email: smookerj@binghamton.edu.
‡University of Southern Denmark. Email: ryan.etierney@gmail.com.

We are extremely grateful to Josh Kinsler, Gregorio Caetano, Josh Hall, Roland Hodler, Bhoomija Ranjan, David Slichter, and seminar participants at University of Rochester, Binghamton University (SUNY), CEA Annual Conference, European Winter Meetings of the Econometric Society, the RES Meetings, and the inaugural AASLE conference for helpful suggestions. We thank the Editor and two anonymous referees for their valuable feedback. Sundar Ponnusamy provided excellent research assistance. All remaining errors are our own.
1 Introduction

Considerable research has been devoted to the question of why people turn out to vote on election day. Early on, Downs (1957) highlighted the following paradox: since voting is costly, why do we observe so many people going to the polls? The probability that each vote influences the outcome drops rapidly in overall turnout, such that at observed turnout levels, each vote is essentially worthless \textit{ex ante}. Previous literature on turnout has focused on this fundamental question of why people vote at all, as well as on more detailed questions of what interventions can induce individual voters to vote, but has largely ignored the idea that inducing more people to vote can itself affect the turnout decision of other voters. In this paper, we address this question, the answer to which can have crucial implications for understanding broadly why people vote at all, and more narrowly whether it is reasonable to take partial equilibrium effects of increased turnout to be a good approximation of general equilibrium effects.

We exploit the unique staggered nature of the Indian General Election—voting in the Parliamentary Elections takes place across the country in several different phases, spanning several weeks. Voters in later stages have access to information about the turnout in previous stages, which can potentially affect their decision to vote on polling day in their own constituency. However, unlike other sequential voting settings, such as the US primaries, results of previous rounds are not released; only the level of turnout is public knowledge.

Our empirical strategy aims to uncover how average turnout in a given electoral phase affects voter turnout in subsequent phases of the same election. This estimated effect can, in essence, be interpreted as a ‘peer effect’ or a spillover, where a given constituency going to polls in the current phase has as its peer group all constituencies voting in the previous electoral phase. As established in the literature on social interactions, identification of causal effects in such a framework faces a number of potential threats (Blume et al., 2010). We use aspects of the institutional setting and an instrumental variable strategy to deal with these identification issues. First, our setting is immune to the so-called reflection problem (Manski, 1993) since a current phase constituency does not belong to its own peer group: constituencies that have voted in the previous electoral phase. Second, individual constituencies cannot choose the electoral phase they vote in, which is assigned to them by the Election Commission of India. Security is the primary consideration when assigning constituencies to phases, and the availability of paramilitary forces largely determines which constituency votes when. We present evidence later that this assignment is uncorrelated with contempo-

\footnote{Brock and Durlauf (2001) label such a setting to deal with the reflection problem as dynamic social interactions.}
raneous constituency specific political outcomes, aiding our identification strategy. Finally, 
correlation in constituency-specific unobservable shocks in a given election across phases can 
potentially create a spurious relationship between average turnouts in the previous phase 
and the current phase, which is dealt by employing an instrumental variable.

The construction of our IV exploits a historical feature of the Indian General Election. 
Since 1991, primarily due to the aforementioned security concerns, elections in India have 
lengthened both in duration (from 2-6 days to 2-5 weeks) and in the number of electoral 
phases. We refer to the post-1991 era as the staggered election era. The IV is constructed 
from constituency-specific average historical turnout in elections from 1977-1989, the pre-
staggered era. This captures a constituency’s intrinsic propensity to vote, giving us rele-
vance for our instrument. For validity, we argue that conditional on each current phase 
constituency’s own historical turnout, the average historical turnout of its assigned peer 
constituencies is unlikely to influence its current turnout. This is especially likely to hold 
given the quasi-random assignment of constituencies to electoral phases post-1991. In add-
ition, a new assignment order every election provides crucial variation in the peer group 
faced by each constituency in a given election as well.

We find that a 1 percentage point (pp) increase in turnout in a given phase depresses 
turnout in the subsequent phase by 0.38 pp. To further bolster our findings, we present two 
key falsification tests. First, if agents are learning from past turnout in the same election 
and we are successful in estimating this causally, as opposed to just capturing correlation in 
unobservable constituency-specific shocks, then our estimation strategy should yield no such 
association within the contemporaneous phase. Indeed, when we estimate this specification, 
results from our IV regressions yield coefficients that are statistically insignificant and of the 
opposite sign. Second, we also run a Granger-causality motivated falsification test where 
we estimate whether average turnout in the subsequent phase affects turnout in the current 
phase. Again we find no evidence for such an ‘effect’, with the point estimates being close 
to zero and statistically insignificant.

Although our results are derived from a unique institutional setting, the finding of higher 
turnout dissuading future voters from turning out can potentially be generalizable to a 
variety of settings. For instance, in the last few elections, around one-third of U.S. voters 
cast their ballots well before Election Day, with turnout statistics being readily available 
to future voters (Fullmer, 2015). Our findings also illustrate that targeted programs aimed 
at incentivizing voters to turn out might have a lower overall impact on increasing turnout 
than anticipated by policy makers. Finally, we provide evidence of strategic voting in large 
scale general elections, shedding light on fundamental voting behavior and contributing to 
the search for a micro-founded theory of political participation. In a final section, we also
adapt two influential papers from this theoretical literature, Feddersen and Sandroni (2006) and Coate and Conlin (2004) to our sequential voting setting, which help us interpret our empirical findings in light of their theoretical predictions.

The closest work to ours are studies by Deltas et al. (2015) and Knight and Schiff (2010), who explore the existence of social learning in the sequential voting setup of US presidential primaries. However, our findings are not comparable. For one, in a US presidential primary, results are released for each state as the round is completed. Second, and more importantly, a primary is a cumulative-vote vetting of individual candidates. Since each candidate remains a constant fixture throughout the process, each result also allows future voters to update their priors on candidate quality. In our parliamentary setting, where the constant fixture is the party, and not the candidate, there is no such signal. In contrast, Fafchamps et al. (2016) explore peer effects in voting behavior in an experimental setting providing evidence from an intervention implemented in Mozambique. Similar to our findings, they estimate negative peer effects in voter turnout among subjects with many connections to other surveyed subjects compared to those with fewer connections.

Morton et al. (2015) and Gerber et al. (2008) have studied the strategic response of voters to exit poll results, but responses to actual turnout in the same election has largely remained unexplored. As mentioned above, a huge literature spanning economics, philosophy, and political science has tried to answer the fundamental question of why people turn out to vote. The bulk of the literature has provided either alternative theoretical models or evidence from an experimental setting. Determinants of voter turnout studied previously include social image (DellaVigna et al., Forthcoming, Ali and Lin, 2013), population size and stability (Mueller, 2003, Ashworth et al., 2006), past turnout and habit formation (Denny and Doyle, 2009, Fujiwara et al., 2016), competitiveness of past elections (Blais, 2000, Fraga and Hersh, 2010), and campaign expenditure (Lau and Pomper, 2001, Kirchgässner and Schulz, 2005) among others. Geys (2006) and Cancela and Geys, 2016, in a review of aggregate-level studies, concludes that there is little consensus on the exact determinants of voter turnout.

The rest of the paper is organized as follows. Section 2 provides details of the institutional setup of the Indian General Election, setting the stage for our identification strategy. Section 3 provides details of the data set and some descriptive statistics about elections in India. Our identification strategy is outlined in detail in Section 4, before moving on to estimation and results in Section 5. Section 6 then provides a discussion of our empirical finding in light of existing models of voter behavior from the political economy literature. Section 7 concludes.
2 Institutional Setup of the Indian Elections

India has a parliamentary electoral system, consisting of two houses, the Lower House called Lok Sabha, and the Upper House called Rajya Sabha. Lok Sabha elections (or Parliamentary elections) are typically held once every five years, and the electoral units are called parliamentary constituencies. Each state is divided into several constituencies in proportion to the size of its electorate, resulting in a total of 543 constituencies, each of which elects its Member of Parliament by plurality voting. India has many national and regional parties (146 in 1991, and as many as 231 in 2004), which have varied representation across the country. The party (or alliance, or coalition) winning more than half the seats forms the Government.

Because of the size of the electorate (a staggering 814 million in 2014, having almost doubled from the 498 million in 1991), the elections in India are now held in multiple phases. Until 1989, the election was held in two phases over four to ten days. For the 1991 election, the number of phases and the total length of the election rose sharply. The schedule of polling is determined by the Election Commission of India (ECI), an autonomous federal authority that is responsible for conducting the elections. T.N. Seshan, the 10th Chief Election Commissioner of India, implemented several changes to boost the transparency of the election (Gilmartin, 2009), and among them was the deployment of federal security forces to stave off violence during elections and maintain impartiality of electoral procedures. The election schedule thereafter has been drawn up in a way that allows the security forces to get from one area to another in time, lengthening the total duration of the election. The schedule is such that different constituencies in the same state often go to the polls on different phases, and the order in which the constituencies are allotted to phases varies across elections. According to the ECI, the assignment is quasi-random, and determined by factors such as examination schedules, weather, and crop harvesting cycles etc., which are arguably orthogonal to prior election outcomes.

Although the elections themselves are staggered across phases, the ECI releases results only after the entire exercise of polling has taken place across the country. Until then, the only information that is available about past phases is total voter turnout, which is widely discussed in newspapers and across digital media.

---

2 An election can be held if the government collapses, for instance due to the withdrawal of support from its allies. Three elections occurred between 1996 and 1999 for this reason.

3 For instance, local police was largely perceived as being partisan in India, a concern endorsed by a former Chief Election Commissioner S. Y. Qureshi himself.

4 We explore this in detail in Section 5.1.

5 Exit polls were only held in 2004, and are not popular in India, with the ECI seeking to prohibit the release of exit poll information before the election is over, and ban opinion polls 45 days ahead of the first phase of polling. Restrictions on opinion and exit polls were placed as far back as 1998, though they were more moderate (Patnam, 2013).
media is critical for the dissemination of this turnout information. In 1991, as part of a broader liberalization program, the government permitted private and foreign broadcasters to operate in India. In the late 1980’s, there was only one television channel; by 1995, there were more than 100 channels catering to more than 400 million viewers. Because of the combined increase in election length, and the expansion in mass media penetration, we focus on the post-1989 elections in our analysis.

Finally, in order to ensure that the constituencies are consistent with the regional demographics, the Delimitation Commission periodically redraws constituency boundaries according to census data. Though the state representation is unchanged, the Delimitation Commission can change the boundaries of a constituency within a state, and the number of seats in a state reserved for minority candidates. Delimitation was suspended after 1976 to ensure that family planning programs adopted by different states, and therefore changes in their populations, would not affect their representation in the Parliament. Delimitation occurred again in 2008, after more than thirty years. We therefore include only elections post 1976 in our analysis, and drop the 2009 and 2014 elections since the delimitation activities left their constituency compositions different from the other election years.

3 Data

The data for this study comes from detailed reports published by the Election Commission of India. The reports provide rich information about the candidates, the overall electorate, the electors who voted, and the exact date of polling for each constituency. Of the total number of votes, we know the exact number that were valid and the number that were rejected. The overall turnout variable is the fraction of the total electorate who cast valid votes.

The actual date of polling at each constituency helps us determine the voting phase to which it belongs, and the number of unique polling dates determines the number of phases in that election.\footnote{A handful of constituencies, from one election to another, went to the polls on a different day, for reasons mostly orthogonal to election turnout, such as local and state holidays. We assign these constituencies to the nearest voting phase.} Figure 1 shows the total length, the number of phases, and the average length of time in between consecutive phases, for all parliamentary elections from 1977 to 2014. We see a distinct shift in 1991: the red bars, denoting the average gap between phases, more than doubles to around 5 days post 1991. The number of phases, denoted by the green bars, grows to at least 3 post 1991, while they’re mostly either 1 or 2 pre 1989. Moreover, there is substantial increase in the overall length of the election as well, captured by the blue bars.\footnote{The increase in the length of the elections is almost monotonic, but there is a spike in 1991, which is...}
Table 1: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># Phases</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Correlation(Phase, Lagged Phase)</td>
<td>-</td>
<td>-0.0308</td>
<td>0.2632</td>
<td>-0.0069</td>
<td>0.1157</td>
</tr>
<tr>
<td># Polling Stations</td>
<td>1106.2</td>
<td>1413.4</td>
<td>1423.0</td>
<td>1426.6</td>
<td>1266.1</td>
</tr>
<tr>
<td>Turnout</td>
<td>0.5687</td>
<td>0.5881</td>
<td>0.6251</td>
<td>0.6060</td>
<td>0.5908</td>
</tr>
<tr>
<td>Average Electorate Size</td>
<td>956,552</td>
<td>1,091,293</td>
<td>1,115,801</td>
<td>1,140,952</td>
<td>1,236,626</td>
</tr>
<tr>
<td>Total Votes Polled</td>
<td>542,612</td>
<td>632,243</td>
<td>691,246</td>
<td>684,474</td>
<td>718,138</td>
</tr>
<tr>
<td>Candidates Contested Per Constituency</td>
<td>16.5</td>
<td>23.9</td>
<td>8.7</td>
<td>8.5</td>
<td>10.0</td>
</tr>
<tr>
<td>Margin of Victory</td>
<td>0.0796</td>
<td>0.0700</td>
<td>0.0615</td>
<td>0.0594</td>
<td>0.0736</td>
</tr>
<tr>
<td>Observations</td>
<td>521</td>
<td>543</td>
<td>543</td>
<td>543</td>
<td>543</td>
</tr>
</tbody>
</table>

Figure 1: Phases and Election Length, 1977-2014
Table 1 provides basic descriptive statistics. The overall turnout is around 60%, ranging between 57% in 1991 to 62% in 1998. The average margin of victory varies between 5.9% of the electorate in 1998 and around 8% in 1991. Crucially, there is very little correlation between the phase that a constituency is assigned to in a particular election and the phase it was allotted in the previous election. The number is highest for 1998, at 0.26, but ranges between -0.03 and 0.11 for the other election years. We explore the dynamics of the assignment of constituencies to electoral phases in detail in Section 5.1.

Figure 2 presents choropleth maps of assignment of constituencies to electoral phases across elections in India. We see considerable variation in phase assignments across elections and different parts of the country go to polls in different phases from election to election. This provides a crucial ingredient to our identification strategy that we outline in detail below. Although the maps underscore this point visually, we undertake a much more detailed exercise in section 5.1 that explores this ‘churning’ of constituencies across elections.

4 Identification Strategy

In this section we detail the empirical methodology that we employ to study how voters learn from observing voting behavior in past phases, and subsequently form their decision of whether to turn out to vote or not on the polling date of their own constituency. Consider the following empirical specification,

\[ V_{iep} = \beta_0 + \gamma \bar{V}_{jep-1} + X'_{iep} \beta_1 + \varepsilon_{iep} \]  

where \( V_{iep} \) represents voter turnout in constituency \( i \) in election \( e \) and in phase \( p \). Our baseline analysis defines voter turnout in the standard way: the total number of votes polled in a given constituency divided by the total number of eligible voters in that constituency. The coefficient, \( \gamma \), for average voter turnout in the previous phase, \( \bar{V}_{jep} \), is our key parameter of interest, measuring the effect of past turnout on current turnout across phases in the explained by the assassination of the President of the Indian National Congress and former Prime Minister Rajiv Gandhi after the first phase of elections, resulting in a postponement of the two later phases.

8For the state of Jammu and Kashmir our maps only show 4 out of 5 constituencies that lie in the Indian Administered Kashmir. The fifth constituency, Ladakh, although part of our main analysis is not shown here since our shape files include Pakistani Administered Kashmir in the geographic confines of Ladakh.

9The 1991 Elections in Jammu and Kashmir and Punjab were canceled or delayed due to security and insurgency related concerns. Therefore, there is no electoral data for them in Figure 2a. We perform a robustness check by dropping the 1991 election altogether and results remain similar to the baseline findings presented later in the paper.
Figure 2: Lok Sabha Constituency Assignments to Electoral Phases
same election.\textsuperscript{10} The turnout in the previous phase of the same election is calculated as the following weighted average,

\[ \bar{V}_{j_{ep}^{-1}} = \frac{\sum_{j \in J_{p-1}} w_j V_{j_{ep}^{-1}}}{\sum_{j \in J_{p-1}} w_j} \]  

(2)

where \( V_{j_{ep}^{-1}} \) refers to the turnout of a constituency \( j \), in election \( e \) that belongs to the set of constituencies \( J_{ep}^{-1} \) going to poll in the previous phase, \( p-1 \). The weights, \( w_j \), are calculated using total electoral size of a given constituency \( j \) divided by the total electoral size of all constituencies going to poll in phase \( p-1 \). This is designed to capture the idea that larger constituencies might feature more prominently in subsequent voters’ beliefs and hence should be given greater weight in the construction of average turnout in the previous phase.\textsuperscript{11} \( X'_{iep} \) controls for other covariates outlined below including a measure of persistence in turnouts widely documented in the literature. Finally, \( \varepsilon_{iep} \) is an observation level idiosyncratic shock.

Estimation of equation (1) through OLS suffers from a number of endogeneity concerns that are likely to render the estimation of \( \gamma \) biased. Given that we are essentially estimating a spillover parameter, below we outline the threats to identification in a framework outlined by Manski (1993) in his influential work on the identification of endogenous social effects.

### 4.1 The Reflection Problem and Endogenous Sorting into Electoral Phases

In our setup, each constituency goes to poll in a given phase, \( p \), and observes the turnout of constituencies who have already gone to polls in the previous phase, \( p-1 \). In this sense, the set of constituencies belonging to \( J_{ep}^{-1} \) comprise the ‘peer group’ of constituency \( i \) in phase \( p \). Given the nature of the setup, constituency \( i \) does not belong to its own peer group and hence breaks the link that generates the reflection problem, as defined by Manski (1993).\textsuperscript{12} However, more crucially our setup allows us to bypass concerns regarding potential endogenous sorting or selection into electoral phases. As discussed in Section 2, a given constituency has no control over the phase it goes to polls in, which is decided by the Election Commission of India (ECI) before each election. If assignment to electoral phases was a function of expected turnout, political beliefs, etc. then it would have added a further

\textsuperscript{10}We experiment with various ways of defining past turnout including a cumulative average up to phase \( p \), and twice lagged specifications. The results are largely similar across specifications although owing to our total number of observations, sample size in the twice lagged specifications fall considerably. We therefore focus on the once lagged specification throughout the rest of the paper.

\textsuperscript{11}We repeat our analysis using the unweighted mean as well and the results and conclusions drawn are qualitatively similar.

\textsuperscript{12}Our estimates, therefore, even from an OLS treatment of equation (1), would not suffer from a simultaneity problem common in the peer effects and spillover literature.
layer of complication in identifying the causal effect of interest. According to the ECI, it considers a number of factors that determine this assignment, the primary factor being the logistics of deploying paramilitary troops to ensure a free, fair, and peaceful election. Overall, these factors are out of the control of individual political parties and voters in a given constituency, and hence the assignment to ‘peer groups’ can be considered as good as random. In section 5.1 we provide evidence that this is indeed likely to be the case in our setting. However, it is important to note that employment of instrumental variables in the context of peer effects can help deal with both endogenous sorting into peer groups (phases in our case) as well as the existence of correlated effects, which we outline next.

4.2 Correlated Effects

A major threat to identification in models with spillovers is the potential existence of correlated effects and homophily channels, where units in the same peer group experience common shocks or behave similarly due to similar characteristics (Manski, 1993). In other words, the covariance between error terms of \( i \) and constituencies in \( J_{ep-1} \) may be non-zero, or formally,

\[
\text{Cov}(\varepsilon_{iep}, \frac{1}{J_{p-1}} \sum_{j \in J_{p-1}} \varepsilon_{jep-1}) \neq 0. \tag{3}
\]

In this case, OLS estimation of equation (3) will yield biased estimates of \( \gamma \). We therefore require an exogenous source of variation that affects average turnout in phase \( p-1 \) but is uncorrelated to election and phase specific shocks. For instance, if there is election related violence in a given region just before phase \( p-1 \), that can deter voters in both subsequent phases from turning out to vote creating a spurious relationship between turnout in phase \( p-1 \) and phase \( p \). Moreover, such shocks cannot be taken care of even with phase specific fixed effects and hence the need for an instrument becomes crucial. Homophily related arguments follow a similar reasoning, where constituencies that are alike in observable or unobservable ways turn out to vote based on those similar characteristics rather than learning from the turnout of the previous phase. Our instrumental variable strategy is designed to tackle such endogeneity concerns and is outlined next.

4.3 Instrumental Variable Strategy: Historical Turnout in Elections from 1977-1989

The instrumental variable we use is the average historical turnout of a given constituency in the four elections from the pre-staggered era, over the years 1977 to 1989. Previous
literature (see Fujiwara et al., 2016 and references therein) has found persistence in turnout across time for the same electoral district, and as a result we can expect a strong first stage, demonstrating the relevance of the IV.

The validity of our instrument is governed by two related factors. First, turnout in constituency $j$ in historical elections is unlikely to influence electoral outcomes in constituency $i$ in the current regime, especially after conditioning on the average historical turnout of $i$ itself. This is further bolstered by the observation that the contemporaneous order and assignment of constituencies to phases is itself quasi-random and, as outlined in section 5.1, there is sufficient churning of constituencies from election to election. In other words, a constituency $i$ gets a different mix of ‘peer constituencies’, i.e. those who vote before $i$, across elections. Second, as discussed earlier, Indian elections in the pre-1991 period had typically only 2 phases spaced over 4-6 days as compared to being drawn out over several weeks post 1991. Therefore, existence of learning opportunities for voters from past phases would be minimal. We construct the average turnout variables in both our structural equation and for our IVs by weighting each constituency in the calculation of the average turnout by its total number of registered voters in the respective phase or election. This would be particularly crucial if voters pay more attention to turnout numbers from larger constituencies compared to smaller ones.

In our set of covariates in equation (1), we add a historical average turnout variable for constituency $i$ to control for habit formation in voting. Furthermore, it is likely that constituencies are heterogeneous in their evaluation of the subjective costs and benefits of voting, and to the extent that this remains static over time, this can be proxied by the historical turnout control as well.\textsuperscript{13,14} In addition to this, we include a phase fixed effect to control for phase specific phenomenon, such as voter fatigue. Concerns regarding a lower probability of being the pivotal voter would also contribute to potentially depressing turnout in later phases. Finally, we add a control for the number of days that have elapsed since constituencies in phase $p - 1$ voted to capture any recall bias related channels.\textsuperscript{15}

\textsuperscript{13}We thank an anonymous referee for pointing this out.
\textsuperscript{14}Another way to deal with such concerns is to introduce constituency fixed effects to our model. We implement this as a robustness exercise and the point estimates diminish slightly but remain statistically significant.
\textsuperscript{15}We also try specifications where we control for average voter density per polling booth in a given constituency. Higher density would imply a higher social pressure and civic responsibility effect and would lead more individuals to turnout to vote. At the same time it can also lead to higher costs of voting by increasing wait times at polling station. Our results are robust to this inclusion with the estimated effects from the IV specifications being slightly higher in magnitude.
Table 2: Descriptive Statistics - Phase Orders

<table>
<thead>
<tr>
<th>Year</th>
<th>1996</th>
<th>1998</th>
<th>1999</th>
<th>2004</th>
</tr>
</thead>
<tbody>
<tr>
<td># Phases</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td># Changed Phase from e-1</td>
<td>439</td>
<td>193</td>
<td>457</td>
<td>372</td>
</tr>
<tr>
<td>Total pairs changed order</td>
<td>97,017</td>
<td>72,924</td>
<td>96,422</td>
<td>108,293</td>
</tr>
<tr>
<td># Changed Order wrt &lt;200</td>
<td>17</td>
<td>350</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td># Changed Order wrt 200-300</td>
<td>94</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td># Changed Order 300-400</td>
<td>298</td>
<td>19</td>
<td>407</td>
<td>182</td>
</tr>
<tr>
<td># Changed Order wrt &gt;400</td>
<td>134</td>
<td>174</td>
<td>136</td>
<td>361</td>
</tr>
<tr>
<td>Observations</td>
<td>543</td>
<td>543</td>
<td>543</td>
<td>543</td>
</tr>
</tbody>
</table>

5 Results

5.1 Phase Assignment

We first present several pieces of evidence that the order in which constituencies vote is not systematically related to their potential turnout or other political variables. We begin by providing suggestive evidence that our setup does not face one of the fundamental identification problems in the causal analysis of social interaction. Specifically, individual constituencies, in our setup, do not have control over choosing their election order and hence their ‘peer group’ of previous-phase constituencies. First, for each constituency, there is substantial variation in its phase assignment across years. As Figure 2 showed above, constituencies’ phase orders change considerably across elections. To visualize this more clearly, Appendix Figure A.1a plots the phase to phase transition of every constituency across elections. Here the x-axis indexes constituencies and the y-axis plots the difference between assigned phases across the relevant elections: a positive number indicating that the constituency has moved to a later phase, negative indicating an earlier phase, and zero indicating no change. As is evident, very few constituencies keep their order and the movements across elections do not show a systematic pattern. We plot this more formally in terms of histograms in figures A.1b-A.1d. The proportion of constituencies that are assigned the same phase is only around 0.33, and falls further, to around 0.2, if we remove the oddly high rate for the 1998 elections. Our results remain robust to the exclusion of the 1998 elections altogether as well.\(^\text{16}\)

We further explore the phase assignment of constituencies in Table 2, and show that

\(^\text{16}\)Our results are robust to the inclusion of election year fixed effects.
Table 3: Descriptive Statistics by Electoral Phases

<table>
<thead>
<tr>
<th>Phase</th>
<th>1</th>
<th>2</th>
<th>&gt;3</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagged Turnout</td>
<td>0.590</td>
<td>0.589</td>
<td>0.585</td>
<td>0.4184</td>
</tr>
<tr>
<td></td>
<td>(0.099)</td>
<td>(0.118)</td>
<td>(0.120)</td>
<td></td>
</tr>
<tr>
<td>Lagged Turnout Squared</td>
<td>0.358</td>
<td>0.360</td>
<td>0.357</td>
<td>0.5930</td>
</tr>
<tr>
<td></td>
<td>(0.115)</td>
<td>(0.140)</td>
<td>(0.142)</td>
<td></td>
</tr>
<tr>
<td>Lagged Turnout Cubed</td>
<td>0.223</td>
<td>0.228</td>
<td>0.226</td>
<td>0.4398</td>
</tr>
<tr>
<td></td>
<td>(0.104)</td>
<td>(0.131)</td>
<td>(0.133)</td>
<td></td>
</tr>
<tr>
<td>Lagged Margin of Victory</td>
<td>0.117</td>
<td>0.131</td>
<td>0.127</td>
<td>0.0926</td>
</tr>
<tr>
<td></td>
<td>(0.106)</td>
<td>(0.116)</td>
<td>(0.111)</td>
<td></td>
</tr>
<tr>
<td># Candidates</td>
<td>11.6</td>
<td>14.2</td>
<td>14.0</td>
<td>0.3458</td>
</tr>
<tr>
<td></td>
<td>(9.4)</td>
<td>(11.9)</td>
<td>(10.2)</td>
<td></td>
</tr>
<tr>
<td># Male Candidates</td>
<td>11.0</td>
<td>13.7</td>
<td>13.3</td>
<td>0.2003</td>
</tr>
<tr>
<td></td>
<td>(9.1)</td>
<td>(11.6)</td>
<td>(9.9)</td>
<td></td>
</tr>
<tr>
<td># Polling Stations</td>
<td>1388.1</td>
<td>1336.1</td>
<td>1291.0</td>
<td>0.1184</td>
</tr>
<tr>
<td></td>
<td>(340.9)</td>
<td>(271.9)</td>
<td>(296.9)</td>
<td></td>
</tr>
<tr>
<td># States</td>
<td>31</td>
<td>30</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td># Constituencies</td>
<td>389</td>
<td>463</td>
<td>535</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>670</td>
<td>817</td>
<td>1206</td>
<td></td>
</tr>
</tbody>
</table>

constituencies not only change phases with respect to their own phase assignment in previous election years, but also change the order in which they vote with respect to other constituencies. This is important since it ensures that each constituency $i$ gets a different group of peer constituencies every election, which gives us crucial variation to exploit. The second row of Table 2 shows the number of constituencies that, in each election, have been assigned to a different phase with respect to their phase assignment in the previous election. The rest of the rows in Table 2 explore flux in the pairwise ordering of constituencies. Two constituencies, A and B, can vote in one of the following three orders in each election: A votes before B, they vote in the same phase, or A votes after B. The third row documents the number of pairs of constituencies that changed orders in this sense, compared to their respective ordering in the previous election. Rows 4 through 7 finally report the numbers of constituencies that changed orders with respect to different numbers of other constituencies. Overall this table establishes that we observe substantial churning in the order of constituencies, which gives us the required variation in the composition of the peer group of a given constituency going to polls in a given phase.

So far we have established that there is considerable variation in phase assignments
across elections for individual constituencies as well as for pairs of constituencies. Never-
theless, there might still be concerns that the Election Commission of India (ECI) itself is
using predetermined political variables to determine phase orderings. The validity of our
instrument hinges on constituencies not being systematically assigned to phases based on
their potential turnout. We explore this by considering constituency level electoral variables
and examining whether they vary significantly with phase assignment. Table 3 presents
statistics for key variables at the constituency level by electoral phase, aggregating across
all elections. Perhaps the closest proxy in our data of a constituency’s potential turnout is
its turnout in the previous election (see for instance, Fujiwara et al. (2016)). The lagged
turnout (i.e. the turnout of the same constituency in the previous election year) is almost
identical across phases, as are its higher order terms, suggesting that phase assignment is
not only uncorrelated with but also likely to independent of it. We also see that a similar
pattern is observed for margin of victory in the previous election, number of candidates that
contest from each constituency, their gender composition, as well as for the total number
of polling stations established in each constituency. To summarize this information statisti-
cally, we regress each variable on phase indicators and conduct an F-test for their joint
significance. The fourth column in Table 3 reports this p-value. All the p-values are well
above the conventional significance levels, except for the lagged margin of victory which is
significant only at the 10% level.\footnote{Our results are robust to controlling for margin of victory.}

5.2 Baseline Results

Table 4 presents our baseline results. The unit of analysis is a constituency-election observa-
tion with the outcome variable being the turnout in constituency \(i\) that went to the polls in
phase \(p\) in election \(e\). Our explanatory variable of interest is the average turnout across all
constituencies that were assigned to phase \(p-1\) in the current election \(e\). Column (1) shows
a negative and significant effect of higher average turnout in the previous phase on turnout in
the current phase. We control for historical propensity to vote and the gap between phases
in Column (2), where the negative effect attenuates, and for phase fixed effects in Column
(3) where it is stronger However, as outlined in the previous section, OLS is likely to be
biased due to the existence of unobservables that can affect turnout of constituencies in both
phase \(p\) and \(p-1\). Columns (4) to (6) present our IV analysis. As can be seen from Table
4, our first stage has high predictive power and is very precisely estimated. We now see a
substantially stronger effect compared to OLS, especially in our fully controlled specification
in column (6), with a 1 percentage point (pp) increase in average turnout in the previous
Table 4: Effect of Voter Turnout in Previous Phase

<table>
<thead>
<tr>
<th>Turnout$_{iep}$</th>
<th>OLS</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Mean\ Turnout_{ep-1}$</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$Mean\ Turnout_{i77-89}$</td>
<td>$-0.100^*$</td>
<td>$-0.016$</td>
</tr>
<tr>
<td></td>
<td>$(0.045)$</td>
<td>$(0.039)$</td>
</tr>
<tr>
<td></td>
<td>[0.448]</td>
<td>[0.823]</td>
</tr>
<tr>
<td>Phase Fixed Effect</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$First\ Stage$</td>
<td>$-0.847^{**}$</td>
<td>$0.835^{**}$</td>
</tr>
<tr>
<td></td>
<td>$(0.056)$</td>
<td>$(0.056)$</td>
</tr>
<tr>
<td>Observations</td>
<td>2,023</td>
<td>2,010</td>
</tr>
</tbody>
</table>

**, * indicate significance at the 1%, and 5% level, respectively. Columns (1) and (4) report results with no controls. Columns (2) and (5) control for the gap in days between phases, and Columns (3) and (6) additionally control for a phase level fixed effect. Standard errors in parentheses are clustered at the constituency level. Square brackets report p-values from a wild cluster bootstrap routine clustered at the state level.

An important point to make here is the comparison between the OLS and IV estimates. The OLS estimates in table 4 are likely biased towards zero, which is why instrumenting yields stronger negative effect sizes. This makes intuitive sense given that the former do not deal with the bias induced by correlated effects in constituency specific unobservables, which

18These results are robust to dropping the 1984 election (which had an unusually high turnout in favor of the INC following the assassination of Prime Minister Indira Gandhi) while constructing the historical turnout.

19In addition to just clustering at the State level, we also check the robustness of our results by clustering at the State*election level. Given that constituencies in the same state sometimes go to polls in different phases, one can expect correlation in error terms within State in a given election year. Results are robust to the alternative way of statistical inference as well.
is likely to lead to a positive spurious relationship between turnouts across phases. This
would be true as long as two randomly chosen constituencies, say in the first two phases,
respond similarly to a pre-election shock, all else constant. It is worth noting also that the
OLS estimates themselves are negative, while any potential bias arising from election-specific
correlated effects are likely to bias them upwards. The instrumental variable therefore simply
yields more precise estimates for the magnitude of the effect, but reinforces the idea that the
overall effect of a higher early turnout is to reduce the later voters’ incentives to vote, social
interactions notwithstanding.

5.3 Falsification Tests

Our estimation strategy argues that we are identifying the effect of turnout in the past
phase, which is observed well before the constituencies in the current phase go to the polls,
on turnout in the current phase. If this is indeed the correct causal mechanism, then our setup
suggests two natural falsification tests to check the soundness of our identification strategy.
First, turnout in constituencies that go to polls contemporaneously should not affect each
others’ turnout. Second, turnout in future phases should not impact turnout in the current
phase. Under both these scenarios, voters in a particular constituency would not observe
the relevant turnout statistics at the time they are making their turnout decision. This is
the basis of the falsification tests for our specification, which we now present. In particular,
these falsification tests will help ascertain whether the instrument used is adequately able
to deal with correlated effects and any potential violation of the quasi random sorting of
constituencies to electoral phases. If either of these effects are manifesting themselves in
our estimated coefficient, then we should see similar results for both contemporaneous and
future phases. Fortunately, as we explain next, we do not face this problem.

To test for the above mentioned effects, we estimate similar specifications using con-
temporaneous turnout and future turnout (in separate regressions) as the main explanatory
variable. Table 5 reports these results: columns (1) to (3) correspond to our OLS specifi-
cation where we expect the falsifications to fail, while columns (4) to (6) correspond to the
instrumental variable specification. For both specifications, in order to facilitate comparison,
we report the results for our original variable of interest from the fully controlled specifica-
tion from table 4, in the first column. The second columns report the coefficients for average
turnout of constituencies, excluding $i$, in the same (current) phase, while the third columns
report the coefficients for the next (future) phase. The second specification is analogous to
a classical peer-effects model, where $i$ is part of its own peer group.

Column (2) shows an estimated effect which switches sign and is statistically significant
Table 5: Effect of Voter Turnout in Previous Phase - Falsification Tests

<table>
<thead>
<tr>
<th>Turnout_{iep}</th>
<th>OLS (1)</th>
<th>OLS (2)</th>
<th>OLS (3)</th>
<th>OLS (4)</th>
<th>OLS (5)</th>
<th>OLS (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Turnout_{iep-1}</td>
<td>−0.132**</td>
<td>−0.132**</td>
<td>−0.375**</td>
<td>−0.375**</td>
<td>−0.375**</td>
<td>−0.375**</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.072)</td>
<td>(0.050)</td>
<td>(0.050)</td>
<td>(0.050)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>Mean Turnout_{~iep}</td>
<td>0.565**</td>
<td>0.565**</td>
<td>0.134</td>
<td>0.134</td>
<td>0.134</td>
<td>0.134</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.078)</td>
<td>(0.078)</td>
<td>(0.078)</td>
<td>(0.078)</td>
<td>(0.078)</td>
</tr>
<tr>
<td>Mean Turnout_{iep+1}</td>
<td>0.266**</td>
<td>0.266**</td>
<td>0.266**</td>
<td>0.266**</td>
<td>0.266**</td>
<td>0.266**</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.061)</td>
<td>(0.061)</td>
<td>(0.061)</td>
<td>(0.061)</td>
<td>(0.061)</td>
</tr>
<tr>
<td>Mean Turnout_{77-89}</td>
<td>0.835**</td>
<td>0.795*</td>
<td>0.782**</td>
<td>0.828**</td>
<td>0.829**</td>
<td>0.781**</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.055)</td>
<td>(0.057)</td>
<td>(0.057)</td>
<td>(0.057)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>First Stage</td>
<td>0.588**</td>
<td>0.905**</td>
<td>0.987**</td>
<td>0.987**</td>
<td>0.987**</td>
<td>0.987**</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.020)</td>
<td>(0.039)</td>
<td>(0.039)</td>
<td>(0.039)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,010</td>
<td>2,010</td>
<td>1,190</td>
<td>2,010</td>
<td>2,010</td>
<td>1,190</td>
</tr>
</tbody>
</table>

**, * indicate significance at the 1%, and 5% level, respectively. All regressions control for the gap in days between phases and a phase level fixed effect. Standard errors in parentheses are clustered at the constituency level. Square brackets report p-values from a wild cluster bootstrap routine clustered at the state level.
for contemporaneous constituencies, as one would expect, given correlated effects across constituencies within the same phase. However, the IV result from column (5) presents a more promising picture, with the point estimate falling towards zero and becoming statistically insignificant. Similarly, column (3) presents the Granger-causality motivated test for the OLS specification and the uncorrected estimate is significant at the 5% level. However, once we instrument, the point estimate falls very close to zero and is statistically insignificant as well. Overall, both these tests provide compelling evidence that our empirical strategy is indeed capturing the effect of turnout in phase \( p - 1 \) on constituencies that go to polls in phase \( p \), as opposed to some spurious relationship between turnouts across constituencies and phases.

### 5.4 An Alternative Way of Defining Turnout

The layman understands that, in a plurality voting system, there are typically only two candidates with a realistic chance of winning (this phenomenon is sometimes known as Duverger’s Law). Thus, a voter who cares about the policy outcome post-election, and who is informed about the competitive milieu, will choose her favorite of these two candidates, even if neither of them are her favorite among all candidates. Those who vote for noncompetitive candidates clearly have other motivations, from the desire to signal support for less-powerful politicians to the simple desire to socialize. Clearly, the behavior of this latter group will be less sensitive to changes in the competitive environment and so their presence might introduce noise into our analysis. Thus, we partition voters accordingly: those who voted for one of the top-two vote-getting candidates and those who did not. The latter group also includes invalid ballots. Surely there is a difference between voters who cast invalid ballots, voters who cast a protest ballot, and voters who simply vote for their favorite candidate, but this behavior falls outside the scope of this paper.\(^{20}\) Nevertheless, the following analysis is unchanged by the exclusion of rejected ballots.

Table 6, Panel A, reports the results for the proportion of the electorate who voted for either of the top two candidates, while Panel B reports the results for the proportion casting non-top-two votes. For brevity, we only report the estimates for the main variable of interest. As expected, the response to past turnout on current turnout is stronger among top-two voters, with the magnitude of the coefficient increasing to -0.66 compared to -0.38 in the baseline instrumented results. This provides suggestive evidence that more informed

\(^{20}\)We also analyze how voters respond to prior turnout by different political ideologies, for a more nuanced understanding of variations in voting behavior by political affiliations of voters, but due to the small sample size and the large number of regional parties in India, results are very noisy and a clear picture does not emerge.
Table 6: Effect of Voter Turnout in Previous Phase - Alternative Turnout Definitions

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A: Top Two Candidates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Turnout ep−1</td>
<td>−0.291∗∗</td>
<td>−0.177∗∗</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.047)</td>
</tr>
<tr>
<td><strong>B: Non-Top Two Candidates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Turnout ep−1</td>
<td>0.191∗∗</td>
<td>0.161∗∗</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.037)</td>
</tr>
</tbody>
</table>

Observations 2,023 2,010 2,010 2,023 2,010 2,010

** and * indicate significance at the 1%, and 5% level, respectively. Columns (1) and (4) report results with no controls. Columns (2) and (5) control for own historical turnout and the gap in days between phases, and Columns (3) and (6) additionally control for a phase level fixed effect. Standard errors in parentheses are clustered at the constituency level. Square brackets report p-values from a wild cluster bootstrap routine clustered at the state level.

On the other hand, Panel B, provides some evidence for social pressure based channels in operation among a subset of the electorate. In contrast to the previous result, the non-top-two voters are more likely to vote in response to a higher turnout in the previous phase. The magnitudes are slightly smaller in size but are precisely estimated, with a 1 pp point increase in average turnout in the past phase leading to a 0.29 pp increase in turnout of voters voting for non-top-two candidates in the subsequent phase. One might think that the proportion of these votes might increase simply because a higher turnout in the previous phase implies that the electorate is enthusiastic and so will turn out in the subsequent phase. This suggests to the voter that his vote will be less important in his own constituency, so he may be less strategic. An obvious response to this, however, would be to not vote at all. So we believe that restricting the turnout to voters casting votes for candidates unlikely to win, or votes that are rejected, gives us a reasonable measure of voters who indeed responded to the festivity effect of the election, turning out to vote just for the sake of participation in the event.21

Furthermore, there is evidence in the literature that lower voter costs can

We also try to identify if there is a stronger effect corresponding to a higher turnout in one of the more ‘important’ constituencies, such as the capital constituencies of the different states. We find that there is a strong significant effect on the turnout in constituencies in phase p if there is a higher turnout in the capital constituency of the same state which was assigned to phase p − 1. We also find a significant positive effect, though smaller in magnitude, corresponding to a higher turnout in any key constituency (not necessarily from the same state) that went to the polls in phase p − 1.

21
Table 7: Effect of Voter Turnout in Previous Phase by Geographic Distance

<table>
<thead>
<tr>
<th></th>
<th>Neighboring States</th>
<th>Non-Neighboring States</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Turnout_{iep} )</td>
<td>OLS</td>
<td>IV</td>
</tr>
<tr>
<td>( Mean\ Turnout_{ep-1} )</td>
<td>0.507**</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.081)</td>
</tr>
<tr>
<td></td>
<td>[0.008]</td>
<td>[0.919]</td>
</tr>
<tr>
<td>( Mean\ Turnout_{177-89} )</td>
<td>0.758**</td>
<td>0.883**</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,737</td>
<td>1,737</td>
</tr>
</tbody>
</table>

** and * indicate significance at the 1% and 5% level, respectively. All regressions control for the gap between phases and a phase level fixed effect. Standard errors in parentheses are clustered at the constituency level. Square brackets report p-values from a wild cluster bootstrap routine clustered at the state level.

bring more uninformed voters to the ballots on election day (Hodler et al. (2015)), and these voters, in turn, would be more likely to vote for non-competitive candidates as well.

5.5 Neighboring vs. Non-Neighboring States

Finally, we present some results for whether voters respond differently to high turnout in places about which they might have different kinds of prior beliefs. Note that our instrument exploits variation arising from the differences in the historical propensity of the voters in the early-phase constituencies to turn up; a high early turnout in this context simply means that the constituencies which are historically inclined to vote in high numbers have been assigned to the early phases. Thus the IV works through the divergence in the priors of the later-phase voters regarding how high early turnout should be, and how high it actually is. So, we should expect a stronger response to a high turnout in places about which the voters have a weaker prior, so that they revise their expectations regarding the number of people likely to turn up in their own phase, compared to places where the voters are familiar with their historical turnout rates and the high numbers don’t appear a surprise. We provide suggestive evidence for this using state-wise geographical proximity: it may be reasonable to suppose that voters have a more accurate prior for the expected turnout in constituencies which are in the same state as theirs or in an adjacent state, compared to constituencies which are in non-neighboring states, and therefore we should observe stronger results if there’s a high turnout in non-neighboring states, which would constitute a larger positive shock to expected turnout. Table 7 documents these results. Indeed, we see a robust negative effect of a high turnout in non-neighboring state constituencies that were assigned
to vote in the earlier phase, while the effect of corresponding constituencies in the same state or in neighboring states is statistically indistinguishable from zero.

6 Discussion

We find compelling models in Feddersen and Sandroni (2006) and Coate and Conlin (2004) that can help shed light on potential mechanisms to explain our findings. Each voter, in their setting, acts as a social planner, prescribing rules of behavior for society at large that maximize what he or she perceives as social welfare, and then following the rules they prescribe. All voters measure social welfare as the social utility of a policy minus the total social costs required to achieve that policy, which in this case is the cost of voting. Not all voters agree on the social utility of various policies, but they share a common belief about the distribution of voting costs in society.

The Indian election, as mentioned, is largely a competition between two coalitions. Accordingly, we think of voters being partitioned into two camps, each camp in agreement on the optimal policy, and so reasoning identically about social welfare. Each voter deduces the following rule: “all agents in my camp whose cost of voting does not exceed a given threshold ought to vote.” This threshold of course depends on prior beliefs about the competitiveness of the election and the distribution of voting costs. The result is a kind of “equilibrium” in which each voter sets their own threshold to maximize the expected probability of their coalition winning, minus the expected cost to society of voters going to poll. Some basic randomness in the turnout of agents ensures that both camps set non-zero thresholds: even if a one knows that it is in the minority, it may rely on this randomness to steal a win if its turnout threshold is sufficiently high. In sum, all voters of a given camp take the cost threshold of the opposite as given and choose their own threshold to maximize social welfare.

Based on this theoretical setting two potential mechanisms emerge that can explain why past turnout might influence current turnout.

The Cost Pathway: Agents view turnout as a signal of the distribution of voting costs. These costs can come from several sources, such as distance and waiting time, and may be lessened by preferential factors, such as fervor or social interactions. If turnout is high, the voter thinks that the distribution of costs is skewed more toward zero implying a lower best response threshold. The voter, however, knows their own cost and may find themselves on the margin: prior to observing the turnout in the previous phase she planned to go to polls, but having adjusted the cost threshold she anticipates that other voters in her camp will turn out on polling day and she doesn’t have to incur her voting cost. This can be thought of as a sort of by-stander effect.
The Competitiveness Pathway: Agents view turnout as a signal of the true tightness of the race between two coalitions. Note that the true tightness is a measure of underlying preferences and does not necessarily correlate with the margin of victory observed since, to minimize costs, voters would choose to win by smaller margins if possible. In fact, when the underlying preferences indicate a tight race, turnout is high as each side thinks randomness might give them the win. Thus, high turnout yesterday might induce high turnout tomorrow.

In our setting it is difficult to empirically differentiate between the above two mechanisms. However, the negative effect of turnout in the previous phase on current turnout that we estimate can potentially be explained by the cost pathway. If this is indeed the case then this staggered nature of polling in the Indian setup is unlikely to introduce any asymmetries in our setup, as it suggests the agents are not taking turnout as a signal of the competitive milieu. This is plausible since the only publicly available information at the end of each electoral phase is the overall turnout, not the party specific break down.

7 Conclusion

Voting behavior is rich and driven by complex strategic decisions. We have studied a previously unexploited source of variation arising from the staggered nature of the Indian General Elections where voting occurs in phases, to dig deeper into voting behavior. We use the quasi-random allocation of constituencies to electoral phases in India to avoid endogenous sorting into phases coupled with an instrumental variable strategy to deal with potential correlated effects across phases. Results show that a 1 percentage point (pp) increase in average turnout in the previous phase leads to a reduction of around 0.38 pp in turnout in the current phase. This signifies that voters are in fact learning from observing turnouts in past phases and thus forming their decision on whether to turn out on polling day.
References


URL http://www.jstor.org/stable/30034339


Patnam, M., 2013. Learning from exit polls in sequential elections.

A Appendix

Figure A.1: Phase to Phase Constituency Transitions Across Elections

(a) $\text{Phase}_e$ by $\text{Phase}_{e-1}$
(b) Changes in Phases across Elections - Overall
(c) Changes in Phases across Elections - By Year
(d) Changes in Phases across Elections - ~1998