Candidate Selection by Parties: Crime and Politics in India*

Arvind Magesan
Economics Department, University of Calgary

Andrea Szabó
Economics Department, University of Houston

Gergely Ujhelyi
Economics Department, University of Houston

June 21, 2024

Abstract

We study how parties choose candidates, a key issue to understand political selection and ultimately policy choices. Do parties select candidates that voters like, or are their choices shaped by other considerations? What is the impact of policies that limit parties' choice sets, such as restrictions on candidates with a criminal history? To study these questions, we combine rich candidate-level data from India with a model in which parties trade off the electoral appeal of candidates against internal party preferences in a strategic game of candidate selection. We find that parties' preferences systematically deviate from voters'. While parties select candidates who are likely to win, all else equal they prefer those who are not overly popular. Selection decisions are also driven by strategic considerations, as well as factors that are independent of voter preferences, such as the ease of recruiting certain candidates. Our estimates provide a nuanced explanation for parties' motivation to run criminal candidates, and, through counterfactual simulations, shed light on the potential impacts of banning criminals from contesting elections.

*For useful comments and suggestions we thank Victor Aguirregabiria, Brad Hackinen, Jean-François Houde, Robert Miller, Paul Novosad, Amil Petrin, Francesco Trebbi, and participants at the 2023 Canadian Public Economics Group Meetings, the 2024 APSA Virtual Research Meeting and the 7th Banff Empirical Microeconomics Conference. Magesan gratefully acknowledges funding from SSHRC through IG grant 435-2019-0437.

1 Introduction

Do parties select candidates that voters like, or are their choices shaped by other considerations? In a representative democracy, elections aggregate voter preferences over candidates *who appear on the ballot*. Therefore the question of whom parties choose to run is key to understand political selection, and ultimately policy choices.

Although a sizeable literature now explores individuals' own decision to enter politics (Besley 2005; Dal Bó and Finan 2018), we know much less about how parties select their candidates. As Dal Bó and Finan (2018, p.566) write: "political parties likely play a major role in who becomes a politician, and yet we have a very limited understanding of how political parties recruit and screen their candidates."

Clearly, parties like to win, which requires running candidates that voters will support. At the same time, there are several reasons why a party may not necessarily want to run the most popular candidate.

Party elites may value candidate traits like loyalty, influence or wealth even if these do not lead to more votes. For example, wealthy candidates can bring resources to the party or self-finance their campaign. Similarly, candidates from influential groups can provide valuable connections and access for party leaders. Such benefits have been proposed as an explanation for the widespread nomination of (wealthy and well-connected) criminal politicians in India (Vaishnav 2017).

Sometimes a candidate who enjoys "too much" voter support may in fact threaten the position of party elites, or steer the party's policies in a direction that is inconsistent with the elite's preferences. In the US, these concerns are well illustrated by changes in the Republican party since the 2016 electoral campaign of Donald Trump. In India, where parties are highly centralized, candidates who become too popular can also give rise to internal power struggles and weaken the party leader (Chandra 2016).[1]

In this paper, we study candidate selection in Indian national elections, combining a discrete-choice model of voter preferences over candidates with a strategic game of candidate selection between the two main party alliances. Estimating the model quantifies several factors that cause party choices to systematically deviate from the maximization of winning probabilities (or vote shares). We use our framework to model the impact of policies that

[1]This idea is consistent with existing theories on the internal organization of parties. In Caillaud and Tirole (2002), parties ensure that low-quality candidates stand down and support a high-quality nominee by sharing the "spoils" from winning. It is not difficult to imagine that a candidate who wins with a landslide will be less likely to share. In Mattozzi and Merlo (2015), party members exert effort that benefits the party and increases their chances of obtaining the nomination. "Superstar" candidates reduce others' incentives to provide effort through a discouragement effect. In both of these cases, running highly popular candidates is costly for the party.

affect the set of candidates available to parties, in particular the banning of candidates with criminal backgrounds.

Our primary source of data is the Election Commission of India (ECI). We focus on the 2009 and 2014 national elections, using additional information from state elections held between 2008-2017. The dataset contains vote shares, number of eligible voters, and candidate characteristics. We include information on candidate wealth, education, and criminal backgrounds from affidavits that candidates are required to file with the Election Commission, and create an indicator for candidates with Muslim names.[2] We match to this data constituency characteristics from the Indian Census using village level information from the SHRUG database (Asher et al. 2020).

We begin by estimating a comprehensive BLP discrete-choice demand system describing voter preferences among candidates in Indian national elections (Berry et al. 1995). Voters have preferences over several (endogenous) candidate characteristics, such as education, wealth, whether the candidate has a criminal background, and whether the candidate is Muslim. These preferences are also shaped by constituency characteristics and unobserved candidate characteristics.

Our specification of the "supply side" focuses on candidate choices by the two main party alliances in national politics, the NDA and the UPA, led by the two largest parties, the BJP and the INC, respectively. We use a simultaneous game of incomplete information to model the strategic interaction between these players, and estimate their preferences over candidates using Aguirregabiria and Mira (2007)'s Nested Pseudo-Likelihood procedure.

To make this game and its estimation feasible, we assume that parties' choice sets are comprised of clusters of candidate characteristics, or candidate "types." We approximate the pool of potential candidate types that parties select from in the national election using the set of candidates contesting elections to Indian state legislatures.[3] We use a machine learning algorithm to identify candidate types in this state election data - essentially, these are combinations of candidate characteristics that tend to occur together in the data.[4] This

[2]Muslims are a salient group in Indian politics, but this characteristic is rarely used in academic research due to a lack of data. We create an indicator for Muslim candidates based on their names using methods from text analysis, assigning candidates to different groups based on the "distance" of their name from libraries of Muslim and non-Muslim names. Throughout we use "Muslim" to describe ethnicity, proxied by name, rather than religion.

[3]A large fraction (almost 25% in our sample years) of candidates for state elections are independent candidates - many of these wanted to run under the banner of a major party but were not selected. In addition, many national level candidates begin their political careers at more local levels of politics (Dar 2019) further supporting the idea that candidates for state election approximate the pool from which national candidates are drawn.

[4]This approach is similar in spirit to Bandiera et al. (2020) who reduce high dimensional data on CEO activities to a small set of CEO types in order to study how CEO behavior affects firm performance. Hamilton et al. (2021) use a clustering approach to reduce the choice set of patients choosing between different medical

clustering algorithm identifies four candidate types, all of which turn out to have clear interpretations: an "educated type" (educated non-Muslim with no criminal history), an "uneducated type" (uneducated non-Muslim with no criminal history), a "Muslim type," and a "criminal type" (non-Muslim with a criminal history who is also relatively wealthy).

Using our demand estimates, we construct counterfactual vote shares corresponding to all combinations of candidate types potentially chosen by the competing parties. This allows us to compute expected vote shares and win probabilities. Parties' objective functions nest these expected vote shares and win probabilities, as well as a set of heterogeneous costs of running different candidates.

Estimating parties' objective functions reveals that while parties prefer candidate types with better chances of winning, all else equal, they prefer candidates with a lower expected vote share. In other words, parties prefer to win with less popular candidates. This finding is consistent with the idea that party elites trade off anticipated electoral performance against the threat of a candidate becoming too powerful and undermining the leadership or its policies, as suggested by the literature on Indian electoral politics (see Section 3.1).

We also find that considerations that are independent of voter preferences (and hence win probabilities or vote shares) matter in parties' objective functions. All else equal, candidate types that are more common in the relevant local candidate pool are less costly for parties to recruit and run. Parties also have direct preferences over candidate types: for example, the NDA has a particularly large direct cost of running a Muslim type, which is in line with the BJP's declared Hindu nationalist profile. Interestingly, we find that both parties obtain a positive direct payoff from running criminal types relative to others. This is consistent with Vaishnav (2017), where criminals bring benefits to the party in the form of (organized crime) networks and an ability to finance their own campaigns.

The pattern of estimated payoff parameters has a striking equilibrium implication: according to our results, criminal candidates are often strategic complements. In other words, parties are often compelled to run a criminal candidate because the other party is running a criminal as well. Intuitively, this happens when, faced with an opposing criminal candidate, the party's only hope for a win is to also select a criminal candidate. From the parties' joint perspective, running criminals may thus be inefficient.

Motivated by this last observation as well as recent policy proposals in India and around the world, we use our framework to model the impact of a ban on criminal candidates.[5] One

treatments.

[5]Restrictions on candidates with a criminal history is also a salient issue outside of India. In Brazil, where 40% of candidates running for governor in 2014 had pending court cases against them, the 2014 Clean Record Law banned convicted criminals from running for a period of 8 years (Paiva et al. 2014). In the US, states vary in whether they allow convicted criminals to run for office. In Louisiana and Maine, convicted felons

implication of the ban is to change the distribution of candidate types contesting elections, leading to higher fractions of educated, uneducated, and Muslim types. According to our results, the ban also has implications for parties' expected winning probabilities. We find that, with a criminal ban in effect, the vote share of third party candidates rises, lowering the winning probability of both major parties. It appears that some voters' preference for the major parties is conditional on these parties' ability to run criminal types.

In spite of this, we find that in 34% of constituencies *both* parties benefit from a criminal ban. The reason for this is the strategic complementarity of criminal candidates. Once their opponent is banned from running a criminal, each party finds other candidate types more attractive. Thus, banning criminals allows parties to compete with candidates who deliver higher payoffs in equilibrium. This provides an explanation for why political parties could support a ban on criminal candidates while at the same time running such candidates in elections. More generally, our results demonstrate how candidate eligibility regulations interact with determinants of party preferences, including the internal organization of parties.

2 Related literature

Our paper contributes to a growing literature that studies the role of parties in political selection. We consider an important setting, India, where the previous literature on this issue has been largely descriptive. Our structural approach provides a methodological contribution, and allows us to ask several new questions.

In the political selection literature, several papers analyze the tradeoffs that parties make between electability and other considerations, and how this impacts the "quality" of candidates. For example, Galasso and Nannicini (2011) present a model where high quality candidates are valued by swing voters but are expensive for parties to recruit. This tradeoff leads to parties running high quality candidates in more competitive districts, which is consistent with data from Italy. Besley et al. (2017) study a model where running the best candidate would jeopardize party leaders' survival, and find evidence consistent with their predictions in Sweden.[6]

In this set of papers, candidate "quality" is measured in terms of education or residuals from a Mincerian wage regression. By contrast, our approach makes it possible to study selection on multiple dimensions simultaneously. We do not take an a priori stance on what constitutes a high quality candidate, and we also let the data tell us what the relevant types

are eligible to run; in Massachusetts, they are ineligible to run while incarcerated, while in Texas they are ineligible even after the completion of their sentence. See https://restoration.ccresourcecenter.org.

[6]See also Casas-Arce and Saiz (2015), who show evidence that, prior to the introduction of a gender quota, Spanish parties failed to maximize votes by not running more female candidates.

of candidates are in parties' choice sets.[7]

In the developing country context, two recent papers conduct field experiments that create changes in the candidate nomination process to study party leaders' behavior. Gulzar et al. (2021) provide party leaders in Nepal with information on potential candidates, including information on their party service, competence, and popularity among voters. Casey et al. (2021) survey voters in Sierra Leone about their preferred candidate, and then randomize an intervention where this information is shared with party leaders, and potential candidates present their qualifications and debate each-other in a public forum. Both papers find that providing information resulted in parties fielding more candidates preferred by voters, which is consistent with a lack of information about voter preferences at baseline.

Our approach is complementary to this strand of the literature. First, we quantify several different mechanisms behind candidate selection, including parties' payoff from winning, vote shares, direct costs, and recruitment costs. Second, neither experiment was designed to study parties' strategic behavior in the candidate selection process, something we explicitly account for. Third, while both experimental studies identify the average treatment effect of the intervention, neither of them was designed to measure the weights of different objectives in parties' candidate selection procedures. Our structural approach makes it possible to answer questions regarding *counterfactual* policy interventions, including interventions in different settings.[8] Fourth, creating experimental changes in real-world nomination procedures must necessarily be limited due to ethical considerations (see Appendix 3 of Gulzar et al. (2021) for a detailed discussion), which creates issues for the replicability of these studies in other time periods or other countries. Our study does not face this limitation, as our methodology can easily be applied to other contexts.

Conceptually, the paper closest to ours is Iaryczower et al. (2024), who present a structural analysis of candidates' policy positions (ideology) in Brazil. In their framework, candidates are exogenously given, and choose ideological positions constrained by other candidates' ideology from the same party. By contrast, we focus on the game played by parties as they select their candidates. In doing so, we take candidate ideology as exogenously given, which is appropriate in a setting where the central party organization sets the policy platform and

[7]Like us, Dal Bó et al. (2017) also considers candidate selection on multiple dimensions in their comprehensive study of Swedish politicians. While their focus is on *self*-selection, they show some evidence that parties are more willing to promote individuals who are competent (as measured by cognitive and leadership abilities) independently of their background. This is consistent with the idea that in Sweden socioeconomic and ethnic background likely play a considerably smaller role in politics than in India or in developing democracies more generally.

[8]For example, in Section 7.3 we will use our model to draw policy conclusions in a setting where party preferences take different values from what we estimate. The possibility of this type of inference adds to the external validity of our findings, and is one of the advantages of a model-based approach.

the elected candidates follow it, as is the case for the two major Indian parties we study. Candidate selection is a discrete game that results in endogenous candidate characteristics over multiple dimensions, and we use estimation methods that account for these features in the analysis.

3 Background and data

3.1 Background

3.1.1 Parties and elections

We study general elections to India's national legislature (the *Lok Sabha*), a near ideal setting for analyzing strategic candidate selection by political parties.

The Lok Sabha is comprised of a large number of single member districts (called *constituencies*). In each election, voters in a given constituency elect one and only one representative from the available choices on the ballot. In contrast to a setting with proportional representation, each competing party selects a single candidate to run in the constituency.

India is a multiparty system with over 10 candidates contesting the average constituency. In most cases, two of these candidates represent the two main competing (pre-election) alliances, the United Progressive Alliance (UPA) and the National Democratic Alliance (NDA), led by the two main national parties, the INC and the BJP, respectively. These alliances run candidates in almost all constituencies in each election and together win the majority of seats. In every constituency, the alliance contains a group of parties (and occasionally some independent candidates) that enter into a pre-election agreement about which candidate will run to represent the alliance, without competition from other members of the group. Because our model will treat alliances as the players, we will refer to the two alliances throughout simply as *parties*.

Although we study parties' choices in national elections, we will also make use of data from state elections (elections to states' legislative assemblies). These are separate elections, and in most cases are held in different years from the national election. The constituencies in the two elections are different, as each national constituency is subdivided into several state constituencies. The set of parties competing in national and state elections can also be different (state elections tend to be contested by a large number of regional parties), but the UPA and the NDA are major forces in state elections as well.

3.1.2 Party goals and procedures for candidate selection

Indian parties are known for their centralized organizations in which a central committee, or in some cases a charismatic leader, dictates all major decisions, including candidate selection. Farooqui and Sridharan (2014) review nomination procedures used in different countries, noting that "the USA represents the decentralised extreme, that of party primaries" while "India lies near the other extreme in that most of its major parties are at the completely or near-completely top-down of the six types of party nomination processes, with the national party leadership having the final say." (p.80) Although both the INC and the BJP have formal consultation procedures that involve local party organizations in the candidate selection process, in practice decisions are ultimately made by each party's central committee (Roy 1966; Farooqui and Sridharan 2014).

Apart from electability, important factors in the candidate selection process include loyalty to the party leadership and service to the party organization. These considerations are often explicit in parties' written procedures on candidate selection (Roy 1966). For example, the INC has declared: "Winnability alone should not be the benchmark for deciding nominees of the party during elections. There should be a balance required between loyalty and winnability." (Jaipur Declaration of the AICC, 2013, quoted in Chandra (2016, p.41)). The main source of this tradeoff is threats to the party from factions that undermine its leadership, or from candidates who defect to other parties (Chhibber et al. 2014).[9] Chandra (2016) summarizes this tradeoff as follows:

> "The central leadership's decisions are influenced not only by the anticipated electoral performance of the candidate, but by two intra-party considerations: (1) to ensure the compliance of powerful factions within the party so that they work for the party candidate or at a minimum do not work against the candidate and do not defect from the party and (2) to undercut factions that are becoming too powerful so that they do not threaten the leadership or its loyalists. Expectations of the electoral performance of the candidate are often subordinated to these two criteria." (p.224)

India is rife with examples of successful candidates defecting and imposing costs on their original party. One particularly striking example is the case of Sarekoppa Bangarappa, who started as a state-level politician in Karnataka but eventually rose to represent the constituency of Shimoga as an elected MP for the BJP in 2004, winning the election with over 50% of the votes. (It is notable that Bangarappa was facing two criminal charges in

[9] The leadership's desire to undercut internal rivals also explains why Indian parties often run newcomers instead of incumbent candidates who may be perceived as becoming too powerful (Lee 2020).

this period, making him a "criminal type" in our framework below.) The following year, while a sitting member of parliament with the BJP, Bangarappa defected to another party (the Samadjwadi Party), forcing a by-election in the constituency, which he won, costing the BJP its seat in the Lok Sabha.

Another important factor influencing parties' candidate selection is financial considerations: financial contributions to the party and a candidate's ability to finance their own campaign. Farooqui and Sridharan (2014, p.87) describe, in the case of the BSP party, the process through which candidates effectively bid to receive the nomination. Similarly, Vaishnav (2017) argues that the main appeal of criminal politicians to Indian parties stems from the fact that these individuals can finance their own campaigns, including by breaking campaign finance laws if necessary.

3.1.3 Rules on criminal politicians

As we shall see, our results confirm the importance of candidates with criminal backgrounds for Indian parties. Based on the 1951 The Representation of the People Act, criminally convicted politicians are eligible to run for election six years after the completion of their sentence. In practice, many indicted politicians also run, and win, while undergoing lengthy trials. In 2019, 43% of candidates elected to the national legislature had been indicted on criminal charges at some point.[10]

Over the last several decades multiple commissions tasked with electoral reform in India have recommended tightening the restrictions on criminal candidates.[11]

Perhaps surprisingly, the two main parties themselves have at different times expressed a desire to ban criminal candidates - even while actively running such candidates in elections.[12] Our results below will help rationalize this phenomenon.

[10] https://adrindia.org/content/nearly-50-cent-mps-new-lok-sabha-have-criminal-records

[11] For example, in 2004 the Election Commission unsuccessfully recommended that indicted politicians be banned from elections (https://prsindia.org/files/bills_acts/bills_parliament/2008/bill200_20081202200_Election_Commission_Proposed_Electoral_Reforms.pdf). In 2013, in Lily Thomas v Union of India, the Indian Supreme Court closed a loophole allowing some politicians to run while appealing their convictions.

[12] In a 2010 speech Sonia Gandhi (then chief executive of the INC) discussed the "need to build a consensus on how to prevent individuals with a criminal record from contesting elections." This sentiment was echoed by members of the rival BJP, including then leader of the opposition Sushma Swaraj (https://economictimes.indiatimes.com/news/politics-and-nation/bar-criminals-from-fighting-polls-sonia-gandhi/articleshow/5500935.cms).

3.2 Data and sample

We study candidate selection in India's 2009 and 2014 national elections using a dataset of official election returns combined with candidate and constituency characteristics. Summary statistics are in Appendix Table A.1.

Elections. Election returns come from the Election Commission of India (ECI) and for each constituency they contain turnout, each candidate's name, party, and number of votes. Constituency boundaries were set in 2008 and are unchanged throughout our sample period.

Our specification requires information on local (state legislative) elections matched to "corresponding" national elections.[13] Each national election constituency contains a subset of the state election constituencies (this assignment is also constant after April 2008). In most states, state elections are held in different years from the national election, every 5 years. For each state, we assign the first state election held after 2008 to the 2009 national election and the second state election to the 2014 national election. In practice this means that state elections held between 2008-2012 are assigned to the 2009 national election and state elections held between 2013-2017 are assigned to the 2014 national election.

Candidate characteristics. The ECI data contains information on candidates' gender, age, and caste (Scheduled Caste, Scheduled Tribe, or General). We supplement this with information on candidates' education, wealth, and criminal history collected and published by the civil group ADR at www.myneta.info. The latter is based on affidavits that all Indian candidates for national and state elections are required to file with the ECI.[14]

The criminal histories that candidates are required to report include previous criminal convictions as well as pending cases, i.e., cases where a judge decided to proceed with a criminal charge beyond the initial police investigation and prosecutorial action (roughly equivalent to an "indictment" in the US system). See Vaishnav (2017, p.318) for details.

Although this dataset on candidates is already quite rich, it does not contain information on one of the most important characteristics in Indian politics: whether the candidate is an ethnic Muslim. We construct a Muslim indicator based on candidate names using tools from text analysis. Specifically, we categorize candidates as Muslim or non-Muslim based on the "distance" between their name and text fragments commonly found in Muslim names. The details are in Appendix 1.

Constituency characteristics. The average constituency has 1.4 million eligible voters and

[13] India's state elections are notorious for the sheer number of candidates who contest them, which provides us with rich variation in candidate characteristics. The ECI had to increase the deposit that candidates pay to contest elections due primarily to the number of state election candidates. See Kapoor and Magesan (2018) for details.

[14] Previous studies using this dataset include Fisman et al. (2016), Prakash et al. (2019), Ujhelyi et al. (2021), and Asher and Novosad (2023), among others.

14 candidates running for election. In the Indian electoral system, some constituencies are reserved and can only be contested by Scheduled Caste or Scheduled Tribe candidates and the ECI data contains indicators for these reserved constituencies.

For demographic and other characteristics of each constituency, we use the SHRUG dataset (Asher et al. 2020). Specifically, we use village-level information from the 2011 Indian Census, which the SHRUG allows to be matched to constituencies. We use the following characteristics: literacy rate, share of working population, share of Scheduled Caste and Scheduled Tribe population, whether the village has access to paved roads, and whether the village is located in a rural or urban area.

We use both the village level information, and also aggregate it up to the constituency level. This creates some missing constituencies when villages could not be uniquely matched to constituencies (see Asher et al. (2020)).

Sample construction. Details of our sample construction are in Appendix 1. We drop constituencies with missing demographic information, and states with very few constituencies. Our final sample includes the 15 largest Indian states, and contains 232 national constituencies and 1629 state constituencies (about half of the constituencies in these states). These national constituencies are contested by a total of 3208 candidates in 2009 and 3373 candidates in 2014. The state constituencies are contested by 17,965 candidates in the 2009 election period and 18,801 in the 2014 election period.

4 Model

We consider a simultaneous Bayesian game of candidate selection between competing parties. Candidates are described by a set of characteristics. In selecting a candidate, each party weighs its internal preferences against the preferences of voters, and thus the probability of winning. We discuss the decision problem of parties and voters in turn.

4.1 Parties

Consider an electoral constituency where competing parties choose which candidate to run. Each party p chooses one candidate out of a set of potential candidates $\mathcal{A}_p$. Let the choice of party p be given by a_p, and denote the vector of choices of p's opponents by $\mathbf{a}_{-p}$. Voters cast their votes based on the candidates that parties choose to run: let $s_p(a_p, \mathbf{a}_{-p})$ represent the votes cast for party p given a candidate selection profile $(a_p, \mathbf{a}_{-p})$, and let $w_p(a_p, \mathbf{a}_{-p})$ represent its corresponding winning probability (both of these functions will be derived

endogenously below).[15]

A key innovation of our approach is to allow for the fact that parties may care about the candidate they run beyond its effect on votes. A party may experience *direct* costs or benefits from running specific candidates. This could reflect considerations such as the availability of certain types of candidates (e.g., a party with few Muslim members may find it more costly to run a Muslim candidate), internal politics (e.g., some candidates may be loyal to the party leadership, while others may come from a competing faction within the party) or party finances (e.g., some candidates may be able to finance their own campaigns, making them a "cheaper" choice for the party).

To capture these considerations, we specify party p's payoff from choosing candidate $a_p \in \mathcal{A}_p$ as

$$b^w w_p(a_p, \mathbf{a}_{-p}) + b^s s_p(a_p, \mathbf{a}_{-p}) + c_p(a_p) + \varepsilon_p(a_p). \tag{1}$$

The party cares about its winning probability, with weight b^w, as well as its vote share. The weight b^s on the latter could be positive if, e.g., the party leader's status is enhanced by a large vote share. It could also be negative if popular candidates may challenge the leader's authority or defect and form new parties, as discussed in Section 3.1. More generally, a winning candidate with a high vote share could take the party in directions that are disliked by the elite. Similarly, a losing candidate with a high vote share may be more difficult to replace in the next election.[16]

The term $c_p(a_p) + \varepsilon_p(a_p)$ captures payoffs from a_p that are independent of w_p and s_p. For clarity, we will refer to these direct payoffs as "costs." While $c_p(a_p)$ is observable to all competing parties, $\varepsilon_p(a_p)$ is party p's private information. The private component $\varepsilon_p(a_p)$ is i.i.d. across parties and candidates, with cdf $G(\cdot)$.

An important factor affecting $c_p(a_p)$ is the pool of potential candidates available to the party. This will be determined by who the party's members are, and who among its members has both the motivation and ability to run for office. For example, finding a Muslim candidate could be easier if the pool includes more Muslims. Or, running a candidate with a criminal history could be more socially acceptable within the party if the party has many such candidates in the pool.

As explained below, we will measure a party's pool of potential candidates using characteristics of the party's candidates in state elections. This is motivated by the fact that (i) many Indian parties competing in state elections have clear affiliations to a national party

[15] Throughout, s_p will correspond to the share of eligible voters who voted for party p (rather than the share among those who turn out to vote). We will refer to s_p as "vote share" for simplicity.

[16] Instead of vote shares, these considerations could alternatively be captured by vote margins (difference in vote share relative to the winner) or by the number of votes (implying higher values for larger constituencies). We find that using any of these alternatives makes little difference to the results - see Appendix 8.

(either the party is the same, or they belong to the same electoral alliance), and (ii) it is common for national politicians to begin their political careers in state elections. To highlight this, write $c_p(a_p) = c_p(a_p, L_p)$, where L_p denotes the pool of party p's candidates in relevant state elections.

Given the presence of private information, this setup gives rise to a simultaneous game of incomplete information between parties competing in the constituency. The solution concept is Bayesian Nash Equilibrium (BNE) in pure strategies.[17] For a realization of the private costs $\boldsymbol{\varepsilon}_p \equiv \{\varepsilon_p(a)\}_{a \in \mathcal{A}_p}$, a party chooses candidate $a_p(\boldsymbol{\varepsilon}_p)$. Let $P(\mathbf{a})$ denote the ex ante probability of a profile of choices $\mathbf{a}$. Then given ε_p and $P(\mathbf{a}_{-p})$, in a BNE party p chooses a_p to maximize its expected payoff given all other parties' strategies:

$$a_p \in \arg\max_a U_p(a, P(\mathbf{a}_{-p})),$$

where

$$U_p(a, P(\mathbf{a}_{-p})) \equiv E_P[b^w w_p(a, \mathbf{a}_{-p}) + b^s s_p(a, \mathbf{a}_{-p})|a] + c_p(a, L_p) + \varepsilon_p(a) \tag{2}$$

is the expected value of (1) over the possible realizations of opponents' choices $\mathbf{a}_{-p}$.

For the purposes of estimation it is convenient to express strategies as *choice probabilities* (CPs). In particular, define payoffs net of the private cost as

$$\tilde{U}_p(a, P(\mathbf{a}_{-p})) \equiv U_p(a, P(\mathbf{a}_{-p})) - \varepsilon_p(a) \tag{3}$$

so that a_p maximizes party p's expected payoffs iff:

$$\tilde{U}_p(a_p, P(\mathbf{a}_{-p})) + \varepsilon_p(a_p) \geq \tilde{U}_p(a, P(\mathbf{a}_{-p})) + \varepsilon_p(a) \quad \forall a \in \mathcal{A}_p.$$

The probability of party p choosing action a_p given the opponent's strategy $P(\mathbf{a}_{-p})$ is then:

$$\begin{aligned} P(a_p) &= \int_{\boldsymbol{\varepsilon}_p} \mathbf{1}\Big\{\varepsilon_p(a) - \varepsilon_p(a_p) \leq \tilde{U}_p(a_p, P(\mathbf{a}_{-p})) - \tilde{U}_p(a, P(\mathbf{a}_{-p})), \quad \forall a \in \mathcal{A}_p\Big\} dG(\boldsymbol{\varepsilon}_p) \\ &\equiv \Lambda_p(a_p; \mathbf{P}_{-p}) \end{aligned} \tag{4}$$

Equilibrium in the game is fully characterized by a fixed point in $P(\mathbf{a})$ of the system of equations defined by (4) for all p. Stacking equations by actions and parties, an equilibrium vector of CPs $\mathbf{P}^*$ satisfies $\mathbf{P}^* = \boldsymbol{\Lambda}(\mathbf{P}^*)$.

Under the assumption that $\varepsilon_p(a)$ follows the Type-I Extreme Value Distribution, the

[17] As long as $G(\cdot)$ is atomless, the existence of a pure-strategy BNE is guaranteed: see Fudenberg and Tirole (1991, Ch 6.8).

equilibrium CPs satisfy

$$\Lambda_p(a_p; \mathbf{P}_{-p}) = \frac{\exp\left\{\tilde{U}_p(a_p, \mathbf{P}(\mathbf{a}_{-p}))\right\}}{\sum_a \exp\left\{\tilde{U}_p(a, \mathbf{P}(\mathbf{a}_{-p}))\right\}} \ \forall p.$$

An important assumption, and limitation, of the above model is that it considers each constituency in isolation. That is, conditional on observables, choosing a candidate in one constituency has no bearing on the candidate chosen in another constituency. In practice it is possible that even after conditioning on observables, party decisions in one constituency affect decisions in another. Allowing for a party to jointly decide on candidates across constituencies (a "Colonel Blotto" type game) would render the model inestimable, as this would leave us with as many markets as we have national elections (two).

4.2 Voters

To model parties' winning probabilities and vote shares as a function of the set of candidates running, we consider the individual decisions made by a continuum of voters. We assume expressive voting with a flexible specification of voter preferences over candidates' characteristics (Ujhelyi et al. (2021) - USC (2021) from now on).[18]

Specifically, each candidate a_p can be described by a vector of characteristics $\mathbf{x}_p = \mathbf{x}(a_p)$, such as their education level or criminal history. Given a set of candidates that parties have chosen to run, voter i's utility from voting for the candidate of party p is

$$V_{ip} = \boldsymbol{\beta}_i \mathbf{x}_p + \xi_p + \eta_{ip}. \tag{5}$$

The first term represents voters' (potentially heterogenous) preferences over the characteristics of p's candidate. The second term, ξ_p, allows for unobserved (to the researcher) candidate characteristics valued by voters or, equivalently, shocks to parties' popularity in the given constituency. The distribution of ξ_p is left unspecified, and it can be correlated with $\mathbf{x}_p$. Finally, η_{ip} are individual preference shocks drawn from a Type-I Extreme Value distribution. To model the sources of preference heterogeneity among voters, write

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Pi} \mathbf{d}_i, \tag{6}$$

[18]The assumption of expressive voting is supported by extensive survey evidence on Indian voters' motivations. Banerjee (2017) provides a book-length discussion of the meaning that voters attach to the act of voting. Based on a recent survey, Heath and Ziegfeld (2022) estimate that at most 1.1% of individuals vote strategically.

where $\mathbf{d}_i$ is a vector of voter demographics, while $\boldsymbol{\beta}$ and $\boldsymbol{\Pi}$ contain the parameters.

To complete the voter's choice set, let $p = 0$ indicate the option to abstain and $V_{i0} = \boldsymbol{\pi}_0 \mathbf{d}_i + \eta_{i0}$ the voter's associated utility. This allows for the utility of abstention (hence the cost of voting) to vary across voters.

Voter i chooses option p (vote for one of the parties or abstain) if $V_{ip} > V_{ip'}$ for all $p' \neq p$. Thus, voters choose between their options based on the observed and unobserved candidate characteristics, the benefit of abstention, and their idiosyncratic shocks. This implicitly defines the set for which voter i will choose option p, $\{(\mathbf{d}_i, \boldsymbol{\eta}_i) | V_{ip} > V_{ip'} \text{ for all } p' \neq p\}$. Given a distribution of $\mathbf{d}_i$ and $\boldsymbol{\eta}_i$, integrating over this set yields parties' vote shares as a function of their candidate choices. Under the assumed Type-I EV distribution for η_{ip} and given a distribution $F(\mathbf{d}_i)$, these vote shares can be written as

$$s_p(\mathbf{x}_p, \mathbf{x}_{-p}) = \int \frac{\exp\left[\boldsymbol{\beta}_i \mathbf{x}_p + \xi_p - \boldsymbol{\pi}_0 \mathbf{d}_i\right]}{1 + \sum_{q>0} \exp\left[\boldsymbol{\beta}_i \mathbf{x}_q + \xi_q - \boldsymbol{\pi}_0 \mathbf{d}_i\right]} dF\left(\mathbf{d}_i\right). \tag{7}$$

This setup allows the domain of voter preferences to be different from the parties' choice sets (voter preferences are defined over characteristics $\mathbf{x}$ while the parties' choice set is $\mathcal{A}_p$). This is realistic because not every party may have access to candidates with all possible combinations of characteristics.[19]

To model the relationship between party choices and candidate characteristics, we simply assume that, after choices are made, candidate characteristics are drawn from a distribution $H(\cdot|a)$ for each party. Thus, party p's vote share is the expected vote share over the realizations of these characteristics:

$$s_p(a_p, \mathbf{a}_{-p}) = \int_{\mathbf{x}_p} \int_{\mathbf{x}_{-p}} s_p(\mathbf{x}_p, \mathbf{x}_{-p}) dH_p(\mathbf{x}_p|a_p) dH_{-p}(\mathbf{x}_{-p}|\mathbf{a}_{-p}) \tag{8}$$

where $s_p(\mathbf{x}_p, \mathbf{x}_{-p})$ is given in (7). Similarly, party p's winning probability is given by

$$w_p(a_p, \mathbf{a}_{-p}) = \int_{\mathbf{x}_p} \int_{\mathbf{x}_{-p}} \mathbf{1}\left\{s_p(\mathbf{x}_p, \mathbf{x}_{-p}) > s_{-p}(\mathbf{x}_p, \mathbf{x}_{-p})\right\} dH_p(\mathbf{x}_p|a_p) dH_{-p}(\mathbf{x}_{-p}|\mathbf{a}_{-p}) \tag{9}$$

In equilibrium, a party's expected vote share and winning probability takes into account

[19] This feature also allows our model to potentially be extended to situations where a party leader picking candidates may not have full control over their characteristics (for example, the leader might delegate candidate choice to subordinates). Modeling this explicitly could be an interesting avenue for future work.

its opponent's strategy, captured by $P(\mathbf{a}_{-p})$ from equation (4):

$$E_P[s_p(a_p, \mathbf{a}_{-p})|a_p] = \sum_{\mathbf{a}_{-p}} s_p(a_p, \mathbf{a}_{-p}) P(\mathbf{a}_{-p}) \tag{10}$$

$$E_P[w_p(a_p, \mathbf{a}_{-p})|a_p] = \sum_{\mathbf{a}_{-p}} w_p(a_p, \mathbf{a}_{-p}) P(\mathbf{a}_{-p}). \tag{11}$$

5 Specification and estimation

5.1 Overview

Our ultimate goal is to estimate parameters in the parties' objective function (2). To do this, we proceed in two stages. In the first stage, we estimate voters' utility functions (5) with a BLP procedure, using as instruments variables that enter parties objective function but do not directly enter voter utilities. This yields estimates of the voter preference parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Pi}$, as well as the popularity shocks $\boldsymbol{\xi}$. By construction, for the candidates observed in the data, the vote shares predicted with these estimates perfectly match the observed vote shares.

With these estimates, we can use our model of voters to predict parties' vote shares given *any* combination of candidates, based on (7). In the second stage, we use these estimated vote share functions to estimate benefit and cost parameters in (2) using a Pseudo-Maximum-Likelihood procedure.

5.2 Estimating voter preferences

Estimation of voter preferences follows the Generalized Method of Moments (GMM) algorithm proposed by Berry et al. (1995). Detailed treatments of the procedure can be found in Berry et al. (1995), Nevo (2000) and Nevo (2001). Here we modify a previous application of this procedure to Indian *state* elections in USC (2021).

5.2.1 Specification and endogenous characteristics

We focus on four candidate characteristics $\mathbf{x}$: education, Muslim, crime, and assets.[20] To be consistent with the estimation of party objectives below, we standardize all these variables to have 0 mean and unit standard deviation. We also include in this vector an indicator for

[20] We also considered three other characteristics observed in the data: caste, gender, and age. Gender has very little variation (almost all candidates are male). Caste has very little variation once we control for constituency reservation. Age does not seem to be an important characteristic for either voters or parties in these elections.

candidates where one or more characteristics were imputed. In addition, we include in (5) the following control variables: party and alliance fixed effects (to control for a portion of ξ_p that is common across constituencies), state and year fixed effects and an indicator for reserved constituencies.

To deal with the presence of many small parties and independent candidates, we follow USC (2021) and aggregate these candidates in each constituency. Specifically, in each constituency we aggregate into one "small party" category parties that are not part of either the UPA or the NDA alliance and only run a few candidates in the data.

The BLP procedure requires the use of instrumental variables (IV) for two reasons: first, to identify the "nonlinear" parameters $\mathbf{\Pi}$, and second, to identify the parameters on any variables in $\mathbf{x}$ that parties can adjust in response to the popularity shocks ξ_p (the typical source of endogeneity in the identification of demand). Because the focus of our study is parties' choice of their candidates, we treat all four candidate characteristics as endogenous.

Variables that enter parties' objective function (2) but do not directly enter voter utility are valid instruments. In our specification, the characteristics of a party's candidates in state elections, L_p, satisfy this condition. Recall the idea behind the presence of these variables in (2): a party's available pool of candidates affects its cost of choosing candidates with specific characteristics. For example, a party has a lower cost of finding an educated candidate when most candidates in its pool are educated. The prevalence of education (for example) in the pool of candidates is measured by the prevalence of this characteristic among the candidates a party runs in state elections in the same geographic area.[21]

The characteristics of a party's candidates in state elections are valid instruments as long as they are uncorrelated with voter valuations ξ_p in the national election. One potential threat to the identifying assumption is if L_p reflects party choices in the state election that are also affected by popularity shocks, and shocks at the national and state level are correlated. For example, a scandal involving a criminal politician could reduce the popularity of parties running criminal candidates at both the state and national level, and parties could respond by reducing the number of criminal politicians in both types of elections. Note, however, that (i) most state elections are held in different years than the corresponding national election, and (ii) the competitive environment (e.g., the number of competing parties, or voter priorities) can be quite different between national and state elections. For these reasons, it is not clear that popularity shocks are generally correlated, nor that parties would respond to similar shocks in the same way at the national and state levels.

[21]The idea corresponds to Industrial Organization applications that use cost shifters that affect firm profits but not consumer utilities as instruments for endogenous variables (typically, prices). USC (2021), which studied state elections, used candidate characteristics in neighboring constituencies based on a similar logic.

We create our instruments as a function of the party alliance (UPA/NDA/neither) and the state assembly constituencies overlapping with the national election constituency. For example, we instrument the assets of a UPA candidate with the average assets of all UPA candidates running in the state constituencies contained in the given national constituency. We create these instruments both for the same election and for the other election in the data.

To get a preliminary sense of the relevance and strength of these instruments (which we also explore in more detail in the next subsection), we regress each characteristic on the corresponding instruments and control variables. The results are shown in Table 1. In columns 1 and 2, the state election averages of the education characteristic have large and significant association with the education level of a party's candidates in the national election. Columns 3-8 show similar patterns for the other characteristics as well. These correlations, which are interesting in their own right, provide support for the idea that state candidate averages proxy for the pool of characteristics that a party is able to draw from at the national level.

Table 1: Characteristics regressed on instruments

Dep. var.:	Education		Muslim		Crime		Assets	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IV same election	0.19	0.16	0.54	0.34	0.29	0.28	0.32	0.31
	(0.05)	(0.05)	(0.07)	(0.08)	(0.06)	(0.06)	(0.05)	(0.05)
IV other election		0.34		0.44		0.10		0.10
		(0.08)		(0.10)		(0.09)		(0.08)
Adj. R^2	0.10	0.11	0.09	0.09	0.14	0.14	0.37	0.37
F	13.23	15.22	60.73	42.73	22.94	13.34	42.73	20.90
Mean dep. var.	0.14	0.14	-0.03	-0.03	0.05	0.05	-0.10	-0.10
Std. dev.	0.84	0.84	0.88	0.88	0.96	0.96	1.00	1.00

Notes: IV same election is the average of the given characteristic among an alliance's candidates in state election constituencies in the given year. *IV other election* is the same variable for the other election in the data (2009 for 2014 and vice versa). Regressions include fixed effects for state, year, party, alliance, imputed characteristics, and reserved constituencies. F is the Olea and Pflueger (2013) effective F statistic. Robust standard errors in parentheses. N = 2,649.

5.2.2 Differentiation IVs and specification choice

In order to identify nonlinear parameters, we use the "differentiation IVs" proposed by Gandhi and Houde (2019). The idea behind these instruments is to use the menu of choices available to each decision-maker to identify preference heterogeneity among decision-makers.

In our application, preference heterogeneity among voters for a candidate's education level (say) is identified based on how many candidates in a voter's choice set have similar education levels.

To construct the differentiation IVs, we first predict each endogenous candidate characteristic using the above instruments (and all exogenous variables). We then use these predicted characteristics to form instruments: for each characteristic of a candidate that enters the nonlinear part of voter utility, we compute the number of candidates in the constituency whose corresponding characteristic is within one standard deviation.

To guide our specification choice and evaluate the strength of our instruments, we use the methods proposed by Gandhi and Houde (2019). The details are in Appendix 2. The idea is to evaluate whether the instruments are "strong enough" to reject the linear (Logit) specification, i.e., $\mathbf{\Pi} = \mathbf{0}$. First, we enter the differentiation IVs as controls in a Logit specification. We find that the differentiation IVs for Muslim and assets are statistically significant while the differentiation IVs for education and crime are not. This suggests that the former two are capable of capturing departures from the Logit model. As an alternative diagnostic, we also run a specification that includes the differentiation IVs as instruments instead of controls. The overidentification J-test clearly rejects this specification, which also provides support for focusing on the nonlinear specifications (Gandhi and Houde 2019).

To further evaluate the nature of preference heterogeneity, we estimate a random coefficients specification where voter demographics $\mathbf{d}_i$ in (6) are replaced with Normally distributed random variables. This specification indicates the presence of significant heterogeneity in voters' preference for candidate assets, but not for the other three characteristics.

5.2.3 Adding voter demographics

Based on the specification checks described in the previous section, we first focus on identifying the sources of preference heterogeneity in voters' valuation of candidate assets. Our main demographic variables are literacy, rural population, presence of paved roads, working population, and lower caste population. We interact the differentiation IV for assets with the average value of each of these demographic variables in the constituency.[22]

We again evaluate these instruments using a Logit specification. This supports using the interaction of the asset differentiation IV with literacy, rural population, and presence of paved roads (Table A.3 of the Appendix). Estimating nonlinear specifications using different combinations of these instruments and corresponding nonlinear parameters yields a clear

[22]Following Gandhi and Houde (2019), the idea is to identify a parameter π_m^k on the interaction of the demographic d_m with the characteristic x^k using the instrument $\bar{d}_m \hat{x}^k$, where $\bar{d}_m$ is the average value of the demographic in the constituency and $\hat{x}^k$ is the differentiation IV for the candidate characteristic.

favorite, shown in column 1 of Table 2. This specification passes the overidentification J test, and results in nonlinear coefficients that are jointly statistically significant based on the Newey-West test.

Table 2: Voter preference parameter estimates

	(1)	(2)	(3)	(4)
Linear parameters (β)				
Education	-0.19	0.94	-0.59	-0.57
	(0.69)	(0.82)	(0.73)	(0.80)
Muslim	-0.50	-0.37	-0.54	-0.54
	(0.28)	(0.27)	(0.29)	(0.30)
Crime	0.66	0.73	0.56	0.59
	(0.51)	(0.57)	(0.50)	(0.50)
Assets	2.40	-2.00	2.81	4.84
	(3.48)	(1.14)	(2.72)	(6.05)
Nonlinear parameters (π)				
Assets × Literacy	-3.45	4.15		-2.42
	(3.49)	(1.85)		(5.95)
Assets × Road		2.59	-3.78	-3.60
		(1.50)	(2.85)	(3.37)
Assets × Rural	3.84		3.97	3.26
	(1.27)		(1.21)	(2.47)
J	1.481	9.715	8.880	9.407
df	4	4	4	4
p-value	0.83	0.05	0.06	0.05
Newey-West p-value	0.001	0.005	0.007	0.001

Notes: BLP estimates. Specifications include state, year, party and alliance fixed effects, indicators for imputed characteristics, and reserved constituencies. J is the overidentification J-statistic with its degree of freedom (df) and p-value. The bottom row shows the the p-value of the Newey-West D-test for the null that all nonlinear parameters are jointly 0. Robust standard errors clustered by constituency in parentheses. N = 2,649.

According to the estimates in column 1 of Table 2, all else equal voters dislike Muslim candidates and like candidates with a criminal history (consistent with Vaishnav (2017) and USC (2021)) though the latter is not statistically significant. Rural voters have a preference for candidates with more assets. A possible explanation is that wealthier candidates have higher social status and thus easier access to other government officials to advance local interests and "get things done." According to Vaishnav (2017), providing these connections between the local community and the state apparatus is very important in voters' evaluation of the candidates.

5.3 Specifying the parties' choice sets

In our specification of voter preferences, we conceptualized parties' choice of candidates a_p as a choice of a bundle of candidate characteristics $\mathbf{x}_p$. Applied directly to parties' problem, this would imply very large choice sets $\mathcal{A}_p$, containing all the possible combinations of characteristics. This is neither practical for estimation, nor realistic as a model of party choices. For example, it is unlikely that a party views two candidates who are identical in all dimensions but whose assets are slightly different as substantively different options.

For a more conceptually appealing (and computationally feasible) model of parties' problem, we assume the existence of a smaller set of candidate "types" that parties consider when choosing who to run. For example, a type could be an "educated non-Muslim with some criminal history in the second quartile of the asset distribution." Rather than making ad hoc assumptions about these types, we use machine learning tools to let the data tell us what they should be.

5.3.1 Data and variables for constructing candidate types

Our goal is to describe the pool of potential national candidates, as opposed to the set of candidates actually selected by the parties. As argued above, the candidates running in state elections provide a good proxy for the pool of candidates that national parties can select from. Thus, we define candidate types based on the characteristics of the candidates running in state elections.

There are many state candidates (36,766 in our data), which yields rich variation in candidate characteristics in the pool of potential candidates. This data also has many (13,787) independent candidates. These candidates are often individuals who wanted to run with a major party but were not selected and are thus highly informative about the type of candidates parties choose from.

As above, we use the candidate characteristics education, assets, criminal history and Muslim. Of these, education, assets, and criminal history have missing values. In order to assign each candidate to a type, we impute these missing characteristics as described in Appendix 1. [23] We standardize these variables to have 0 mean and unit standard deviation.

[23] An alternative approach we considered is to use a Missing indicator as an additional characteristic. However, there are differences between missing values in the state and national election data (for example, most independent candidates' education, assets, and criminal history is missing in the former). Thus, using Missing as an additional characteristic would mechanically make candidate types in the two datasets less comparable.

5.3.2 Clustering algorithm

We use k-means clustering to create the types. This iterative procedure partitions the data into K clusters based on the 4 variables described above (Muslim, education, assets and crimes). The algorithm begins by specifying K initial centroids and forming clusters by assigning each candidate to the closest centroid. Throughout, we use Eucledian distance to compute candidates' distance from a centroid. Next, new centroids are computed based on the average characteristics of the candidates assigned to each cluster. Using these centroids, candidates are reassigned to the closest cluster, and the process continues until no candidate is reassigned from their current cluster.

To use k-means clustering, one must first choose the number of clusters K. There is currently no cross-validation method to assess the relative performance of different values, and one option is to choose the number K based on substantive considerations (Athey and Imbens 2019).[24] Alternatively, we can use a set of common measures from the machine learning literature to select K. The first measure relies on the Within Cluster Sum of Squares (WCSS), also referred to as the Inertia score, which measures the similarity of units within the same cluster. The second measure is the Silhouette Coefficient (SC), which measures how far apart the clusters are from one another. The formula for each measure is given in Appendix 6. Ideally, the clusters should result in a low WCSS and a large SC.

The WCSS is shown on the left panel of Figure 1 for values of K ranging from $K = 1, ..., 20$. Notice that this measure is decreasing in K by construction - the more clusters, the more similar the members of the cluster will be. As can be seen in the figure, up to $K = 4$ each additional cluster creates large drops in the WCSS. By comparison, the gains from additional clusters beyond 4 are much smaller.[25]

On the right panel of Figure 1 we display the Silhouette Coefficient over the same range of values of K. The largest gain in SC, by far, occurs when moving from $K = 3$ to $K = 4$, consistent with the results for WCSS above. When moving from $K = 4$ to $K = 5$, SC actually decreases slightly, and increases in SC are relatively small for $K > 4$.

Each of these validation checks supports using $K = 4$.[26] In addition, as discussed in the

[24] Alternative unsupervised methods such as Density Based Clustering (DBSCAN) do not require the researcher to input the number of types K. The drawback of this type of method is that the researcher must select other hyperparameters, and the resulting clusters can be highly sensitive to these choices. Moreover, DBSCAN does not classify all points in the sample. Trebbi and Weese (2019) develop an interesting method to identify the number of organized insurgent groups using correlations in the timing of attacks over space. It is not obvious how to apply their method in our setting however, as identification of the number of clusters would require covariates that vary independently across our candidate characteristics, which we do not have.

[25] This approach for validating the choice of K is often referred to as the "elbow method" (Thorndike 1953).

[26] We also found that with $K = 4$, the clusters resulting from the algorithm were invariant to the starting values of the centroids. This was not the case with $K = 5$ and $K = 6$.

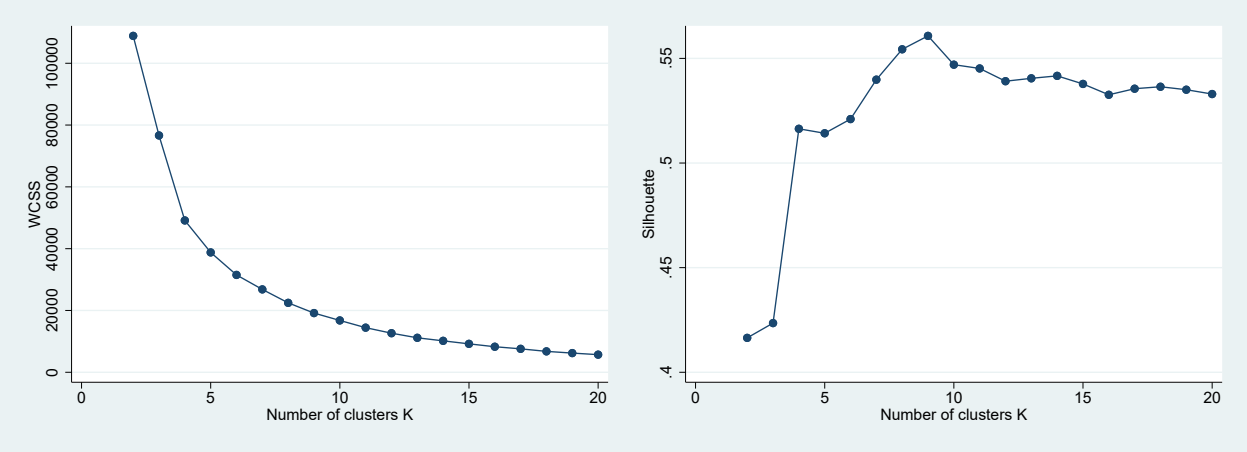


Figure 1: WCSS and Silhouette scores for different values of K

next section, $K = 4$ also yields candidate types that are easy to interpret and match well some of the important categories of politicians mentioned in the literature.

5.3.3 Candidate types

Table 3 shows the centroids of each of the four candidate types resulting from the k-means clustering algorithm. The types resulting from the algorithm turn out to have very clear interpretations. Types 1 and 2 are both non-Muslim candidates with no criminal history, but Type 1 has high education while Type 2 has low education. Type 3 contains all the Muslim candidates. Type 4 contains all the non-Muslim candidates with criminal history, and these candidates are also richer than the other three types. For simplicity, we will refer to the four types as *educated*, *uneducated*, *Muslim*, and *criminal* types, respectively.

These candidate types obtained from the clustering algorithm appear quite sensible. Given the salience of the Muslim characteristic in Indian politics, it is not surprising that one of the types we obtain is defined by this characteristic. The positive correlation between criminality and wealth, reflected in our criminal type, is consistent with Vaishnav (2017) and Asher and Novosad (2023). Finally, the importance of education among non-Muslim non-criminal candidates is consistent with a large literature that uses education as the main measure of candidate valence (e.g., Galasso and Nannicini (2011); Besley and Reynal-Querol (2011)).

In sum, the types resulting from the clustering reflect meaningful choices for the parties, and not simply some artificial combination of candidate characteristics. To further support this, in Appendix 6 we ask how robust the types are to using a different clustering algorithm (agglomerative hierarchical clustering). Remarkably, we find that the two algorithms perfectly agree on the Muslim and criminal types. They differ somewhat in the allocation of

the remaining candidates, but we show that k-means clustering dominates the hierarchical algorithm based on WCSS and Silhouette scores.

Table 3: Centroids of candidate types

	Assets	Crimes	Education	Muslim
Type 1	14.35	0.00	1.00	0.00
Type 2	14.27	0.00	0.00	0.00
Type 3	14.30	0.17	0.44	1.00
Type 4	15.06	1.00	0.68	0.00

Notes: Centroids resulting from the k-means clustering algorithm. The algorithm is run on standardized variables; the table shows the centroids transformed back to the original scale for ease of interpretation. Assets: real Rp in logs; Crimes: 1 if has at least one criminal case; Education: 1 if completed high school; Muslim: 1 if Muslim name.

The left panel of Table 4 shows the distribution of candidate types among all candidates, as well as the candidates of the UPA and the NDA. Compared to the UPA, the NDA has a somewhat higher share of uneducated types and (not surprisingly) a smaller share of Muslim types.

Table 4: Distribution of candidate types

	State elections			National elections		
	All candidates	UPA	NDA	All candidates	UPA	NDA
Type 1	34.74	46.78	44.15	45.16	50.98	52.95
Type 2	39.63	16.37	20.76	20.63	5.47	9.09
Type 3	11.20	11.27	6.55	12.48	10.94	5.23
Type 4	14.43	25.58	28.55	21.73	32.6	32.73
Total	100	100	100	100	100	100
N	36,764	3,788	4,109	6,577	457	440

Notes: Candidates in state and national elections assigned to each type by the clustering algorithm. Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively.

5.3.4 Types in the national elections

Next we assign national candidates to the candidate types created from the state elections data. To do this, recall that the types in Table 3 have clear definitions in terms of characteristics: Type 1 contains all the educated, non-Muslim, non-criminal candidates; Type 2 all the uneducated, non-Muslim, non-criminal candidates; Type 3 all the Muslim candidates; and Type 4 all the non-Muslim criminal candidates. We assign national candidates to types

using these definitions.[27]

The resulting distribution of candidate types is shown in the right panel of Table 4. Relative to state candidates, there are relatively more candidates of the educated type (and fewer of the uneducated type). There are also more of the criminal type, but this difference between state and national elections is less pronounced for candidates of the UPA and the NDA. The share of the Muslim type is similar between state and national elections. As in the state elections, the most pronounced difference between the NDA and the UPA is the former's lower share of the Muslim type and, to a lesser extent, its higher share of candidates of the uneducated type.

Table 5: Electoral performance and frequency of different candidate types in the raw data

	Winner	Closeness to winner	Vote share	Choice share
Type 1	0.31	0.65	0.19	0.52
Type 2	0.35	0.67	0.19	0.07
Type 3	0.23	0.62	0.19	0.08
Type 4	0.35	0.70	0.20	0.33

Notes: Average values from the data for UPA and NDA candidates, by candidate type. Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively. *Winner* is an indicator for candidates who won. *Closeness to winner* is the candidate's vote share divided by the winner's vote share. *Vote share* is the vote share observed in the data. Choice share is relative frequency in the data. N = 897.

Table 5 presents information on the average electoral performance of each candidate type in the raw data. The first two columns are measures of winning probability: "winner" is an indicator equal to 1 if the candidate won, and "closeness to winner" is the candidate's vote share divided by the winner's vote share. These values suggest that different candidate types have different winning probabilities. The criminal type has the highest average on both measures. In spite of these differences, the vote share of different types is very similar on average. A possible explanation is that parties do not attempt to increase their vote shares beyond the minimum necessary to win, in line with the discussion in Section 3.1. The choice frequencies of the different types, shown in the last column, indicate a very different pattern from either win probabilities or vote shares. This is suggestive of the fact that parties' choices are driven by other considerations.

[27] We also tried an alternative assignment where we assign each candidate to the type with the closest centroid. This creates a very similar assignment: in particular, only 3 UPA or NDA candidates are assigned to a different type.

5.4 Specification, identification, and estimation of party objectives

We focus on candidate selection by the two major party alliances in Indian politics, the UPA and the NDA. Other parties are included in the model as non-strategic players, with their candidates' characteristics fixed at their values observed in the data.

The choice set of each party $p = UPA,\ NDA$ contains the $K = 4$ candidate types obtained from the clustering algorithm. Based on (2), we specify party p's objective function in constituency c as

$$U_p(a_{pc}, P(\mathbf{a}_{-p,c})) = \sum_{\mathbf{a}_{-p,c}} \left[b^w w_{pc}(a_{pc}, \mathbf{a}_{-p,c}) + b^s s_{pc}(a_{pc}, \mathbf{a}_{-p,c})\right] P(\mathbf{a}_{-p,c}) + \sum_{k=2}^{4} (c^0_{kp} + \mathbf{c}_k \mathbf{L}_{pc}) \mathbf{1}\{a_{pc} = k\} + \varepsilon_{pc}(a_{pc}), \tag{12}$$

where $\mathbf{a}_{-p,c}$ includes all other parties' choices (including non-strategic parties'), $\mathbf{L}_{pc}$ contains the proxies for the pool of candidate characteristics (the average of candidate education, assets, Muslim and criminal history in the assembly constituencies corresponding to constituency c), and c^0_{kp} and $\mathbf{c}_k$ are type-specific cost parameters. As will become clear below, the cost parameters are only identified for three of the four types; we use $k = 1$ as the excluded category in (12). We will refer to the costs that depend on the candidate pool, $\mathbf{c}_k$, as *recruitment costs* and the costs c^0_{kp} as *direct costs*.

Our goal is to estimate the parameters $\boldsymbol{\theta} = (b^w, b^s, \{c^0_{kp}, \mathbf{c}_k\}^4_{k=2})$, which include parties' payoffs from increasing their probability of winning and their vote share, and their costs of running different candidate types.

To compute the vote shares and winning probabilities corresponding to different hypothetical candidate choices by the parties, we use our demand estimates. We compute the vote shares and associated winning probabilities for all possible action profiles $(a_{pc}, \mathbf{a}_{-p,c})$ of the players in a given constituency (16 profiles), holding fixed all exogenous variables (including non-strategic parties' choices) as well as the popularity shocks ξ_{pc}.

To use our demand estimates, we also need to assign a vector of characteristics to each type choice in the data. As described in Section 4.2, when parties choose types, there is some uncertainty over the actual realization of characteristics. We capture this uncertainty using the distribution of candidate characteristics for each type in the state election data. For each combination of types, we draw 1000 values of their characteristics from the data, and compute vote shares and winning probabilities as expected values across these draws (equations (8) and (9), respectively).

We provide a formal discussion of the identification of the party objective function parameters in Appendix 3, but it is worth briefly discussing the sources of identification here. As is typically the case in models of discrete choice, the cost parameters are identified only up to differences with respect to a reference alternative. That is, we identify $c_k - c_1$ for $k = 2, 3, 4$ where Type 1 is the reference type. The identification of these differences, for given parameters (b^w, b^s) is standard (see Appendix 3) and depends on the magnitude of the observed probability of selecting type k, $P(a_{pc} = k)$ relative to the probability of selecting the reference type.

By contrast, the benefit parameters are pinned down in levels, not differences. To see this, note that given the Type-I Extreme Value assumption we can express the relationship between choice probabilities and party payoffs as:

$$\begin{aligned} \ln(P(a_{pc} = k)) - \ln(P(a_{pc} = 1)) &= b^w \times (E[w_{pc}(k, \mathbf{a}_{-p,c})] - E[w_{pc}(1, \mathbf{a}_{-p,c})]) \\ &+ b^s \times (E[s_{pc}(k, \mathbf{a}_{-p,c})] - E[s_{pc}(1, \mathbf{a}_{-p,c})]) \\ &+ c_k - c_1 + \eta_{pc}(k) \end{aligned} \tag{13}$$

where we have assumed a single type-specific cost term c_k for simplicity. As we can treat choice probabilities and win probabilities as known, this can be viewed as a regression of $\ln(P(a_{pc} = k)) - \ln(P(a_{pc} = 1))$ on $(E[w_{pc}(k, \mathbf{a}_{-p,c})] - E[w_{pc}(1, \mathbf{a}_{-p,c})])$ and $(E[s_{pc}(k, \mathbf{a}_{-p,c})] - E[s_{pc}(1, \mathbf{a}_{-p,c})])$ where the intercept is the cost difference $c_k - c_1$. Then b^w is, loosely, identified as the covariance across constituencies between the probability of selecting type k relative to the reference option, $\ln(P(a_{pc} = k)) - \ln(P(a_{pc} = 1))$, and the difference in expected win probabilities, $E[w_{pc}(k, \mathbf{a}_{-p,c})] - E[w_{pc}(1, \mathbf{a}_{-p,c})]$. If the parties tend to select candidate k in the constituencies where they are likely to win, b^w will be positive. See Appendix 3 for a more formal and detailed discussion.

Estimation proceeds by recursively updating the choice probabilities using (pseudo) maximum likelihood estimates of the parameter vector $\boldsymbol{\theta}$ up to convergence as in Aguirregabiria and Mira (2007). Specifically, consider an initial choice probability estimate $\hat{P}^0(\mathbf{a}_{-p,c})$. In constituency c, party p chooses a_{pc} to maximize (12). Again defining utility net of the unobservable as $\tilde{U}_p(a, P(\mathbf{a}_{-p,c}); \boldsymbol{\theta})$, the implied probability that $a_{pc} = a$ is:

$$P(a|\hat{\mathbf{P}}^0, \boldsymbol{\theta}) = \frac{\exp\left\{\tilde{U}_p(a, \hat{\mathbf{P}}^0(\mathbf{a}_{-p,c}), \boldsymbol{\theta})\right\}}{\sum_{a'} \exp\left\{\tilde{U}_p(a', \hat{\mathbf{P}}^0(\mathbf{a}_{-p,c}), \boldsymbol{\theta})\right\}} \tag{14}$$

where we emphasize the fact that the choice probabilities are a function of the estimates $\hat{\mathbf{P}}^0$ as well as the parameters $\boldsymbol{\theta}$. Denoting parties' choices observed in the data with a^*_{pc}, the

estimates $\hat{\boldsymbol{\theta}}^0$ maximize the log likelihood

$$\ell(\boldsymbol{\theta},\hat{\mathbf{P}}^0) = \sum_{pc}\sum_{a} \mathbf{1}\{a = a^*_{pc}\} \ln P(a|\hat{\mathbf{P}}^0, \boldsymbol{\theta}).$$

With these estimates, we can construct new estimates of the choice probabilities as

$$\hat{P}^1(a|\hat{\mathbf{P}}^0) = \frac{\exp\left\{\tilde{U}_p(a, \hat{P}^0(\mathbf{a}_{-p,c}), \hat{\boldsymbol{\theta}}^0)\right\}}{\sum_{a'}\exp\left\{\tilde{U}_p(a', \hat{P}^0(\mathbf{a}_{-p,c}), \hat{\boldsymbol{\theta}}^0)\right\}}.$$

Given these new choice probabilities, the equilibrium probability that $a_{pc} = a$ in equation (14) becomes $P(a|\hat{\mathbf{P}}^1, \boldsymbol{\theta})$. In turn, this yields an updated log likelihood $\ell(\boldsymbol{\theta}, \hat{\mathbf{P}}^1)$, which is maximized to obtain a new estimate $\hat{\boldsymbol{\theta}}^1$. We iterate in this way until convergence. The resulting estimator, $\hat{\boldsymbol{\theta}}_{NPL}$ is asymptotically equivalent to the Maximum Likelihood estimator. See Aguirregabiria and Mira (2007) for details. Standard errors are obtained from the inverse Hessian of the likelihood function evaluated at $\hat{\boldsymbol{\theta}}_{NPL}$.

6 Estimation results

6.1 What do parties maximize?

In Table 6, we first estimate a version of the model where parties care only about their expected winning probability and vote share. We then introduce the recruitment costs $\mathbf{c}_k\mathbf{L}_{pc}$ in column 2, and also add the direct costs c^0_{kp} in column 3. Allowing for these costs substantially improves the model's fit: as we move from column 1 to 3, the Log Likelihood increases by 29.7%.[28] Chi-square tests at the bottom of columns 2 and 3 indicate that the cost parameters are highly jointly significant. Costs, which are independent of voter preferences, are important in explaining parties' choices of which candidates to run.

According to our estimates, parties like to win ($b^w > 0$), but, all else equal, choosing candidates with higher vote shares has a disutility ($b^s < 0$). This is consistent with the idea that party elites balance anticipated electoral performance of a popular candidate against the possibility that the candidate becomes too powerful and threatens the elite - as discussed in the literature on Indian parties (see Section 3.1). To interpret the magnitude of the estimates, we can use the standard deviation of parties' equilibrium payoffs, which is 1.49. All else equal, winning raises parties' payoff by 2.2 standard deviations (3.24/1.49). Winning with a one

[28] See Section 6.3 below for a more detailed evaluation of model fit. In Appendix 5 we provide more results illustrating how the fit improves as we introduce costs.

standard deviation higher vote share (7 percentage points in the data) lowers payoffs by two thirds of a standard deviation ($-14.15 \times 0.07/1.49 = -0.66$).

Table 6: Party objective function estimates

	(1)	(2)	(3)		(1)	(2)	(3)
b^w	4.26	3.08	3.24	c_4^{educ}		-0.64	-0.41
	(0.84)	(0.95)	(1.02)			(0.20)	(0.23)
b^s	-5.35	-9.20	-14.15	c_2^{crime}		-0.47	-0.03
	(1.22)	(1.45)	(1.73)			(0.22)	(0.24)
$c_{2,NDA}^0$			-1.57	c_3^{crime}		-0.31	0.36
			(0.27)			(0.21)	(0.23)
$c_{3,NDA}^0$			-2.58	c_4^{crime}		1.06	1.10
			(0.30)			(0.14)	(0.15)
$c_{4,NDA}^0$			0.58	c_2^{asset}		-0.73	0.32
			(0.23)			(0.19)	(0.26)
$c_{2,UPA}^0$			-2.09	c_3^{asset}		-1.08	-0.12
			(0.32)			(0.20)	(0.26)
$c_{3,UPA}^0$			-1.93	c_4^{asset}		0.06	-0.06
			(0.29)			(0.14)	(0.16)
$c_{4,UPA}^0$			0.40	c_2^{Muslim}		0.44	-0.07
			(0.23)			(0.32)	(0.41)
c_2^{educ}		-2.01	-0.53	c_3^{Muslim}		1.74	1.48
		(0.26)	(0.34)			(0.27)	(0.29)
c_3^{educ}		-2.30	-0.63	c_4^{Muslim}		0.12	0.03
		(0.26)	(0.34)			(0.24)	(0.24)
Log likelihood					-1189.00	-919.41	-836.21
Joint significance of c's					-	355.02	482.45

Notes: Estimates of party objective functions in (12). The cost parameters c are measured relative to the excluded category, Type 1. $c_{k,p}^0$ is party p's direct cost of choosing a type k candidate. c_k^l is the impact of characteristic l in the pool of candidates on the recruitment cost of a type k candidate. Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively. Standard errors in parentheses. Number of markets: 434.

The recruitment cost parameters $\mathbf{c}_k$ show that a higher prevalence of some candidate characteristic in a party's candidate pool lowers the party's cost of selecting a candidate type with that characteristic. For example, the positive estimate for c_4^{crime} shows that a higher prevalence of criminal candidates in the pool increases a party's payoff from choosing the criminal Type 4. The negative estimates of c_k^{educ} for $k = 2, 3, 4$ indicate that the payoff from choosing the educated Type 1 (the excluded category) increases with a higher prevalence of educated candidates in the pool. In this way, the supply of candidate characteristics available to parties affects who they choose to run, mirroring the patterns seen in Table 1.

Column 3 shows that parties' costs of running specific candidates are not restricted to the recruitment costs. According to these estimates, the NDA has the largest direct cost from running a Muslim type (Type 3): running a Muslim candidate relative to an educated candidate lowers its payoff by 1.7 standard deviations, all else equal (2.58/1.49). This is in line with the Hindu nationalist profile of the BJP, the NDA's leading party, and indicates that the NDA's aversion to running Muslim types is not due simply to voter preferences.

The UPA has the largest direct cost from uneducated types (Type 2), and both parties incur the lowest direct cost (or highest benefit) from running a criminal type (Type 4), followed by the (excluded) educated type. The idea that parties have relatively low costs of running criminal candidates is in line with Vaishnav (2017)'s argument that these candidates are willing to fund their own campaigns, perhaps circumventing campaign finance laws, which makes them cheaper for the parties.

Figure 2 plots the recruitment costs for each type, and compares them to the direct costs (indicated with vertical lines). Both recruitment costs and direct costs matter; for Type 2 and 3, direct costs are larger in magnitude compared to recruitment costs. The median recruitment cost implies a payoff reduction of 0.05 and 0.30 standard deviations for Types 2 and 3, respectively, and a payoff increase of 0.12 standard deviations for Type 4. Interestingly, the recruitment cost parameters for Type 4 are mostly positive (i.e., a benefit). In most constituencies, parties incur relatively low costs from selecting criminal types compared to other types due to the ample supply of criminality in the candidate pool.

6.2 Why do parties choose criminal candidates?

Our estimates reveal a nuanced set of reasons that guide parties' choice of criminal candidates. This formalizes previous explanations proposed in the literature, and adds several new considerations.

First, according to our estimates, voters like the criminal type, particularly because these candidates are wealthy. Since parties like to win, they have an incentive to choose criminal types to increase their winning probability. At the same time, criminal types can generate particularly large vote shares, and our results show that this creates a disutility for party leaders for a given winning probability.[29] As discussed above, the disutility from vote shares that are "too large" can reflect threats from factionalism and defections from the party. Therefore, the finding that this disutility is particularly important for criminal types links together two influential literatures on candidate selection in India, one on criminal candidates

[29]To quantify this effect for the criminal type, in Appendix Table A.15 we show that without the distutility of large vote shares (setting $b^s = 0$), the share of criminal candidates chosen by each party would more than double.

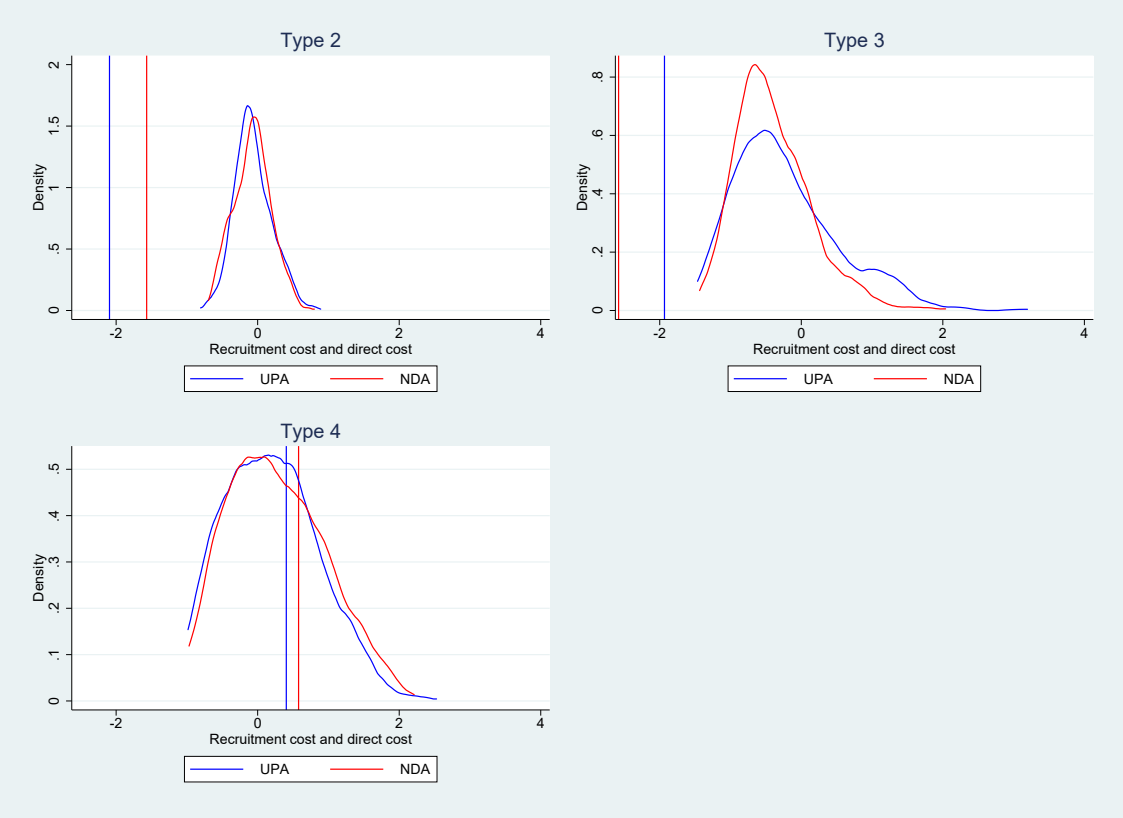


Figure 2: Distribution of each party's costs of different types across constituencies

Based on column 3 of Table 6. Kernel density plots of recruitment costs $\mathbf{c}_k\mathbf{L}_{pc}$. Vertical lines are direct costs c^0_{kp}. Types 2-4 are the Uneducated, Muslim, and Criminal types, respectively (Type 1 is the excluded category).

(Vaishnav 2017), and one on party factions (Chandra 2016).

Second, criminal types also affect party payoffs independently of votes. We found that in most constituencies recruitment costs favor this type: the large supply of criminal candidates makes it relatively cheap to run them (Figure 2). We also found that criminal candidates have the lowest direct costs among the four types. This is consistent with the descriptive literature on how criminals' private wealth and connections can benefit the party. In Appendix 4 we further illustrate the importance of wealth as a determinant of candidate choice by sorting parties' equilibrium payoff by the wealth of their chosen candidate.

Importantly, this pattern of costs and benefits raises the possibility that running criminals is a *strategic* response to the other party's criminal candidate. A party could run a criminal simply because its opponent is doing that as well. Intuitively, there are two reasons why a party could find it relatively advantageous to run a criminal when its opponent is running a criminal. First, because voters like the criminal type, running a different type against the opponent's criminal would yield a low probability of winning. Second, when the opponent is running a criminal, choosing a criminal increases the probability of winning without generating vote shares that are "too large" and therefore costly for the party.

Thus, criminals may be strategic complements for the parties. This possibility is supported by the fact that the correlation between the two parties' choice probability of a criminal type is positive (0.30), while the correlation between one party's choice probability for a criminal type and the other party's choice probability for any other type is always negative (ranging between -0.14 and -0.29). See Appendix Table A.9.

To investigate this directly, we compute $\frac{\partial P_p(4)}{\partial P_{-p}(4)}$, the change in party p's probability of choosing a criminal (Type 4) in response to a change in its opponent's probability of choosing a criminal. A positive derivative for both parties indicates strategic complementarity (see Appendix 7). Figure 3 shows the distribution of this derivative across constituencies. For the UPA (NDA), the derivative is positive in 85% (87%) of the constituencies. The derivative is positive for both parties in 77% of the constituencies (and negative for both in only 6%). In most cases, criminal candidates are strategic complements.

Because running a criminal has costs for the party, strategic complementarity raises the possibility that in equilibrium parties run criminals more often than they would like to. As we show below, this has important policy implications for measures that restrict parties' ability to run criminal candidates.

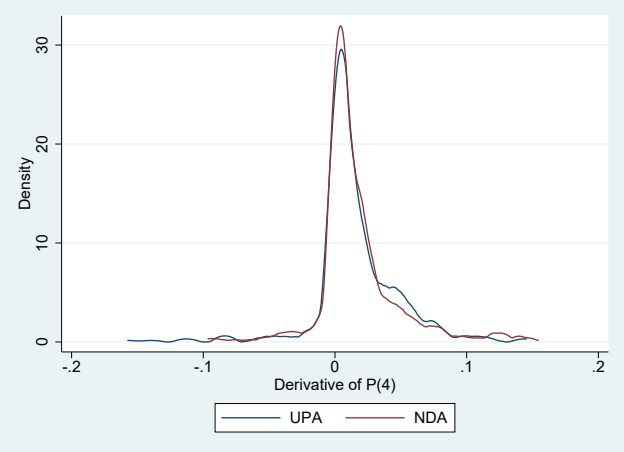


Figure 3: Strategic complementarity of criminals

Notes: Kernel density plots of $\partial P_p(4)/\partial P_{-p}(4)$ based on equation (11) in Appendix 7.

6.3 Robustness and model fit

In Appendix 8 we present a series of robustness checks on our estimates. We replace the vote share terms $b^s s(\cdot)$ in parties' objective function with terms allowing for further nonlinearities, with vote margins, or with the number of votes. We find that these changes make little difference to our estimates. We show that controlling for several sources of observable constituency level heterogeneity, including election year, reserved constituencies, or the share of rural population, does not change our estimates in a material way. We also apply the method of Bonhomme et al. (2022) to control for *unobservable* constituency level heterogeneity, and show that this also leaves our main estimates unaffected.

To explore the model's ability to fit the data, we first use the estimated model to simulate party choices in every constituency. We draw actions using the equilibrium choice probabilities 100 times, and compare the average over simulations with the actual choices observed in the data for each type and each party. The result is in Panel A of Table 7. The model performs well.

To evaluate the model's "out of sample" performance we use a cross-validation procedure. We repeat the following 5 times. First, we hold out the first 20% of the sample, and estimate the model using the remaining 80% of observations. We then solve for the equilibrium in each constituency in the 20% hold-out sample, and use this to simulate party choices and evaluate the model's fit on that sample. We repeat this process for the next 20-80 split, etc., and take the average across the 5 sets of predictions (this is essentially a k-fold cross validation, with $k = 5$). The results are in Panel B of Table 7. While there is variation across the folds in predictive ability, on average over the k-folds the model does just as well

Table 7: Model fit

	UPA actual	UPA predicted	NDA actual	NDA predicted	All actual	All predicted
Panel A: Actual and predicted type choices						
Type 1	217	221	229	234.80	446	455.80
Type 2	24	24	40	41	64	61.60
Type 3	49	46.20	22	19.80	71	66.00
Type 4	144	146.20	143	138.40	287	284.60
Panel B: 5-fold cross validation						
Type 1	41.4	40.68	43.4	42.00	84.80	82.68
Type 2	4.8	3.88	7.8	7.68	12.60	11.56
Type 3	9.2	10.32	4.4	5.64	13.60	15.96
Type 4	27.60	28.12	27.40	27.68	55.00	55.80

Notes: Panel A shows the number of candidates of each type observed in the data and predicted by the model (average across 100 simulations). Panel B uses a 5-fold cross validation procedure as described in the text (values shown are averages across 20% subsamples). Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively.

in predicting outcomes as in Panel A.

7 Policy experiment: banning the criminal type

What is the impact of banning candidates with a criminal history from elections? As discussed above, this is a relevant policy question in India and many other countries.

To model this kind of policy, we consider a counterfactual scenario where we make it prohibitively costly to choose the criminal Type 4 for both parties. Although Type 3 also contains some candidates with criminal history, Type 4 candidates *always* have a criminal history (Table 3). They are also wealthier, and thus more closely match the kind of criminal candidates who are considered problematic in the Indian context (Vaishnav 2017).

Using our parameter estimates, we compute a new equilibrium. This involves computing the expected vote share of candidate types once the criminal type has been banned. We do this in two ways: banning the criminal type only for the UPA and the NDA but leaving third party candidates as is, or also removing third parties that run criminal candidates in the demand model (we focus on the former in the main analysis). We study the change in parties' choices, winning probabilities, and payoffs in the counterfactual equilibrium compared to the baseline equilibrium.

7.1 Changes in candidate characteristics

A first observation is that, because candidates are bundles of correlated characteristics, removing a candidate type directly affects the distribution of characteristics among the candidates running for election. The criminal type is relatively wealthy, and comes from the non-Muslim majority. Removing these candidates may therefore raise the share of less wealthy and minority candidates. Such changes may be unintended side effects of a policy of banning candidates with a criminal history.

Table 8: Different types' choice and winning probabilities, with and without the Criminal type

	Choice probability		Conditional win probability	
	Baseline	Counterfactual	Baseline	Counterfactual
All candidates				
Type 1	51.4	76.8	19.0	20.1
Type 2	7.4	11.5	20.8	22.0
Type 3	8.2	11.7	12.9	13.7
Type 4	33.1	0.0	38.2	-
UPA				
Type 1	50.0	75.4	16.7	17.8
Type 2	5.5	8.7	17.9	19.1
Type 3	11.3	15.9	11.3	12.0
Type 4	33.2	0.0	34.2	-
NDA				
Type 1	52.8	78.2	21.4	22.5
Type 2	9.2	14.4	23.6	24.9
Type 3	5.1	7.4	14.6	15.5
Type 4	32.9	0.0	42.2	-

Notes: Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively. Values shown are the averages across all the constituencies in the data. Conditional win probability is the probability that a party would win conditional on choosing a given type.

The first two columns of Table 8 compare the average choice probabilities (i.e., the predicted share of each candidate type among the candidates contesting the election) in the no-criminal counterfactual equilibrium and the baseline equilibrium. The full distribution of the changes in probabilities is shown in Appendix Figure A.16.

We find that the choice probabilities of candidate types 1-3 all increase. Note that in single-agent models, reducing the choice set could never reduce the choice probability of a remaining option. In the context of our model where parties play best responses to each-other's strategies, this is not necessarily the case. To illustrate with an example that ignores incomplete information, take a constituency where the NDA ran a Type 1 candidate and the

UPA a Type 4 candidate. Losing the option of running a Type 4 candidate, the UPA might switch to Type 2. If the NDA's best response is to switch to Type 3, then all else equal the share of Type 2 and 3 candidates would increase, but the share of Type 1 would decrease.

Although choice probabilities for types 1-3 could decrease, we find that, for each type, this is the case in less than 2% of the constituencies. In absolute terms, the increase in average choice probabilities is largest for the educated type (Type 1), whose expected share increases by 25.4 percentage points, from 51.4 to 76.8 percent (Table 8). Eliminating criminal candidates increases the share of uneducated candidates (Type 2) by 55 percent (from 7.4 to 11.5), and the share of minority (Muslim) candidates (Type 3) by 43 percent (from 8.2 to 11.7). The latter effect is larger for the NDA, where the share of Muslim candidates was relatively low initially.

As shown in Table 8, these changes also impact the political viability of different candidate types, captured by the conditional winning probabilities $E_P[w_p(a_p, \mathbf{a}_{-p})|a_p]$ (the probability that a type would win conditional on being selected). In each case, this conditional winning probability increases on average. Overall, we find that banning criminal candidates would raise the probability of running, as well as the probability of winning, of educated, uneducated, and Muslim candidates.

7.2 Changes in parties' winning probabilities and payoffs

Figure 4 shows the changes in the two parties' vote share and probability of winning following a ban on criminal types. The left panel shows that vote shares almost always decline, by an average of 3.1 percentage points. To put this in context, the average winning margin in the data is 8.1 percentage points. The reduction in vote shares also means a reduction in the probability of winning. The probability that the UPA wins falls in 93% of the markets, by an average of 3 percentage points. The probability that the NDA wins falls in 92% of the markets, by an average of 5 percentage points.

One reason for the decline in the main parties' vote share is that once criminal types are not available, some voters find the main parties less attractive relative to third parties: the probability that a third party wins increases in 97% of the markets. This pattern is apparent already in the data: In constituencies where both main parties run criminal types, the average total vote share of all third parties is 21.1%; in constituencies where neither main party runs a criminal type, this vote share increases to 28.6%. The main parties' equilibrium response to a criminal ban can mitigate the decrease in popularity, but Figure 4 indicates that they cannot fully offset it, and their winning probability goes down.[30]

[30]It is important to note that the reduction in winning probabilities for the two major parties is *not* due to voters' switching to third parties who run criminals. First, we obtain similar patterns when we restrict

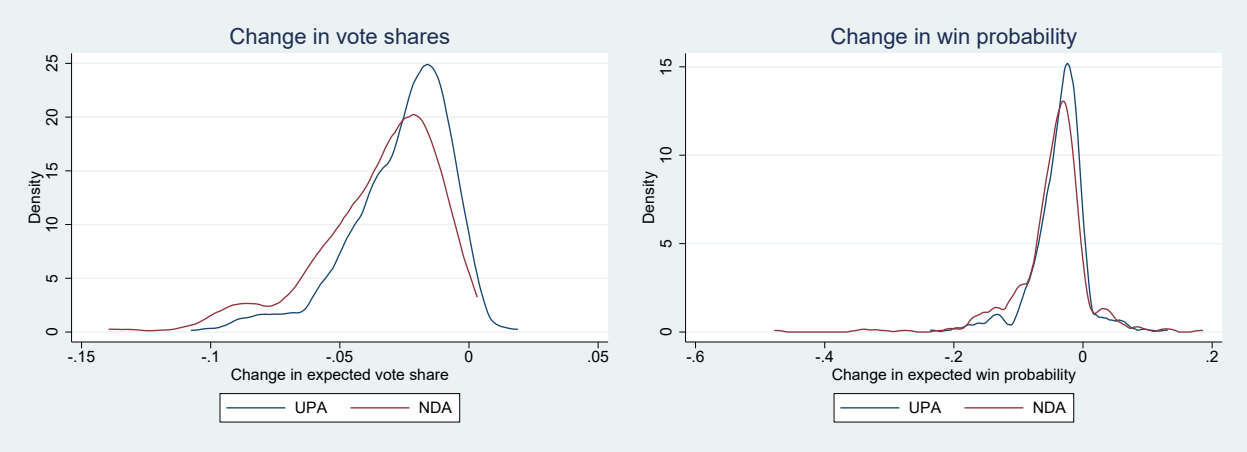


Figure 4: Distribution of changes in parties' vote shares and winning probabilities in the counterfactual

Computing the changes in parties' expected equilibrium payoffs yields the distributions shown in Figure 5. On average, both the UPA's and the NDA's payoffs decrease when the criminal type is banned. In 28% of the constituencies, both parties' payoffs decline. However, there are also many constituencies where payoffs rise: payoffs rise for both parties in 34% of the constituencies, for only the UPA in 17%, and for only the NDA in 21%.

The reason for this is related to the strategic complementarity discussed in Section 6.2. According to our estimates, a party can be incentivized to choose a criminal because its opponent chooses a criminal. Banning the criminal type allows parties to profitably choose different candidates. These candidates yield lower equilibrium vote shares, and in some cases lower recruitment costs, which raises parties' payoffs. Even though their probability of winning goes down, this decline is not large enough to undo the positive effect. These findings suggest that, in some cases, the main Indian parties may be willing to collectively support a ban on criminal candidates. This helps explain instances when these parties express support for a ban on criminal candidates while at the same time running such candidates in elections.

attention to the 78 markets where none of the third parties runs a criminal: here the probability that the UPA wins falls in 91% of cases, the probability that the NDA wins falls in 88% of cases, and the probability that a third party wins increases in all cases but one (99%). Second, we also ran a set of counterfactuals where we removed all third parties running criminals, and measured the impact of the criminal ban relative to this modified baseline. We again found that the criminal ban lowered the winning probability for both the UPA (81% of markets) and the NDA (80% of markets), and increased the probability of third party winners (82% of markets).

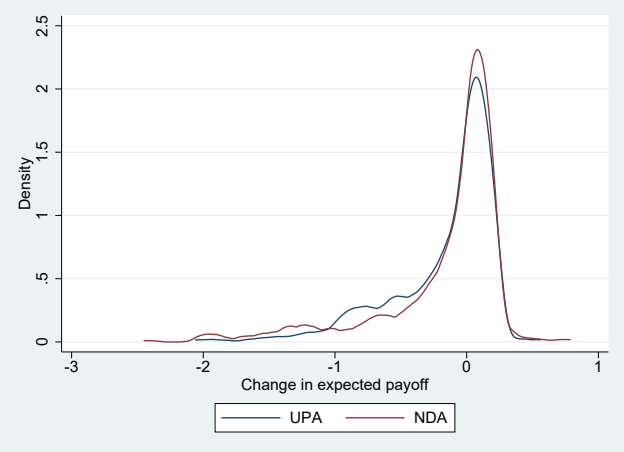


Figure 5: Distribution of the change in parties' payoffs in the counterfactual

7.3 Party organization and candidate eligibility regulations

As we saw above, parties' response to a criminal ban was determined to a large extent by their aversion to running overly popular candidates (i.e., the fact that $b^s < 0$). This illustrates that in order to understand the impact of regulations affecting candidate eligibility, it is important to understand the tradeoffs that guide party nomination strategies.

Because different party organizations imply different tradeoffs, this in turn implies that the impact of candidate eligibility regulations will be mediated by party organizations. To illustrate this point, we now repeat the policy experiment of a criminal ban, but in a setting where parties do not experience the disutility from running popular candidates, i.e., when $b^s = 0$. This could capture a setting where intraparty competition of the form described by Caillaud and Tirole (2002) and Mattozzi and Merlo (2015) is less prevalent, or where defectors from a party cannot easily join or establish alternative parties.

Equilibrium choices in this case are summarized in Appendix Table A.15. These are remarkably similar to the same policy experiment for the $b^s < 0$ case. This is despite the fact that without the criminal ban, parties' choices in the $b^s = 0$ and $b^s < 0$ case are very different (columns 1 and 2 of Appendix Table A.15. It seems that a criminal ban reduces the importance of organizational features that create a disincentive for running popular candidates. In this sense, candidate eligibility regulations and party organization can be substitutes.

8 Conclusion

We estimate a model of candidate selection by political parties to study why parties in a representative democracy select the candidates they do. Our setting is India, where we combine a rich demand side specification of voter preferences with a supply side game between parties that incorporates direct payoffs from candidate selection. To make the problem feasible, we use a machine learning algorithm to assign candidates to types based on detailed information on their characteristics.

We find that, while parties systematically select candidates with a higher probability of winning the election, selection decisions are shaped by other considerations. All else equal, parties would prefer to win with candidates who are not overly popular with voters. This is consistent with a common perception in Indian politics that the major parties are reluctant to select candidates who might become too powerful and threaten the party and its elites. In addition, we find that selection decisions are shaped by considerations that are independent of voter preferences. Parties are more likely to select candidates whose characteristics are abundant in local candidate pools, and they also have distinct direct preferences over candidate types.

Our estimates provide a detailed explanation for the factors that cause parties to run criminal candidates. Apart from electoral considerations and party costs, we also find an important strategic complementarity. This creates an incentive for parties to run criminals because their opponent does so as well.

We use the estimated model to study the consequences of a ban on criminal candidates, a question of current policy relevance in India and elsewhere. We find that the ban reduces the chances that India's major national parties win the election, as the ban causes more voters to choose third party alternatives. In spite of this, banning criminal candidates can increase the major parties' payoffs by allowing them to compete with more desirable candidates.

Our paper provides insights for other countries where crime in politics is a salient problem, and more generally a method to identify and estimate the forces that guide parties' selection of candidates.

References

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

Asher, S., T. Lunt, R. Matsuura, and P. Novosad (2020): "Development Research

at High Geographic Resolution: An Analysis of Night Lights, Firms, and Poverty in India using the SHRUG Open Data Platform," *working paper*.

Asher, S. and P. Novosad (2023): "Rent-Seeking and Criminal Politicians: Evidence from Mining Booms," *The Review of Economics and Statistics*, 105, 20–39.

Athey, S. and G. W. Imbens (2019): "Machine learning methods that economists should know about," *Annual Review of Economics*, 11, 685–725.

Bandiera, O., A. Prat, S. Hansen, and R. Sadun (2020): "CEO behavior and firm performance," *Journal of Political Economy*, 128, 1325–1369.

Banerjee, M. (2017): *Why India Votes?*, Routledge.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

Besley, T. (2005): "Political Selection," *Journal of Economic Perspectives*, 19, 43–60.

Besley, T., O. Folke, T. Persson, and J. Rickne (2017): "Gender quotas and the crisis of the mediocre man: Theory and evidence from Sweden," *American Economic Review*, 107, 2204–2242.

Besley, T. and M. Reynal-Querol (2011): "Do democracies select more educated leaders?" *American political science review*, 105, 552–566.

Bonhomme, S., T. Lamadon, and E. Manresa (2022): "Discretizing unobserved heterogeneity," *Econometrica*, 90, 625–643.

Caillaud, B. and J. Tirole (2002): "Parties as Political Intermediaries," *The Quarterly Journal of Economics*, 117, 1453–1489.

Casas-Arce, P. and A. Saiz (2015): "Women and power: unpopular, unwilling, or held back?" *Journal of political Economy*, 123, 641–669.

Casey, K., A. B. Kamara, and N. F. Meriggi (2021): "An experiment in candidate selection," *American Economic Review*, 111, 1575–1612.

Chandra, K. (2016): *Democratic dynasties: State, party, and family in contemporary Indian politics*, Cambridge University Press.

Chhibber, P., F. R. Jensenius, and P. Suryanarayan (2014): "Party organization and party proliferation in India," *Party Politics*, 20, 489–505.

Dal Bó, E. and F. Finan (2018): "Progress and perspectives in the study of political selection," *Annual Review of Economics*, 10, 541–575.

Dal Bó, E., F. Finan, O. Folke, T. Persson, and J. Rickne (2017): "Who becomes a politician?" *The Quarterly Journal of Economics*, 132, 1877–1914.

Dar, A. (2019): "Parachuters vs. Climbers: Economic Consequences of Barriers to Political Entry in a Democracy," *working paper*.

Farooqui, A. and E. Sridharan (2014): "Incumbency, internal processes and renomination in Indian parties," *Commonwealth & Comparative Politics*, 52, 78–108.

Fisman, R., F. Schulz, and V. Vig (2016): "Financial disclosure and political selection: Evidence from India," *working paper*.

Fudenberg, D. and J. Tirole (1991): *Game theory*, MIT press.

Galasso, V. and T. Nannicini (2011): "Competing on good politicians," *American Political Science Review*, 105, 79–99.

Gandhi, A. and J.-F. Houde (2019): "Measuring substitution patterns in differentiated-products industries," *NBER Working paper w26375*.

Gulzar, S., Z. Hai, and B. K. Paudel (2021): "Information, candidate selection, and the quality of representation: Evidence from Nepal," *The Journal of Politics*, 83, 1511–1528.

Hamilton, B. H., A. Hincapié, R. A. Miller, and N. W. Papageorge (2021): "Innovation and Diffusion of Medical Treatment," *International Economic Review*, 62, 953–1009.

Heath, O. and A. Ziegfeld (2022): "Why So Little Strategic Voting in India?" *American Political Science Review*, 116, 1523–1529.

Iaryczower, M., G. Kim, and S. Montero (2024): "Representation Failure," *working paper*.

Kapoor, S. and A. Magesan (2018): "Independent candidates and political representation in India," *American Political Science Review*, 112, 678–697.

Lee, A. (2020): "Incumbency, parties, and legislatures: Theory and evidence from India," *Comparative Politics*, 52, 311–331.

Mattozzi, A. and A. Merlo (2015): "Mediocracy," *Journal of Public Economics*, 130, 32–44.

Nevo, A. (2000): "A practitioner's guide to estimation of random-coefficients logit models of demand," *Journal of Economics & Management Strategy*, 9, 513–548.

——— (2001): "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 69, 307–342.

Olea, J. L. M. and C. Pflueger (2013): "A Robust Test for Weak Instruments," *Journal of Business & Economic Statistics*, 31, 358–369.

Paiva, N., J. Sakai, and L. Schoenster (2014): *Quatro em cada dez candidatos a governador têm ocorrências na Justiça ou nos Tribunais de Contas*, Transparencia Brasil, https://www.transparencia.org.br/publicacoes.

Prakash, N., M. Rockmore, and Y. Uppal (2019): "Do criminally accused politicians

affect economic outcomes? Evidence from India," *Journal of Development Economics*, 141, 102370.

Roy, R. (1966): "Selection of Congress candidates. I: The formal criteria," *Economic and Political Weekly*, 833–840.

Thorndike, R. (1953): "Who belongs in the family?" *Psychometrika*, 18, 267–276.

Trebbi, F. and E. Weese (2019): "Insurgency and small wars: Estimation of unobserved coalition structures," *Econometrica*, 87, 463–496.

Ujhelyi, G., S. Chatterjee, and A. Szabó (2021): "None of the Above: Protest Voting in the World's Largest Democracy," *Journal of the European Economic Association*, 19, 1936–1979.

Vaishnav, M. (2017): *When crime pays: Money and muscle in Indian politics*, Yale University Press.

Online Appendix to "Candidate Selection by Parties: Crime and Politics in India"

Arvind Magesan
Economics Department, University of Calgary

Andrea Szabó
Economics Department, University of Houston

Gergely Ujhelyi
Economics Department, University of Houston

June 21, 2024

Contents

1 Data and sample **3**
1.1 Construction of the Muslim characteristic 3
1.2 Sample construction 4
1.3 Missing characteristics 4

2 Specification choice for voter preferences **7**

3 Identification of model parameters **7**

4 Candidate wealth and party payoffs **13**

5 Model fit and validation **13**

6 Clustering **15**
6.1 Details of k-means clustering 15
6.2 Alternative clustering procedure 15

7 Criminals as strategic complements **17**

8 Other robustness checks and counterfactuals **18**
8.1 Robustness 18
8.1.1 Alternative payoff specifications 18
8.1.2 Constituency-level heterogeneity 20
8.2 Additional counterfactual results 27

1 Data and sample

1.1 Construction of the Muslim characteristic

We construct a Muslim indicator based on observed candidate names and common fragments (substrings) of names. We proceed as follows. First, with the aid of a research assistant from India, we used the candidate affidavit data, covering all elections from 2008-2019, to obtain a library of Muslim names and common substrings, or "name fragments," found in Muslim names. Indian Muslim names have Arabic or Farsi origins but are often spelled differently or modified from the original name in some other way. For this reason, in constructing our library we used names of candidates in India as opposed to a general list of Muslim names to ensure a higher degree of name matching accuracy.

Muslim names are quite distinctive from other Indian names and contain fragments that clearly distinguish them, making it quite easy to identify Muslim names in the affidavit data. Conversely, non-Muslim fragments would also be easy to isolate. For example "shankar" would not show up in a Muslim name and "ahmed" would not show up as part of a non-Muslim name. There are a total of 470 names and fragments in the Muslim library. We do the same for non-Muslim names.[1] There are a total of 1210 names and fragments in the non-Muslim library. We then compute two measures for every single name in our data which we wish to classify as Muslim or non-Muslim. The first measure is the "distance" between the name to each of the two libraries. Denote the library of Muslim names as M and the library of non-Muslim names as H. Then, for every candidate name $name_i$ in our data, we calculate the Levenshtein distance to every item in each of M and H.[2] We take the distance of $name_i$ and library M to be the minimum of these distances for names in library M. Similarly, the distance between $name_i$ and library H is the smallest distance between $name_i$ and all names in H. Let $d(name_i, M)$ and $d(name_i, H)$ denote these distances.

Next, we use the name fragments to construct another measure. Specifically count, for every name in the data, how many Muslim fragments and how many non-Muslim fragments appear in the name, and divide this by the number of fragments in the respective library to get a frequency. Denote these as $frag(name_i, M)$ and $frag(name_i, H)$ respectively.

Finally, $name_i$ is assigned "Muslim" identity if either $frag(name_i, M) > frag(name_i, H)$ or $\{frag(name_i, M) = frag(name_i, H)$ and $d(name_i, M) < d(name_i, H)\}$, and it is assigned "non-Muslim" identity otherwise.

[1] While the majority of non-Muslim names will have Hindu or Sikh origins, there is also a substantial population with Christian names.

[2] The Levenshtein distance simply counts the number of single edits required to turn one string into another. For example the Levenshtein distance between "car" and "stare" is 3.

1.2 Sample construction

Our dataset is limited by the state constituencies for which we can obtain demographic information from the SHRUG. Because we analyze national elections, we also need to aggregate the state constituencies up to the national constituency level. There are reasons to believe that state constituencies that are missing in a national constituency are systematically different from other state constituencies (e.g., they are more likely to be large urban areas). Therefore we only include in our analysis national constituencies for which we have information on all the state constituencies they contain. This drops from the sample several states in their entirety (mostly small states with only a few national constituencies).

The states excluded and the total number of their national constituencies are: Arunachal Pradesh (2), Goa (2), Manipur (2), Meghalaya (2), Mizoram (1), Nagaland (1), Puducherry (1), Punjab (13), Sikkim (1), Tripura (2), Uttarakhand (5), Delhi NCT (7).

From the remaining 18 states, we drop 3 because we either only have less than 20% of their constituencies or because they have very few constituencies to begin with. Specifically, we drop Chhattisgarh, with only 2 out of 11 constituencies, Himachal Pradesh, with 2 out of 4 constituencies, and Jammu & Kashmir, with 1 out of 6 constituencies.

The remaining 15 states contain 478 of the 538 constituencies in India, and we have constituency characteristics from the SHRUG for 234 of these. We drop 2 constituencies because some of their candidates have unrealistically high numbers of criminal convictions,[3] leaving us with a total of 232 constituencies in the dataset.

1.3 Missing characteristics

The specification of both voters' and parties' choices requires that all relevant candidate characteristics be observed. In the data, education, assets, and criminal history have missing values, and we impute these characteristics based on the candidate's gender, age, and caste (which have no missing values). Specifically, we impute assets, number of criminal cases, and number of completed years of education using the average by gender, caste and age range, where the age range is specified as +/- 1 year relative to the candidate's age. For example, a 30 year old male general caste candidate's imputed asset is the average of all male general caste candidates aged 29-31. We then create the variables used in the estimation and the clustering algorithm (log(assets + 1), an indicator for at least one criminal case, and an indicator for at least twelve years of education).

In practice, the main impact of imputing these characteristics is that we are using all candidates in the data when creating types.

[3] Both of these are in Tamil Nadu, and both have a candidate with close to 400 criminal cases.

The state election data has 36,766 observations. We drop 2 observations because they have missing characteristics and have no similar candidates (based on gender, caste and age) that we could use for imputing these missing values. The asset variable has 6,301 missing values and an additional 562 zeros which we also treat as missing. Criminal history has 6,301, and education 7,629 missing values.

The national election data has 6,581 observations. We drop 4 observations with missing characteristics that cannot be imputed. After aggregating small parties and independent candidates, we have 2,649 observations. Assets, criminal history, and education have, respectively, 170, 195 and 250 missing values. Among the 897 candidates of the UPA and the NDA, the corresponding numbers are 43, 41, and 58, respectively.

Table A.1: Summary statistics

	N	Mean	Std. Dev.	Median	10%	90%
A. Candidate characteristics - state elections						
UPA (0/1)	36766	0.10				
NDA (0/1)	36766	0.11				
Education (0/1)	29137	0.62				
Muslim (0/1)	36766	0.11				
Criminal history (0/1)	30465	0.20				
Assets (log)	29903	13.90	2.36	14.00	10.84	16.83
Male (0/1)	36766	0.93				
Age	36766	44.76	11.27	44	31	61
B. Candidate characteristics - national elections						
UPA (0/1)	6581	0.07				
NDA (0/1)	6581	0.07				
Education (0/1)	2900	0.75				
Muslim (0/1)	6581	0.12				
Criminal history (0/1)	3068	0.25				
Assets (log)	3012	14.78	2.52	14.97	11.55	17.71
Male (0/1)	6581	0.93				
Age	6581	46.25	11.96	45	31	63
SC or ST (0/1)	6581	0.35				
C. Constituency characteristics - national elections						
Eligible voters (1000)	464	1421.50	195.59	1426.24	1173.14	1685.34
Turnout (%)	464	64.97	12.39	65.99	47.57	81.05
N. of candidates before aggregation	464	14.18	6.06	14	7	22
N. of candidates after aggregation	464	5.71	1.46	5	4	8
Reserved constituency (0/1)	464	0.28				
Literate population (share)	464	0.61	0.10	0.61	0.49	0.74
ST and SC population (share)	464	0.27	0.14	0.24	0.13	0.46
Rural population (share)	464	0.82	0.11	0.83	0.68	0.94
Population with paved roads (share)	464	0.83	0.19	0.90	0.58	1.00
Working population (share)	464	0.42	0.07	0.43	0.32	0.50

Notes: Education: 1 if completed high school. Criminal history: 1 if has at least one criminal case. N. of candidates after aggregation: number of candidates after small-party candidates are aggregated as described in Section 5.2.1 in the paper. Assets (log): log of real assets in Rp.

2 Specification choice for voter preferences

Following Gandhi and Houde (2019), we first enter the differentiation IVs as controls in a Logit specification. This specification still includes all the controls described above, and instruments the endogenous characteristics with the 8 instruments created from the state election data. The results are in Table A.2. In column 1, the differentiation IVs for Muslim and assets are statistically significant while the differentiation IVs for education and crime are not. This suggests that the former two are capable of capturing departures from the Logit model. As an alternative diagnostic, we also run a specification that includes the differentiation IVs as instruments instead of controls. The last row of the table (IIA p-val) shows the p-value of the overidentification J-test for this specification. The fact that this specification is clearly rejected also provides support for focusing on the nonlinear specifications (Gandhi and Houde 2019). In column 2 we use only the differentiation IVs for Muslim and assets and obtain similar conclusions.

Column 3 of Table A.2 shows a random coefficients specification where voter demographics $\mathbf{d}_i$ in equation (6) in the paper are replaced with random variables drawn from a standard normal distribution. We use a separate i.i.d. variable for each of the four candidate characteristics, and estimate this specification using the BLP procedure with the differentiation IVs as instruments. This specification indicates the presence of significant heterogeneity in voters' preference for candidate assets, but not for the other three characteristics.

3 Identification of model parameters

Here we formally establish identification of the parameters of the model of candidate selection and discuss the intuition of the identification results. Throughout, we assume that choice probabilities, win probabilities, and expected vote shares are known to the researcher - these are estimated in a first stage. Let the probability that party $p \in \{1, 2\}$ chooses action $a_p = k$ for $k = 1, ..., K$ given observable payoff variables $\mathbf{z}$ (i.e., constituency characteristics) be given by $P_p(k, \mathbf{z})$, and write expected winning probability of party p as $w_p^P(k, \mathbf{z}) = E_p[w_p(a_p, a_{-p}, \mathbf{z})|a_p = k]$, where the expectation $E_p[w_p(a_p, a_{-p}, \mathbf{z})|a_p = k]$ is an integration over a_{-p} using player $-p$'s choice probability (see Section 3). Similarly, write the expected vote share as $s_p^P(k, \mathbf{z}) = E_p[s_p(a_p, a_{-p}, \mathbf{z})|a_p = k]$.

We establish identification in the baseline model where parties have preferences over win probability, expected vote share and have type specific costs, as most of the intuition can be gleaned from this case, and allowing for additional cost parameters (i.e., recruting costs) as in our full model does not substantially change the identification argument.

Table A.2: Specification choice: differentiation IVs and random coefficients

	(1)	(2)	(3)		(1)	(2)	(3)
Education	-0.63	-0.62	-0.38	diffIV(Educ)	0.05		
	(0.40)	(0.40)	(0.51)		(0.03)		
Muslim	-0.37	-0.39	-0.42	diffIV(Muslim)	-0.07	-0.08	
	(0.17)	(0.17)	(0.32)		(0.03)	(0.02)	
Crime	-0.00	0.16	0.33	diffIV(Crime)	-0.04		
	(0.34)	(0.31)	(0.38)		(0.03)		
Assets	1.67	1.66	2.65	diffIV(Assets)	-0.08	-0.07	
	(0.27)	(0.28)	(0.56)		(0.04)	(0.04)	
				$\pi_{Education}$			-0.04
							(8.00)
				π_{Muslim}			-0.06
							(6.82)
				π_{Crime}			0.04
							(9.82)
				π_{Assets}			-1.63
							(0.46)
J p-val					0.10	0.07	0.02
IIA p-val					0.00	0.00	

Notes: Columns (1) and (2) are specification checks proposed by Gandhi and Houde (2019). The dependent variable is vote shares. Candidate characteristics are instrumented with the instruments described in section 5.2.1 in the paper, and the "differentiation IVs" are entered as controls. Specifications also control for state, year, party and alliance fixed effects, indicators for imputed characteristics, and reserved constituencies. J p-val is the p-value of the overidentification J test. IIA p-val is the p-value of the overidentification J test when the diffIV variables are used as instruments instead of controls. Column (3) is a random-coefficients specification, using standard Normal draws instead of voter demographics, and using the four diffIV variables as instruments. Robust standard errors in parentheses.

Table A.3: Specification choice: differentiation IVs and constituency demographics

	(1)	(2)	(3)	(4)	(5)	(6)
diffIV(Assets × Literacy)	-0.20					-0.30
	(0.05)					(0.16)
diffIV(Assets × Rural)		-0.09				0.17
		(0.04)				(0.09)
diffIV(Assets × Roads)			-0.15			-0.18
			(0.04)			(0.07)
diffIV(Assets × Workers)				-0.24		0.08
				(0.08)		(0.26)
diffIV(Assets × Caste)					-0.12	0.12
					(0.09)	(0.12)
J p-val	0.07	0.04	0.06	0.05	0.03	0.20
IIA p-val	0.00	0.00	0.00	0.00	0.02	0.00

Notes: Specification checks proposed by Gandhi and Houde (2019). The dependent variable is vote shares. Candidate characteristics are instrumented with the instruments described in section 5.2.1 in the paper, and the "differentiation IVs" are entered as controls. Only the coefficients on the differentiation IVs are shown. Specifications also control for state, year, party and alliance fixed effects, indicators for imputed characteristics, and reserved constituencies. J p-val is the p-value of the overidentification J test. IIA p-val is the p-value of the overidentification J test when the diffIV variables are used as instruments instead of controls. Robust standard errors in parentheses.

To build intuition, we begin with a simple case where $c_k = 0 \quad \forall k$, so that the only parameters to identify are (b^w, b^s). With identification in this simple case established, we then re-introduce cost parameters below and derive full identification results.

In the model with no costs, Party p's choice probability satisfies

$$P_p(k, \mathbf{z}) = \Lambda\big(b^w \times w_p^P(k, \mathbf{z}) + b^s \times s_p^P(k, \mathbf{z})\big)$$

where, given our assumption about the error distribution:

$$\Lambda\big(b^w \times w_p^P(k, \mathbf{z}) + b^s \times s_p^P(k, \mathbf{z})\big) = \frac{\exp\left\{b^w \times w_p^P(k, \mathbf{z}) + b^s \times s_p^P(k, \mathbf{z})\right\}}{\sum_{k'} \exp\left\{b^w \times w_p^P(k', \mathbf{z}) + b^s \times s_p^P(k', \mathbf{z})\right\}}$$

As the argument for identification is symmetric across players, we drop the p subscript in what follows for expositional purposes.

Inverting the choice probability gives:

$$\begin{aligned} \Lambda^{-1}\big(P(k, \mathbf{z})\big) &= \ln\big(P(k, \mathbf{z})\big) - \ln\big(P(K, \mathbf{z})\big) \\ &= b^w \times (w^P(k, \mathbf{z}) - w^P(K, \mathbf{z})) + b^s \times (s^P(k, \mathbf{z}) - s^P(K, \mathbf{z})) \end{aligned} \tag{1}$$

where we have taken type K as the reference type.

Define $\Delta_w^P(k, \mathbf{z}) \equiv w^P(k, \mathbf{z}) - w^P(K, \mathbf{z})$ as the *increased expected probability of winning when selecting type k relative to the reference type K*, and similarly $\Delta_s^P(k, \mathbf{z}) \equiv s^P(k, \mathbf{z}) - s^P(K, \mathbf{z})$.

Now, consider two constituencies, $1, 2$ with values of $\mathbf{z}$ of $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$. Using these two constituencies we get a system of equations

$$\begin{bmatrix} \Lambda^{-1}\big(P(k, \mathbf{z}^{(1)})\big) \\ \Lambda^{-1}\big(P(k, \mathbf{z}^{(2)})\big) \end{bmatrix} = \begin{bmatrix} \Delta_w^P(k, \mathbf{z}^{(1)}) & \Delta_s^P(k, \mathbf{z}^{(1)}) \\ \Delta_w^P(k, \mathbf{z}^{(2)}) & \Delta_s^P(k, \mathbf{z}^{(2)}) \end{bmatrix} \begin{bmatrix} b^w \\ b^s \end{bmatrix} \tag{2}$$

The parameters (b^w, b^s) are identified if the matrix $\begin{bmatrix} \Delta_w^P(k, \mathbf{z}^{(1)}) & \Delta_s^P(k, \mathbf{z}^{(1)}) \\ \Delta_w^P(k, \mathbf{z}^{(2)}) & \Delta_s^P(k, \mathbf{z}^{(2)}) \end{bmatrix}$ has full rank. This is the case if

$$\frac{\Delta_w^P(k, \mathbf{z}^{(1)})}{\Delta_w^P(k, \mathbf{z}^{(2)})} \neq \frac{\Delta_s^P(k, \mathbf{z}^{(1)})}{\Delta_s^P(k, \mathbf{z}^{(2)})} \tag{3}$$

This is the key condition for identification of (b^w, b^s). Essentially, there must be at least two constituencies where the difference across constituencies in win probability associated with type k is not equal to the difference across constituencies in expected vote share associated with type k. As vote shares can change without turning a loss into a win, this condition is easy to satisfy.

Solving the system in (2) we get:

$$b^w = \frac{\Delta_s^P(k, \mathbf{z}^{(2)})\Lambda^{-1}(P(k, \mathbf{z}^{(1)})) - \Delta_s^P(k, \mathbf{z}^{(1)})\Lambda^{-1}(P(k, \mathbf{z}^{(2)}))}{\Delta_s^P(k, \mathbf{z}^{(2)})\Delta_w^P(k, \mathbf{z}^{(1)}) - \Delta_s^P(k, \mathbf{z}^{(1)})\Delta_w^P(k, \mathbf{z}^{(2)})} \tag{4}$$

and:

$$b^s = \frac{\Delta_w^P(k, \mathbf{z}^{(1)})\Lambda^{-1}(P(k, \mathbf{z}^{(2)})) - \Delta_w^P(k, \mathbf{z}^{(2)})\Lambda^{-1}(P(k, \mathbf{z}^{(1)}))}{\Delta_s^P(k, \mathbf{z}^{(2)})\Delta_w^P(k, \mathbf{z}^{(1)}) - \Delta_s^P(k, \mathbf{z}^{(1)})\Delta_w^P(k, \mathbf{z}^{(2)})} \tag{5}$$

To get some intuition for these identified values, assume first that the denominator is positive (the case with a negative denominator is symmetric) so that:

$$\frac{\Delta_w^P(k, \mathbf{z}^{(1)})}{\Delta_w^P(k, \mathbf{z}^{(2)})} > \frac{\Delta_s^P(k, \mathbf{z}^{(1)})}{\Delta_s^P(k, \mathbf{z}^{(2)})}$$

In this case the relative returns to win probability with type k are greater in constituency 1 than constituency 2, and the relative returns to expected vote share are greater in constituency 2 than 1.

First, $b^w > 0$ if and only if:

$$\frac{\Lambda^{-1}(P(k,\mathbf{z}^{(1)}))}{\Lambda^{-1}(P(k,\mathbf{z}^{(2)}))} > \frac{\Delta_s^P(k,\mathbf{z}^{(1)})}{\Delta_s^P(k,\mathbf{z}^{(2)})}$$

This is intuitive - the inverse function $\Lambda^{-1}(\cdot)$ is strictly increasing. What this says then, is, holding fixed the preference for vote share b^s, the ratio of the rate at which the party chooses type k in constituency 1 relative to the same rate in constituency 2 is larger than the ratio of returns in vote share associated with that type. This gap must be explained by a positive preference for win probability b^w.

Similarly, $b^s > 0$ if and only if:

$$\frac{\Lambda^{-1}(P(k,\mathbf{z}^{(2)}))}{\Lambda^{-1}(P(k,\mathbf{z}^{(1)}))} > \frac{\Delta_w^P(k,\mathbf{z}^{(2)})}{\Delta_w^P(k,\mathbf{z}^{(1)})}$$

The intuition is similar here. We have assumed

$$\frac{\Delta_w^P(k,\mathbf{z}^{(1)})}{\Delta_w^P(k,\mathbf{z}^{(2)})} > \frac{\Delta_s^P(k,\mathbf{z}^{(1)})}{\Delta_s^P(k,\mathbf{z}^{(2)})}$$

so that the returns to selecting type k in *expected vote share* are relatively higher in constituency 2 than constituency 1. So $b^s > 0$ if, holding fixed the preference for win probability b^w, the ratio of the rate at which the party chooses type k in constituency 1 relative to the same rate in constituency 2 is larger than the ratio of returns in win probability associated with that type. This gap must be explained by a positive preference for expected vote share b^s.

One of our key findings is that $b^w > 0 > b^s$. Given the derivations above, this means that:

1. The party chooses type k at a relatively higher rate than can be explained by returns to expected vote share, so $b^w > 0$.

2. The party chooses type k at a relatively *lower* rate than can be explained by win probability, so $b^s < 0$

Adding costs to the payoff function is straightforward. In this case the inverted choice probability satisfies:

$$\begin{aligned} \Lambda^{-1}\big(P(k,\mathbf{z})\big) &= \ln\big(P(k,\mathbf{z})\big) - \ln\big(P(K,\mathbf{z})\big) \\ &= b^w \times (w^P(k,\mathbf{z}) - w^P(K,\mathbf{z})) + b^s \times (s^P(k,\mathbf{z}) - s^P(K,\mathbf{z})) + c_k - c_K \end{aligned} \tag{6}$$

Now, consider two values of $\mathbf{z}$, say $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$. Differencing (6) across these two values we eliminate the cost parameters, assumed to be independent of $\mathbf{z}$, and get:

$$\begin{aligned}\Lambda^{-1}\big(P(k,\mathbf{z}^{(1)})\big) - \Lambda^{-1}\big(P(k,\mathbf{z}^{(2)})\big) &= b^w \times \big(\Delta_w^P(k,\mathbf{z}^{(1)}) - \Delta_w^P(k,\mathbf{z}^{(2)})\big) \\ &\quad + b^s \times \big(\Delta_s^P(k,\mathbf{z}^{(1)}) - \Delta_s^P(k,\mathbf{z}^{(2)})\big)\end{aligned}$$

or:

$$\Lambda^{-1}\big(P(k,\mathbf{z}^{(1)})\big) - \Lambda^{-1}\big(P(k,\mathbf{z}^{(2)})\big) = b^w \times \Delta_w^P(k,\mathbf{z}^{(1,2)}) + b^s \times \Delta_s^P(k,\mathbf{z}^{(1,2)}) \tag{7}$$

We can do the same at another pair of value of $\mathbf{z}$, say $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(3)}$, which then gives us a system of two equations and two unknown parameters (b^w, b^s). Defining $\Delta_\Lambda^P(k,\mathbf{z}^{(1,2)}) \equiv \Lambda^{-1}\big(P(k,\mathbf{z}^{(1)})\big) - \Lambda^{-1}\big(P(k,\mathbf{z}^{(2)})\big)$ and $\Delta_\Lambda^P(k,\mathbf{z}^{(1,3)})$ similarly, we can write the system as:

$$\begin{bmatrix}\Delta_\Lambda^P(k,\mathbf{z}^{(1,2)}) \\ \Delta_\Lambda^P(k,\mathbf{z}^{(1,3)})\end{bmatrix} = \begin{bmatrix}\Delta_w^P(k,\mathbf{z}^{(1,2)}) & \Delta_s^P(k,\mathbf{z}^{(1,2)}) \\ \Delta_w^P(k,\mathbf{z}^{(1,3)}) & \Delta_s^P(k,\mathbf{z}^{(1,3)})\end{bmatrix}\begin{bmatrix}b^w \\ b^s\end{bmatrix} \tag{8}$$

The parameters (b^w, b^s) are identified if the matrix $\begin{bmatrix}\Delta_w^P(k,\mathbf{z}^{(1,2)}) & \Delta_s^P(k,\mathbf{z}^{(1,2)}) \\ \Delta_w^P(k,\mathbf{z}^{(1,3)}) & \Delta_s^P(k,\mathbf{z}^{(1,3)})\end{bmatrix}$ has full rank. This is the case if, for at least one triple $(\mathbf{z}^1, \mathbf{z}^2, \mathbf{z}^3)$:

$$\frac{\Delta_w^P(k,\mathbf{z}^{(1,2)})}{\Delta_w^P(k,\mathbf{z}^{(1,3)})} \neq \frac{\Delta_s^P(k,\mathbf{z}^{(1,2)})}{\Delta_s^P(k,\mathbf{z}^{(1,3)})}$$

This is the key condition for identification of (b^w, b^s), analogous to the condition in Equation 3 in the simpler case above, except we now need variation in win probability and expected vote share across more values of $\mathbf{z}$ - this is how we identify b^w, b^s separately from the cost difference $c_k - c_K$.

Solving the system in (8) we get expressions analogous to what we had in the simple case with no costs:

$$b^w = \frac{\Delta_s^P(k,\mathbf{z}^{(1,2)})\Delta_\Lambda^P(k,\mathbf{z}^{(1,3)}) - \Delta_s^P(k,\mathbf{z}^{(1,3)})\Delta_\Lambda^P(k,\mathbf{z}^{(1,2)})}{\Delta_s^P(k,\mathbf{z}^{(1,2)})\Delta_w^P(k,\mathbf{z}^{(1,3)}) - \Delta_s^P(k,\mathbf{z}^{(1,3)})\Delta_w^P(k,\mathbf{z}^{(1,2)})} \tag{9}$$

and:

$$b^s = \frac{\Delta_w^P(k,\mathbf{z}^{(1,3)})\Delta_\Lambda^P(k,\mathbf{z}^{(1,2)}) - \Delta_w^P(k,\mathbf{z}^{(1,2)})\Delta_\Lambda^P(k,\mathbf{z}^{(1,3)})}{\Delta_s^P(k,\mathbf{z}^{(1,2)})\Delta_w^P(k,\mathbf{z}^{(1,3)}) - \Delta_s^P(k,\mathbf{z}^{(1,3)})\Delta_w^P(k,\mathbf{z}^{(1,2)})} \tag{10}$$

With b^w and b^s identified we can now obtain an expression for the cost difference $c_k - c_K$

by substituting the identified values into (6) to get:

$$c_k - c_K = \Lambda^{-1}\big(P(k,\mathbf{z})\big) - b^w \times (w^P(k,\mathbf{z}) - w^P(K,\mathbf{z})) - b^s \times (s^P(k,\mathbf{z}) - s^P(K,\mathbf{z}))$$

where b^w, b^s are as defined in (9) and (10), respectively.

4 Candidate wealth and party payoffs

To illustrate the importance of wealth as a determinant of candidate choice, we compute parties' payoff from the candidates chosen in the data. The upper panel of Figure A.5 shows these (normalized) payoffs ordered by constituency and candidate type for the UPA. Lighter colors correspond to higher payoffs. On the lower panel, we present the same information but ordering candidates by wealth, regardless of type. Figure A.6 shows corresponding graphs for the NDA. These graphs reveal that the wealth of the candidate plays an important role in increasing the payoff of political parties.

5 Model fit and validation

Here we provide further results about how model fit depends on the inclusion of party preferences over candidates.

In Table A.4 we present the analogue of Table 7 in the paper but in a model that assumes parties only care about voter preferences (and thus the probability of winning). This is the model estimated in the first column of Table 6 in the paper.

Table A.4: Model fit with no cost parameters

	UPA actual	UPA predicted	NDA actual	NDA predicted	All actual	All predicted
Type 1	217	106.23	229	106.14	446	212.37
Type 2	24	107.10	43	108.95	64	216.05
Type 3	49	105.39	22	103.76	71	209.15
Type 4	144	115.28	140	115.15	287	230.43

Notes: Number of candidates of each type observed in the data and predicted by the model (average across 100 simulations) using the estimates from column 1 of Table 6 in the paper. Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively.

When parties are restricted to care only about voter preferences, the model considerably under-predicts the selection of the educated type (type 1) and over-predicts the other types, in particular the uneducated type (type 2) and the Muslim type (type 3).

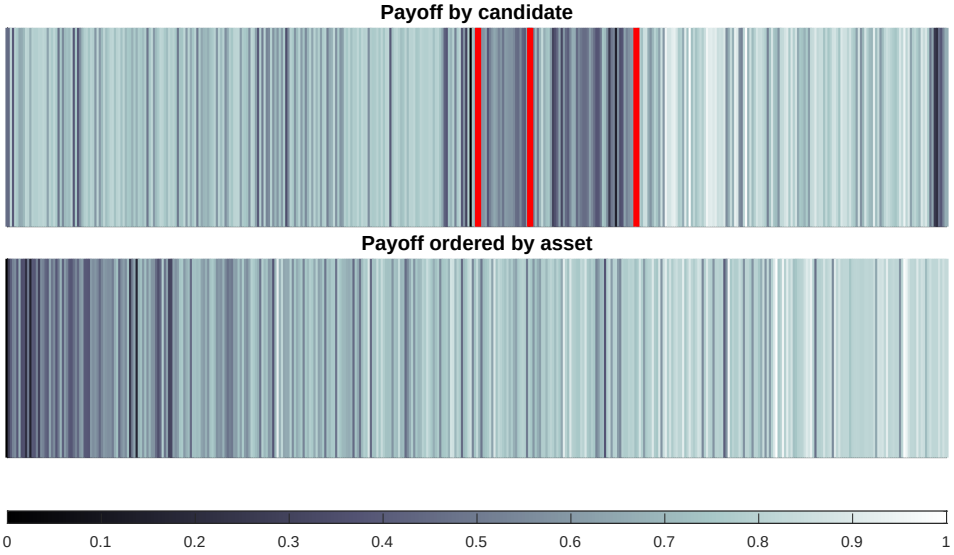


Figure A.5: Candidate wealth and the UPA's payoffs

The upper panel shows the payoffs of candidates chosen in the data grouped by type. Types are separated by the red vertical lines. Payoffs are normalized by subtracting the minimum value and dividing by the range. Lighter colors indicate higher payoffs. The lower panel shows the same payoffs when candidates are ordered by wealth, from lowest to highest.

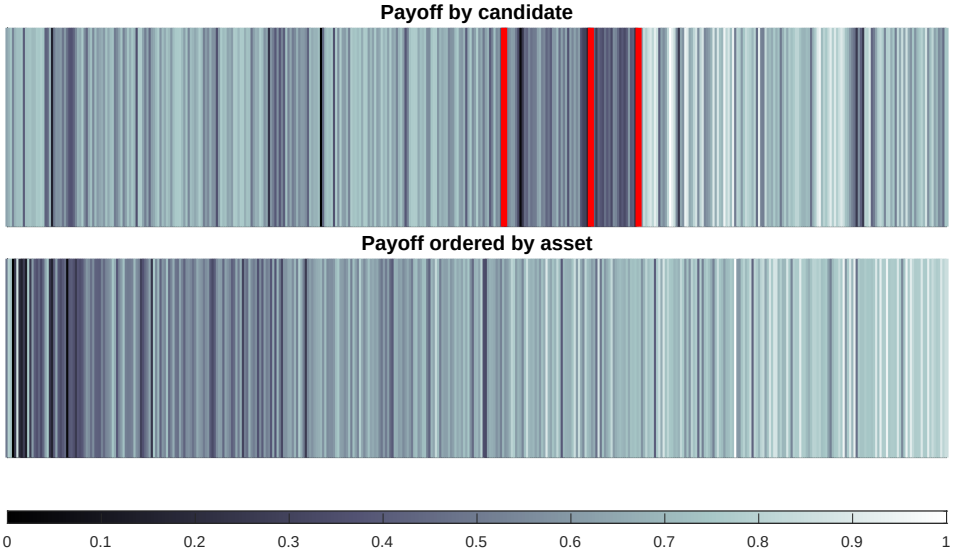


Figure A.6: Candidate wealth and the NDA's payoffs

The upper panel shows the payoffs of candidates chosen in the data grouped by type. Types are separated by the red vertical lines. Payoffs are normalized by subtracting the minimum value and dividing by the range. Lighter colors indicate higher payoffs. The lower panel shows the same payoffs when candidates are ordered by wealth, from lowest to highest.

6 Clustering

6.1 Details of k-means clustering

Let the characteristics of candidate i be given by $\mathbf{x}_i$, and for a given value of K, denote the centroid of cluster C_k by b_k, $k = 1, ..., K$. The Within-Cluster Sum of Squares is then

$$WCSS(K) = \sum_{k=1}^{K} \sum_{i \in C_k} ||\mathbf{x}_i - b_k||^2$$

To define the Silhouette Coefficient, let $C(i)$ denote the cluster of candidate i, and define:

$$S_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}$$

where

$$a_i = \frac{1}{|C(i)| - 1} \sum_{j \neq i, j \in C(i)} ||\mathbf{x}_i - \mathbf{x}_j||$$

is the average distance between i and all other points in the same cluster, and

$$b_i = \min_{k: C_k \neq C(i)} \frac{1}{|C_k|} \sum_{j \in C_k} ||\mathbf{x}_i - \mathbf{x}_j||$$

is the average distance between i and the points in the cluster nearest to i other than the one it was assigned to. Notice that $S_i \in [-1, +1]$, and a value of $S_i \simeq 1$ implies that i is close to other points in its assigned cluster, and far from other clusters, while a value of $S_i \simeq -1$ implies that i is close to points in other clusters relative to points in its own. The Silhouette Coefficient is the average over all candidates in the sample:

$$SC(K) = \frac{1}{N} \sum_i S_i$$

Note that, unlike the WCSS, the Silhouette Coefficient need not be monotonic in K.

6.2 Alternative clustering procedure

In this section we show that a common alternative to the k-means approach for clustering data, Hierarchical Clustering Analysis (HCA), yields clusters that are similar to what we obtain using k-means, but that the k-means approach we take in the paper yields superior validation results.

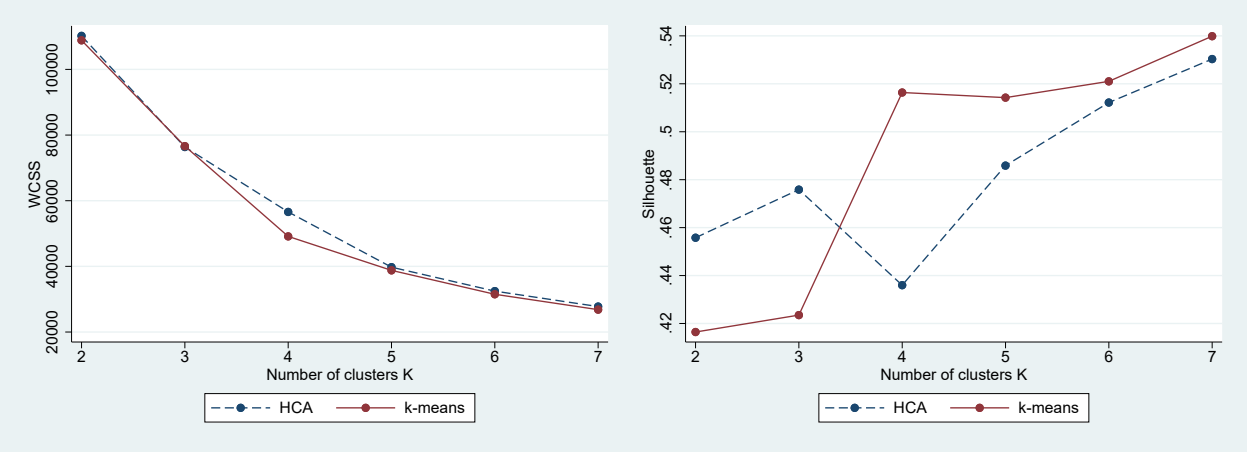


Figure A.7: WCSS and Silhouette: k-means vs HCA

Conceptually HCA, begins by treating each of the N individual observation as its own "cluster," then iteratively agglomerates the clusters.[4] In the first iteration, we take the two points that are closest together in the sense that they have the lowest variance among any pair of points in the sample (the Ward merging criterion), and then combine those to form a new cluster. We now have $N-1$ clusters. Then again, we find the two clusters that are closest together in terms of variance and combine those, and so on. The number of clusters is determined by where we stop this iterative agglomeration process.

First, on the left panel of Figure A.7 we compare the within-cluster sum of squares, WCSS, over a number of possible clusters across the two methods. The k-means approach is uniformly better by this metric (i.e., produces lower WCSS scores).[5]

On the right panel of Figure A.7 we compare the Silhouette score associated with the two approaches across a range of possible numbers of clusters. While HCA performs better at low number of clusters (2 or 3), k-means performs better from 4 clusters on. Moreover, k-means scores with 4 or more clusters dominate any HCA score below 4 clusters. Finally, note that we only ever hit a Silhouette score of 0.5 with k-means and the number of clusters at least 4. In cluster analysis, a Silhouette score of 0.5 is often used as a rule of thumb for "good clustering" (Rousseeuw 1987).

These patterns indicate that k-means is preferable to HCA for our application. We nevertheless computed the types that the HCA algorithm predicts when we set $k = 4$. The types are in Table A.8.

The types recovered by HCA look remarkably similar to what we obtain using k-means.

[4] This is hierarchical *agglomerative* clustering. A much less used alternative that we do not consider here is hierarchical divisive clustering.

[5] Note that k-means is based on "centroids" (the average of all points in the cluster), while HCA is not. To generate Figure A.7, we compute the "centroid" of each HCA cluster analogously.

Table A.8: Centroids of candidate types using HCA

	Assets	Crimes	Education	Muslim
Type 1	13.68	0.00	0.79	0.00
Type 2	15.21	0.00	0.00	0.00
Type 3	14.30	0.17	0.44	1.00
Type 4	15.06	1.00	0.68	0.00

Notes: Centroids resulting from the HCA clustering algorithm with $k = 4$. The algorithm is run on standardized variables; the table shows the centroids transformed back to the original scale for ease of interpretation.

Similar to k-means, HCA clusters all Muslim candidates together, and then among the remaining candidates clusters all criminals together. The one difference is that the HCA algorithm puts some uneducated individuals into Type 1 so that there is a larger gap in assets between the Type 1 and Type 2 candidates than there is using k-means.

7 Criminals as strategic complements

Table A.9: Correlation of estimated choice probabilities

		UPA			
		Type 1	Type 2	Type 3	Type 4
NDA	Type 1	0.27	0.12	0.12	-0.29
	Type 2	0.18	0.42	-0.10	-0.14
	Type 3	0.03	-0.12	0.26	-0.16
	Type 4	-0.26	-0.15	-0.14	0.30

Notes: Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively.

Table A.9 shows the correlation between the two parties' estimated choice probabilities for the different candidate types. The correlation between the two parties' choice probabilities for criminals is positive, while the correlation between criminals and other types is always negative. This provides suggestive evidence of strategic complementarity between criminal candidates.

To obtain a precise measure of strategic complementarity, write party p's choice probability associated with choosing some type a' as

$$P_p(a') = \frac{\exp\left\{\tilde{U}_p(a', P)\right\}}{\sum_{a=1}^{4} \exp\left\{\tilde{U}_p(a, P)\right\}}$$

where $\tilde{U}_p(a, P) = E_P[b^w w_p(a) + b^s s_p(a)] + c_p(a)$. Letting $\Delta^P(a) \equiv \tilde{U}_p(a, P) - \tilde{U}_p(4, P)$, we

have

$$P_p(4) = \frac{1}{1 + \sum_{a=1}^{3} \exp\left\{\Delta^P(a)\right\}}$$

Differentiating with respect to the opponent's probability of choosing type 4, we get

$$\begin{aligned} \frac{\partial P_p(4)}{\partial P_{-p}(4)} &= -\sum_{a=1}^{3} \frac{\exp\left\{\Delta^P(a)\right\} \frac{\partial \Delta^P(a)}{\partial P_{-p}(4)}}{\left(1 + \sum_{a=1}^{3} \exp\left\{\Delta^P(a)\right\}\right)^2} \\ &= -\sum_{a=1}^{3} P_p(4) P_p(a) \frac{\partial \Delta^P(a)}{\partial P_{-p}(4)} \end{aligned} \tag{11}$$

We have

$$\frac{\partial \Delta^P(a)}{\partial P_{-p}(4)} = b^w \frac{\partial E_P[w_p(a) - w_p(4)]}{\partial P_{-p}(4)} + b^s \frac{\partial E_P[s_p(a) - s_p(4)]}{\partial P_{-p}(4)}.$$

where

$$\frac{\partial E_P[w_p(a) - w_p(4)]}{\partial P_{-p}(4)} = w_p(a,4) - w_p(4,4) + \sum_{t=1}^{3} \left[w_p(a,t) - w_p(4,t)\right] \frac{\partial P_{-p}(t)}{\partial P_{-p}(4)}, \tag{12}$$

and similarly for $\frac{\partial E_P[s_p(a) - s_p(4)]}{\partial P_{-p}(4)}$.

Using our parameter estimates, we compute (11) for both parties in every constituency. We do this by assuming that $\frac{\partial P_{-p}(t)}{\partial P_{-p}(4)}$ in (12) is the same for $t = 1, 2, 3$ (if we instead set $\frac{\partial P_{-p}(t)}{\partial P_{-p}(4)} = -1$ for some t we obtain very similar results). A positive derivative $\frac{\partial P_p(4)}{\partial P_{-p}(4)}$ indicates strategic complementarity between criminal candidates.

The result is shown on Figure 3 in the paper: the derivative is positive for both parties in 77% of the constituencies, and positive for at least one of the parties in 94%.

8 Other robustness checks and counterfactuals

8.1 Robustness

In this section we consider a series of other checks of the robustness of our main estimates in Table 6 in the paper.

8.1.1 Alternative payoff specifications

First, In Table A.10 we allow for additional non-linearities in party preferences over vote shares (of course, the winning probability itself is a highly non-linear function of vote shares). As a benchmark, in column 1 we reproduce the estimates from the baseline model. In column

Table A.10: Party objective function estimates - Further non-linearities in vote shares

	(1)	(2)	(3)		(1)	(2)	(3)
b^w	3.24	2.64	2.59	c_3^{educ}	-0.63	-0.62	-0.61
	(1.02)	(1.09)	(1.09)		(0.34)	(0.35)	(0.35)
b^s	-14.15	-9.51	-12.91	c_4^{educ}	-0.41	-0.43	-0.42
	(1.73)	(2.95)	(4.54)		(0.23)	(0.22)	(0.22)
b^s_{sq}		-5.06	5.31	c_2^{crime}	-0.03	-0.03	-0.03
		(2.65)	(10.95)		(0.24)	(0.24)	(0.24)
b^s_{cu}			-10.13	c_3^{crime}	0.36	0.36	0.34
			(10.53)		(0.23)	(0.23)	(0.23)
$c^0_{2,NDA}$	-1.57	-1.57	-1.57	c_4^{crime}	1.10	1.10	1.10
	(0.27)	(0.27)	(0.27)		(0.15)	(0.15)	(0.15)
$c^0_{3,NDA}$	-2.58	2.49	2.54	c_2^{asset}	0.32	0.32	0.32
	(0.30)	(0.31)	(0.31)		(0.26)	(0.25)	(0.23)
$c^0_{4,NDA}$	0.58	0.43	0.53	c_3^{asset}	-0.12	-0.11	-0.11
	(0.23)	(0.25)	(0.27)		(0.26)	(0.26)	(0.26)
$c^0_{2,UPA}$	-2.09	-2.09	-2.09	c_4^{asset}	-0.06	-0.10	-0.09
	(0.32)	(0.32)	(0.32)		(0.16)	(0.17)	(0.17)
$c^0_{3,UPA}$	-1.93	-1.85	-1.90	c_2^{Muslim}	-0.07	-0.07	-0.08
	(0.29)	(0.29)	(0.30)		(0.41)	(0.41)	(0.41)
$c^0_{4,UPA}$	0.40	0.27	0.38	c_3^{Muslim}	1.48	1.46	1.46
	(0.23)	(0.24)	(0.26)		(0.29)	(0.29)	(0.29)
c_2^{educ}	-0.53	-0.53	0.53	c_4^{Muslim}	0.03	0.02	0.03
	(0.34)	(0.34)	(0.34)		(0.24)	(0.24)	(0.24)
Log likelihood					-836.21	-834.16	-833.62

Notes: Estimates of party objective functions allowing for further non-linearities in vote share. Column 1 reproduces column 3 from Table 6 in the paper. Column 2 adds the expected value of s^2 and column 3 also the expected value of s^3, with corresponding parameters b^s_{sq} and b^s_{cu}. Standard errors in parentheses. Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively. Number of markets: 434.

2 we include the expected squared vote share, and in column 3 we add the expected cubic vote share. We find that the linear term is the only one that is always statistically significant. In addition, in column 2 the quadratic term is significant at 10%. Figure A.11 shows the equilibrium choice probabilities implied by each specification. Equilibrium behavior is virtually identical, which further supports our focus on the more parsimonious specification.

Second, in Table A.12 we replace the expected vote share in parties' objective functions with their expected vote share margin. For winners, this is the difference compared to the runner-up; for losers, it is the difference compared to the winner (which is negative). Qualitatively there is no change in the results. Parties have a preference for winning, but are averse to bigger wins as measured by the vote share margin. Third, we consider a model where instead of expected share, parties' payoff is affected by the expected number of votes

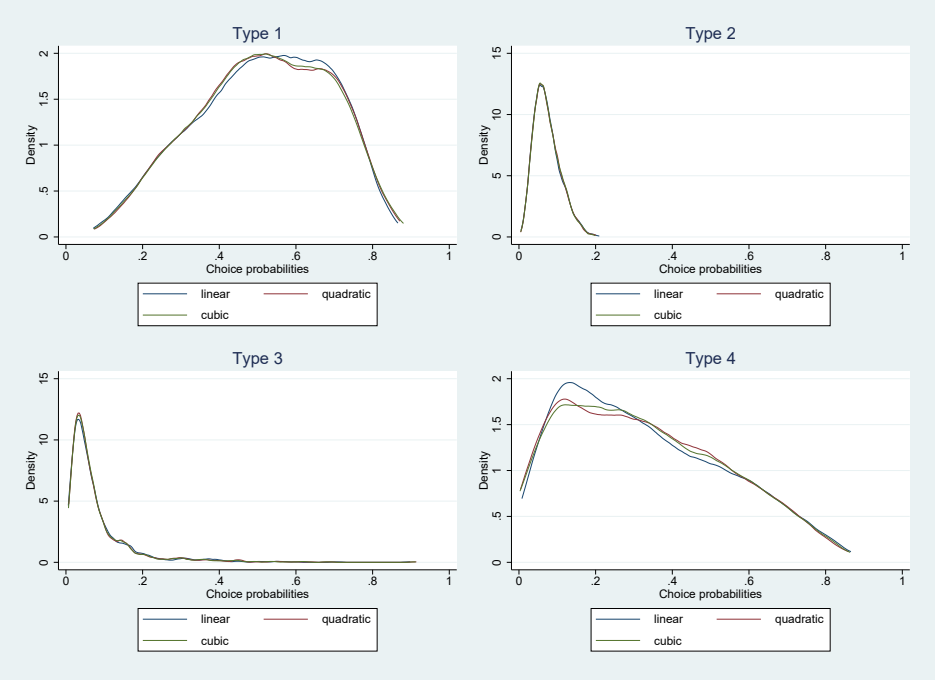


Figure A.11: Distribution of choice probabilities with and without additional non-linearities

Kernel density plots of equilibrium choice probabilities corresponding to the specifications in Table A.10. Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively.

received. This also introduces additional cross-constituency variation depending on the size of the constituency. The estimates are in Table A.13, and they again show no qualitative difference relative to our main results.

8.1.2 Constituency-level heterogeneity

In Table A.14 we study the robustness of our payoff function estimates to various types of observable and unobservable heterogeneity. In column 1 we consider the possibility that selection decisions are systematically different across election years (2009 and 2014). The two elections were different - the 2014 election was a breakthrough year for the BJP (and NDA more generally) and marked the end of the INC's near continuous hold on power since the birth of the Indian state. As such, we allow for the possibility that the payoffs to different types of candidates might vary across the two years. We let c_2^*, c_3^*, c_4^* denote the interaction between the type dummies and an indicator for election year 2014. The estimates suggest that parties were less likely to select type 2 and type 4 candidates in 2014 than they were in the 2009 election. However, there is little change in other estimates relative to column 3 of Table 6 in the paper.

Second, India's constitution allows for reservation of some electoral constituencies for

Table A.12: Party objective function estimates: Expected Margin

	(1)	(2)	(3)		(1)	(2)	(3)
b^w	4.06	2.37	2.24	c_4^{educ}		-0.60	-0.38
	(0.74)	(0.85)	(0.90)			(0.20)	(0.22)
b^m	-3.84	-6.11	-9.65	c_2^{crime}		-0.47	-0.03
	(0.80)	(0.96)	(1.15)			(0.22)	(0.24)
$c_{2,NDA}^0$			-1.56	c_3^{crime}		-0.31	0.37
			(0.27)			(0.21)	(0.22)
$c_{3,NDA}^0$			-2.60	c_4^{crime}		1.04	1.08
			(0.30)			(0.14)	(0.15)
$c_{4,NDA}^0$			0.60	c_2^{asset}		-0.74	0.30
			(0.24)			(0.19)	(0.25)
$c_{2,UPA}^0$			-2.08	c_3^{asset}		-1.04	-0.04
			(0.32)			(0.20)	(0.26)
$c_{3,UPA}^0$			-1.98	c_4^{asset}		0.02	-0.14
			(0.29)			(0.14)	(0.17)
$c_{4,UPA}^0$			0.47	c_2^{Muslim}		0.44	-0.07
			(0.23)			(0.32)	(0.41)
c_2^{educ}		-2.00	-0.53	c_3^{Muslim}		1.71	1.45
		(0.26)	(0.34)			(0.27)	(0.28)
c_3^{educ}		-2.32	-0.65	c_4^{Muslim}		0.15	0.05
		(0.26)	(0.34)			(0.24)	(0.24)
Log likelihood					-1187.20	-920.22	-835.75

Notes: Estimates of party objective functions with expected vote share replaced by the expected vote margin (the difference compared to the runner-up for winners, and the difference compared to the winner for losers). Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively. Number of markets: 434.

Table A.13: Party objective function estimates: Expected number of votes

	(1)	(2)	(3)		(1)	(2)	(3)
b^w	3.74	3.29	3.05	c_4^{educ}		-0.64	-0.36
	(0.83)	(0.94)	(1.00)			(0.20)	(0.22)
b^t	-0.32	-0.70	-0.98	c_2^{crime}		-0.46	-0.03
	(0.09)	(0.11)	(0.12)			(0.22)	(0.24)
$c_{2,NDA}^0$			-1.59	c_3^{crime}		-0.31	0.34
			(0.27)			(0.21)	(0.23)
$c_{3,NDA}^0$			-2.51	c_4^{crime}		1.07	1.12
			(0.30)			(0.14)	(0.15)
$c_{4,NDA}^0$			0.45	c_2^{asset}		-0.73	0.34
			(0.23)			(0.19)	(0.26)
$c_{2,UPA}^0$			-2.12	c_3^{asset}		-1.10	-0.21
			(0.30)			(0.20)	(0.26)
$c_{3,UPA}^0$			-1.84	c_4^{asset}		0.16	-0.14
			(0.29)			(0.14)	(-0.14)
$c_{4,UPA}^0$			0.23	c_2^{Muslim}		0.44	-0.08
			(0.22)			(0.32)	(0.41)
c_2^{educ}		-2.02	-0.53	c_3^{Muslim}		1.75	1.48
		(0.26)	(0.34)			(0.27)	(0.28)
c_3^{educ}		-2.31	-0.66	c_4^{Muslim}		0.11	0.03
		(0.26)	(0.35)			(0.24)	(0.24)
Log likelihood					-1191.70	-916.90	-837.36

Notes: Estimates of party objective functions with expected vote share replaced by the expected number of votes. Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively. Number of markets: 434.

Table A.14: Party objective function estimates: Constituency-level heterogeneity

	(1)	(2)	(3)	(4)		(1)	(2)	(3)	(4)
b^w	2.91	3.24	3.43	3.29	c_2^{educ}	-0.46	-0.55	-0.56	-0.45
	(1.02)	(1.02)	(1.03)	(1.03)		(0.34)	(0.35)	(0.35)	(0.35)
b^s	-14.43	-13.90	-14.45	-14.00	c_3^{educ}	-0.65	-0.60	-0.60	-0.60
	(1.73)	(1.74)	(1.75)	(1.75)		(0.35)	(0.34)	(0.35)	(0.35)
c_2^*	-0.76	0.23	0.57	-	c_4^{educ}	-0.37	-0.40	-0.46	-0.40
	(0.33)	(0.29)	(1.40)			(0.22)	(0.22)	(0.22)	(0.22)
c_3^*	0.19	-0.69	-0.92	-	c_2^{crime}	0.00	-0.02	-0.04	-0.01
	(0.32)	(0.33)	(1.30)			(0.24)	(0.24)	(0.25)	(0.24)
c_4^*	-0.61	-0.37	1.36	-	c_3^{crime}	0.35	0.37	0.40	0.37
	(0.21)	(0.20)	(0.87)			(0.23)	(0.23)	(0.24)	(0.23)
$c_{2,NDA}^0$	-1.41	-1.66	-2.04	-1.94	c_4^{crime}	1.12	1.08	1.06	1.08
	(0.28)	(0.29)	(1.19)	(0.79)		(0.15)	(0.15)	(0.15)	(0.15)
$c_{3,NDA}^0$	-2.67	-2.37	-1.83	-2.08	c_2^{asset}	0.70	0.36	0.36	0.23
	(0.32)	(0.31)	(1.11)	(0.95)		(0.31)	(0.26)	(0.27)	(0.27)
$c_{4,NDA}^0$	0.82	0.66	-0.53	1.22	c_3^{asset}	-0.21	-0.19	-0.18	-0.20
	(0.25)	(0.24)	(0.74)	(0.57)		(0.31)	(0.26)	(0.28)	(0.28)
$c_{2,UPA}^0$	-2.02	-2.20	-2.57	-2.46	c_4^{asset}	0.24	-0.10	0.02	-0.08
	(0.32)	(0.34)	(1.21)	(0.81)		(0.20)	(0.17)	(0.17)	(0.17)
$c_{3,UPA}^0$	-1.99	-1.70	-1.16	-1.40	c_2^{Muslim}	-0.06	-0.04	-0.07	-0.07
	(0.29)	(0.30)	(1.12)	(0.57)		(0.42)	(0.41)	(0.41)	(0.42)
$c_{4,UPA}^0$	0.60	0.51	-0.73	1.07	c_3^{Muslim}	1.45	1.39	1.48	1.37
	(0.24	(0.23)	(0.76)	(0.57)		(0.29)	(0.29)	(0.29)	(0.29)
					c_4^{Muslim}	0.00	-0.02	0.05	0.00
						(0.25)	(0.24)	(0.24)	(0.25)
Log likelihood						-829.48	-831.80	-834.40	-829.69

Notes: Estimates of party objective functions allowing for additional heterogeneity. Columns 1-3 interact candidate types with the following variables, respectively: 1: indicator for election year 2014; 2: indicator for reserved constituencies; 3: rural population. The parameters on the interactions are denoted c_2^*, c_3^*, c_4^*, respectively. In column 4, we include interactions with fixed effects for constituency "types" following Bonhomme et al. (2022). Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively. Number of markets: 434.

Scheduled Castes (SC) and Tribes (ST). In these constituencies all candidates must be from one of the designated minority groups. As the pool of candidates that parties can choose from might vary in an important way between reserved and unreserved constituencies, we allow for costs to depend on reservation status in column 2. The parameters c_2^*, c_3^*, c_4^* now denote the interaction of the type dummies with an indicator for reserved constituencies. The only estimate that is significantly different from 0 (at the 5% level) is c_3^* - the additional cost of selecting the Muslim type in a reserved constituency. This is not surprising as casteism is largely a Hindu phenomenon. Again, none of our estimates is affected by allowing for this type of heterogeneity.

In column 3 we allow for the possibility that selection decisions may depend on the rural/urban split in the constituency. The parameters c_2^*, c_3^*, c_4^* now denote the interaction of the type dummies with the share of rural population. Allowing for these differences causes little change in our estimates.

Finally, it is possible that constituency level *unobservable* factors - fixed effects - confound our estimates. As an example, it could be that parties systematically avoid picking criminal candidates in precisely the constituencies where they are most popular for some unobservable reason. Including a full set of constituency dummies in the model is not an option, however. There are 217 electoral constituencies in our data, and we would need to interact these dummies with $K-1$ of the K type choices, which would imply 651 additional parameters in our model with $K = 4$. The resulting incidental parameters problem makes this infeasible.

As an alternative we consider the approach of Bonhomme et al. (2022), who propose to reduce the dimensionality of the fixed effects by clustering units based on observable characteristics into groups in a first step. Applied to our context, the idea is that constituencies in the same cluster likely have similar *unobservables*. We use kmeans as proposed by Bonhomme et al. (2022) to cluster constituencies. The observables we include are the constituency level demographic characteristics in Table A.1: Rural population share, Literate population share, ST and SC reservation status, Population with paved roads, Working population and Rural working population. The algorithm identifies four constituency types, and we include dummies for these types (interacted with candidate types) and re-estimate the model in column 4 of Table A.14. Controlling for constituency level unobserved heterogeneity yields virtually identical b^w and b^s estimates to our main specification, and causes little change in the cost parameters.

Table A.15: Different types' probability of being chosen when $b^s = 0$, with and without a criminal ban

	Baseline	Setting $b^s = 0$	Baseline ban	Ban with $b^s = 0$
All candidates				
Type 1	51.4	22.5	76.8	79.7
Type 2	7.4	3.6	11.5	13.5
Type 3	8.2	2.0	11.7	6.8
Type 4	33.1	72.0	0.0	0.0
UPA				
Type 1	50.0	25.2	75.4	80.2
Type 2	5.5	3.0	8.7	10.1
Type 3	11.3	3.0	15.9	9.7
Type 4	33.2	68.8	0.0	0.0
NDA				
Type 1	52.8	19.9	78.2	79.2
Type 2	9.2	4.1	14.4	17.0
Type 3	5.1	0.9	7.4	3.8
Type 4	32.9	75.1	0.0	0.0

Notes: Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively. Values shown are the averages across all the constituencies in the data. The first column reproduces the baseline equilibrium from the paper. The next column replaces the b^s parameter estimate with 0. The third column reproduces the counterfactual criminal ban from the paper. The last column shows the results when $b^s = 0$ and criminals are banned.

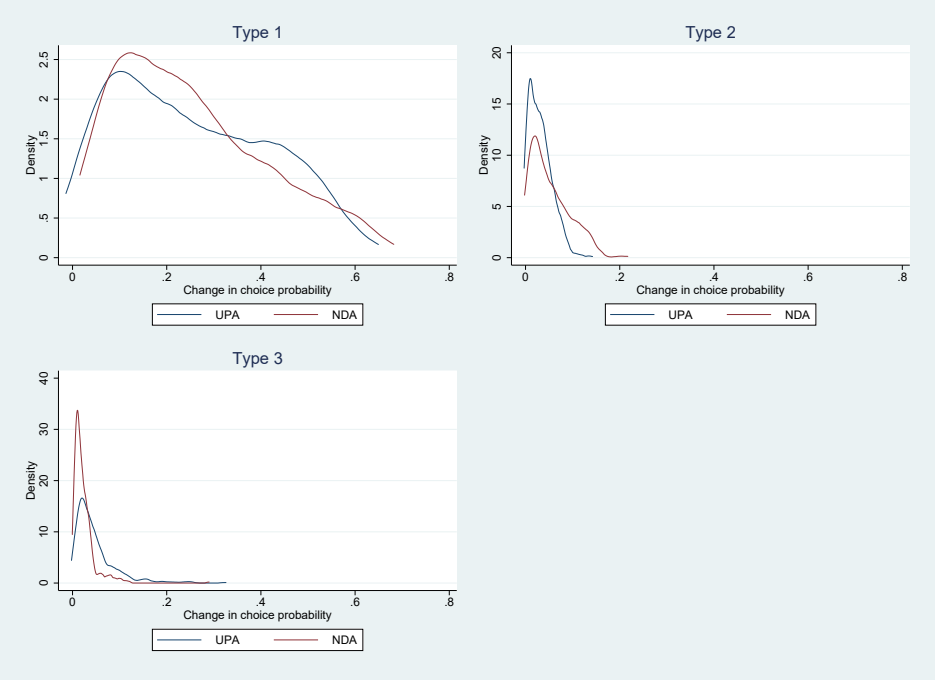


Figure A.16: Distribution of changes in types' choice probabilities after criminal ban

Kernel density plots of the change in choice probabilities for each type as a result of a criminal ban (counterfactual choice probability minus baseline choice probability). Types 1-3 are the Educated, Uneducated, and Muslim types, respectively.

8.2 Additional counterfactual results

In the paper we find that $b^s < 0$, so that parties receive a disutility from candidates with higher expected vote shares (for given winning probability). This result is qualitatively robust across the many specifications we consider. In Table A.15 we present the results of a counterfactual experiment where we set $b^s = 0$ to get a sense of the implications of this finding for different types.

The most striking result is that criminal candidates are more than twice as likely to be chosen in this counterfactual scenario than at baseline. At baseline, the parties each choose criminal candidates 33% of the time, and in the counterfactual scenario the UPA chooses criminals 69% of the time while the NDA chooses criminals 75% of the time.

Next, we repeat our counterfactual exercise of a criminal ban combined with $b^s = 0$. The resulting choice probabilities are summarized in the last column of Table A.15, and the third column shows the counterfactual results from the paper for comparison. As can be seen, under a ban, parties' behavior is much less dependent on party organization as reflected by the value of b^s.

References

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2022): "Discretizing unobserved heterogeneity," *Econometrica*, 90 (2), 625–643.

Gandhi, Amit and Jean-François Houde (2019): "Measuring substitution patterns in differentiated-products industries," *NBER Working paper w26375*.

Rousseeuw, P.J. (1987): "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Computational and Applied Mathematics*, 20, 53–65.