

Who Wins Votes?

Candidate Characteristics in Indian Elections

Adam Ziegfeld¹

February 2014

Abstract

Are certain kinds of candidates more likely to do better in elections than others? Using a unique dataset on the characteristics of candidates in elections in the north Indian state of Haryana, this study finds that certain candidate characteristics are, indeed, associated with greater vote shares. In particular, candidates who previously held high-level state or national office or who had family in politics who held such positions tend to win substantially higher vote shares than those who do not. Additionally, candidates who are natives of the constituencies from which they contest, previously held local office or who had family who held local office, or have occupations in business also tend to do somewhat better than other candidates. Incumbency, gender, membership in a large caste, and occupations in agriculture and the professions are not associated with a candidate's vote share.

¹ Visiting Assistant Professor of Political Science, Beloit College. Email: ziegfelda@beloit.edu. I would like to acknowledge the financial support of the John Fell OUP Research Fund and British Academy Small Grant Scheme. I gratefully thank the dedicated and excellent research assistance of Jyoti, Kashmir Dhankhar, Sankar Prasad Duria, Jyoti Mishra, and Rohit Sharma. Georgia Kernell, Michele Margolis, Irfan Nooruddin, seminar participants at Nuffield College, the Northwestern Parties Working Group, and participants at the EECURI network meeting in Bangalore provided comments on an earlier version of this paper. All errors are my own.

What kinds of candidate are successful in elections? When analyzing electoral outcomes, political scientists and casual observers alike often talk in terms of political parties and their ability to win popular support. However, election outcomes depend not only on the parties competing in elections but also on the candidates. Intuitively, candidate quality seems as though it should matter in determining electoral outcomes, with high quality candidates winning greater vote shares than low quality candidates. Unfortunately, candidate quality is difficult to measure and often relies on subjective assessments that may vary considerably across voters. Nevertheless, are there objective candidate characteristics that are associated with greater electoral success? In other words, do candidates that have certain attributes tend to be more electorally successful than those who do not?

Exploring these questions within the context of state elections in India, I find that certain candidate characteristics are, indeed, associated with higher vote shares, even after controlling for the popularity of the party label on which a candidate contests. More specifically, making use of a unique dataset on candidate characteristics from five state elections in the north Indian state of Haryana, I find that having previously held high-level political office and having family who previously held high-level office are associated with winning substantially higher vote shares. Other characteristics, such as being a native of one's election constituency,² having previously held local political office, having family who previously held local political office, and having an

² India elects its state and national legislators in single-member districts. The candidate winning the plurality of vote in each district wins the seat. In India, these electoral districts are called constituencies. Districts refer to an administrative unit that typically includes multiple state election constituencies (also know as assembly constituencies) and imperfect overlap with national election constituencies (known as parliamentary constituencies).

occupation in business are also associated with somewhat higher vote shares. Meanwhile, being an incumbent state legislator, being a member of a numerically large caste, having an occupation in agriculture or the professions, and being a man are not robustly associated with a candidate's vote share.

The findings presented in this article mark an important contribution to the study of Indian politics. In recent years, scholars have amassed considerable data on the characteristics of elected legislators in India.³ This data is useful for understanding who India's legislators are and the extent to which the composition of legislatures in India mirrors the composition of the electorate. However, without information on losing candidates, we do not know whether legislatures simply reflect the pool of candidates competing in elections or if certain types of legislators are systematically more likely to win elections. We therefore know relatively little about the extent to which certain individual attributes can confer electoral advantage on a candidate. Although election analysts might have some notion of what constitutes a "good" candidate in the Indian context, almost no systematic empirical evidence can corroborate or refute such intuitions. This article's major contributions are therefore in documenting not only the kinds of candidates that contest elections but also in identifying which candidate attributes are associated with higher vote shares.

The findings presented here contribute to a growing literature on candidates in Indian politics. Although journalistic accounts of elections frequently explain electoral outcomes in

³ See Subhash Arora and R.K. Prabhakar, "A Study of Municipal Council Elections in India: Socioeconomic Background of Candidates in Rohtak, Haryana," *Asian Survey* (October 1997), pp. 918-926; S.S. Chahar, *Dynamics of Electoral Politics in Haryana* (Delhi: Sanjay Prakashan, 2004); Kanchan Chandra and Wamiq Umaira, "India's Democratic Dynasties," *Seminar* (June 2011); Christophe Jaffrelot, *India's Silent Revolution* (New York, Columbia University Press, 2003); Christophe Jaffrelot and Sanjay Kumar, eds., *Rise of the Plebeians? The Changing Face of Indian Legislative Assemblies* (New Delhi: Routledge, 2009),

terms of factors related to candidates—such as, a party’s poor candidate selection or the presence of rebel candidates⁴—political scientists have long tended to focus on parties as the main determinants of electoral outcomes. More recently, however, scholars are increasingly focusing on candidates. Many, for example, have examined the effects of India’s system of reservation, which mandates that women or members of particular castes must represent certain electoral constituencies. In effect, these studies consider the impact of the gender or ethnic identity of a legislator on a variety of outcomes.⁵ Some very recent research also considers why voters might prefer certain types of candidates, such as those belonging their own ethnic group.⁶ Most closely related to the findings in this article, however, are a handful of studies that examine whether candidates with certain characteristics are more electorally successful than those who do not. So far, researchers have found that incumbent candidates are electorally disadvantaged,⁷ while those who are wealthier and have criminal backgrounds fare better in elections.⁸ In this article, I

⁴ Rebel candidates are party dissidents who contest as independents or on minor party labels when they do not secure their party’s nomination.

⁵ See Rikhil Bhavnani, “Do Electoral Quotas Work after They are Withdrawn? Evidence from a Natural Experiment in India,” *American Political Science Review* (February 2009), pp. 23-35; Simon Chauchard, “Can Descriptive Representation Change Beliefs about a Stigmatized Group? Evidence from Rural Rajasthan,” *American Political Science Review* (forthcoming); Rohini Pande, “Can Mandated Political Representation Provide Disadvantaged Minorities Policy Influence? Theory and Evidence from India,” *American Economic Review* (September 2003), pp. 1132-1151; Thad Dunning and Janhavi Nilenkani, “Ethnic Quotas and Political Mobilization: Caste, Parties, and Distribution in Indian Villages,” *American Political Science Review* (February 2013), pp. 35-56.

⁶ Simon Chauchard, “Unpacking Ethnic Voting: Theory and Micro-level Evidence from North India,” unpublished manuscript, 2014.

⁷ Leigh L. Linden, “Are Incumbents Really Advantaged? The Preference for Non-Incumbents in Indian National Elections,” unpublished manuscript, 2004; Yogesh Uppal, “The Disadvantaged Incumbents: Estimating Incumbency Effects in Indian State Legislatures,” *Public Choice* (January 2009), pp. 9-27.

⁸ Milan Vaishnav, “The Market for Criminality: Money, Muscle and Elections in India,” unpublished manuscript, 2012.

similarly focus on whether candidates possessing a range of potentially relevant attributes do better in elections than those who do not.

The remainder of this article is divided into four sections. Section 1 describes the data that I use to investigate whether certain types of candidates tend to win more votes than others. In so doing, I explain my rationale for focusing on state-level elections, the state of Haryana, and the specific variables on which I collected data. Section 2 describes the candidate pool in Haryana, providing a broad sense for the kinds of candidates fielded in elections. Section 3 presents the results of the statistical analysis that I conduct to determine whether certain types of candidates fare better in elections than others. In Section 4, I discuss some of the implications of these findings.

1. Data

To ascertain where certain kinds of candidates are more electorally successful than others, I collected data on the characteristics of major candidates contesting state-level elections in the north Indian state of Haryana in the five elections from 1991 through 2009.⁹ The bulk of the data was collected between December 2010 and February 2011 by the author and five research assistants, all of whom were graduate students at Indian universities. The data collected are not available from any public source, so we obtained our information through interviews with politicians across Haryana's 21 administrative districts. Respondents were almost all activists from the state's major parties, and a large number had themselves contested states elections at some point in time. Those that were not former candidates either held posts in the local or state party organizations or held local-level political office.

⁹ I exclude bye-elections.

Why State Elections in Haryana?

The decision to gather data on candidates in the state of Haryana, a relatively small state in north India, reflected a mix of theoretical and logistical reasons. No state in India is representative of the country as a whole; however, some states—such as Kerala with its large non-Hindu population and very high literacy rates—are clearly very non-representative. On many dimensions, Haryana is reasonably representative of India. The state is relatively close to the Indian average in terms of the share of its population that is Hindu, Scheduled Caste, resident in urban areas, and literate. Thanks to the growth of Gurgaon and Faridabad, two satellite cities of Delhi located in Haryana, the state has, in the last decade, become substantially wealthier than the national average. However, the state's relative wealth is fairly recent. In addition, Haryana is part of the Hindi belt, the single largest linguistic region in India. In terms of party systems, India's states exhibit three major patterns: those dominated by national parties, those dominated by regional parties, and those that have important regional and national parties. Haryana falls into the last category, which is the most common among major states.¹⁰ The two largest parties are Congress, the country's largest and most national party, and the Indian National Lok Dal (INLD), a regional party.¹¹ Thus, for the purposes of selecting a state whose politics is maximally generalizable across India, Haryana is as good a state as any. Additionally, from a logistical point of view, Haryana is relatively small in its geography, which increased the feasibility of collecting data on candidates from across the state.

¹⁰ Of India's seventeen major states, ten have party systems including a mix of regional and national parties: Andhra Pradesh, Assam, Haryana, Jharkhand, Karnataka, Kerala, Maharashtra, Orissa, Punjab, and Uttar Pradesh.

¹¹ For consistency, I use the name INLD for the entirety of the period under consideration (1991-2009). However, in 1991, the INLD was known as the Samajwadi Janata Party and in 1996 as the Samata Party.

I focus on state elections because state politics is widely regarded as the most salient political arena in India.¹² National elections represent little more than an aggregation of state-level elections, each of which operates according to its own rhythm. For voters, state governments oversee policy domains relevant to every-day life: health, sanitation, education, and law and order. Politicians too place tremendous weight on state elections, often eschewing national-level posts in favor of state-level positions. Moreover, because state election constituencies are much smaller than national ones, gathering data on state elections permitted me to gather far more data points over the same geographic space than would have been possible if focusing on national elections.

Criteria for Inclusion

The number of candidates contesting Indian elections is vast. In the five elections included in the dataset, there were 7,663 candidacies.¹³ Almost two thirds of these candidacies (5,055) won less than 1% of the vote. I therefore target only electorally relevant candidacies for inclusion in the dataset. I aimed to include candidates who either 1) contested on a party label winning 5% or more in the state or 2) won 10% or more in their constituency, irrespective of party label. Aside from candidates meeting these criteria, I also aimed to include candidates from a handful of parties that did not cross the 5% threshold. I include the Bahujan Samaj Party in 1991 and 2005 since it did not cross the 5% threshold in those years but won more than 5% in 1996, 2000, and 2009. I also include the Janata Dal because it nearly crossed the 5% threshold in

¹² Yogendra Yadav and Suhas Palshikar, “Ten Theses on State Politics in India,” *Seminar* (November 2008).

¹³ All data on electoral results come from election reports published by the Election Commission of India. By candidacies, I do not mean discrete individuals. If the same person contested all five Haryana elections from 1991 through 2009, she would count as five separate candidacies.

1991 and was important in national politics in 1996. Based on these criteria, I targeted 2,113 candidacies for inclusion.

In total, the dataset includes 2,032 candidacies, representing 1,219 individuals. Of the 2,032 candidacies, 1,855 met the criteria for inclusion, meaning that we successfully gathered data on about 88% of the targeted candidacies. Most of the missing candidacies are for candidates winning relatively small vote shares but who were targeted based on their party label. The 177 candidacies that were not targeted but that are included in the data come mainly from individuals who contested multiple times but failed to meet the criteria for inclusion in some elections. In a few instances, data on one or two characteristics may be missing. Rather than drop the candidate entirely from the dataset, I code a candidate as not having a particular attribute. Such codings are relatively rare.¹⁴

Table 1 presents a summary of the data by political party. It lists the number of observations for each party, the election years in which candidates for the party are included, and the average vote share won by candidates in the dataset. Since the Haryana Legislative Assembly has 90 seats, there are a maximum of 450 candidacies for each party, though only Congress fielded candidates in every constituency in all five elections. Data are relatively even distributed across years.¹⁵

Selection Effects

Obviously, the data collection procedure created important selection effects since it systematically excluded poorly performing candidates. The average vote share won by

¹⁴ For instance, caste was missing for 30 candidacies or 1.5% of the total. Caste was one of the variables for which there was a comparatively high incidence of missing data.

¹⁵ The number of observations is 388, 414, 366, 382, and 482 for 1991, 1996, 2000, 2005, and 2009, respectively.

candidates included in the dataset is 20.3%, whereas the average vote share for candidates not intended for inclusion was 0.6%. Although the dataset includes a disproportionate share of successful candidates, many major parties fielded highly unsuccessful candidates, thereby ensuring that a sizeable number of unsuccessful candidacies are included in the dataset. Indeed, 22% (455) of the observations are for candidacies winning less than 5% of the vote. Perhaps more importantly, the exclusion of unsuccessful candidacies should bias the data against finding evidence that candidate characteristics correlate with election outcomes. Suppose that successful and unsuccessful candidates systematically differ in their characteristics. The fact that the data mainly captures relatively successful candidates means that the candidates in the dataset are likely to have fairly similar characteristics. Meanwhile, those systematically excluded should be those highly unsuccessful candidates whose characteristics differ substantially from those of their more successful counterparts. As a result, finding evidence of a correlation between candidate attributes and vote shares should be more difficult given the exclusion of many poorly performing candidates.

The Characteristics

The aim of the data collection was to compile objective information about candidates that voters would plausibly use to arrive at their vote choice. To that end, the dataset includes information on caste, place of birth and residence, previous political experience, relatives in politics, and occupation. I focus on these characteristics because, based on previous fieldwork, both in Haryana and elsewhere in India, I determined that they are characteristics that are 1) relatively objective (unlike honesty or accessibility), 2) widely known (unlike education levels or income), and 3) characteristics potentially valued by voters. Below, I detail the information

included in the dataset, the reasons that I expect the characteristics to be associated with greater electoral success, and the resulting variables used in the subsequent data analysis.

Caste.¹⁶ Caste is thought to be an important factor in vote choice. Popular arguments about caste suggest that voters prefer to vote for politicians of the same caste for expressive reasons. Chandra argues that voters have instrumental reasons to vote based on caste because patronage is often doled out along caste lines.¹⁷ If candidates belong to a numerically large caste, then they should have a sizeable popular base on which to build, making them more electorally viable. Since the census does not collect and publish data on caste, I not only collected information on the caste of candidates but also on the caste demography of each constituency (based on respondents' estimates). The resulting variable is *Large caste*, coded as 1 if a candidate is from a caste described by respondents as one of the three or four largest in the constituency or constituting a sizeable share of the population and 0 otherwise.

Place of birth and place of residence. Voters should prefer local candidates to outsiders. Local candidates are likely to enjoy stronger personal ties to a larger number of voters. Local candidates should also be more accessible to constituents in need of assistance. Finally, because local candidates have a longer history in the community, they are more likely to have a history of public service in the area and to therefore be well known as a problem solver. The resulting variable is *AC native*, coded as 1 if the candidates was either born in or is a longtime resident of the assembly constituency (AC) from which she is contesting and 0 otherwise.

¹⁶ Most castes are embedded with religions, so that knowing a candidate's caste necessarily reveals a candidate's religion. Since some castes span multiple religions, data on religion were also collected.

¹⁷ Kanchan Chandra, *Why Ethnic Parties Succeed: Patronage and Ethnic Headcounts in India* (New York: Cambridge University Press, 2004).

Political experience. Candidates with previous political experience are likely to be well known in the constituency from their history of solving problems in office. They are also more likely to be politically well connected. These connections should make experienced politicians better able to provide the constituency with greater resources (i.e., patronage and pork) and assist constituents in getting special favors, circumventing rules, or rectifying problems with the bureaucracy. Increasingly powerful positions, such as ministerial berths, should only increase a politician's access to power, which can benefit her constituents. Part of the data collection process involved identifying instances in which the same person contested multiple elections, so that I could construct variables related to previous political experience. Currently, election data do not provide any way to verify the identity of a candidate over time. Transliterations of names into English are inconsistent; candidates are inconsistent in how they record their name (sometimes using middle names or caste-based last names and other times not); and many candidates do not have last names, with the effect that elections may feature several different candidates with almost identical names. Therefore, an important part of the data collection involved verifying whether candidates with similar names who appeared in multiple elections were, in fact, the same person.

Several variables result from this information on political experience. *Local government* is coded as 1 if a candidate previously held local political office and 0 otherwise. Since high quality data on local election results over time are not readily available from a centralized government source, this variable is based on respondent responses. *Former MP/MLA* is coded as 1 if a candidate was previously a member of the legislative assembly (MLA) or member of parliament (MP) and 0 otherwise. This variable is based on election results and respondent responses that helped identify instances when the same person contested multiple elections.

Former minister is coded as 1 if a candidate was previously a cabinet minister in the state government and 0 otherwise. This variable is based on both respondent responses as well as data collected on the composition of government ministries included in publications of legislator biographies available in the Haryana Vidhan Sabha library. *Former CM* is coded as 1 if a candidate was previously a chief minister (the state-level analogue of the prime minister) and 0 otherwise. Finally, *Incumbent* is coded as 1 if the candidate is a sitting member of the legislature and 0 otherwise. Note that all candidacies coded as 1 for *Incumbent* are necessarily coded as 1 for *Former MP/MLA*, but many coded as 1 for *Former MP/MLA* are coded as 0 for *Incumbent* because they were not a sitting legislator at the time of a particular election.

Relatives in politics. Political dynasties are very common in India.¹⁸ They are important because they confer a high degree of name recognition even on those who are politically inexperienced. Relatives in politics can also help if political families create “brands” associated with service to the community or proximity to centers of power. Just as political experience may indicate connections to those in power, so too do relatives in power. The data on family in politics produces three variables depending on the level of politics at which family members are involved. *Family: Local* takes a value of 1 if the highest post a candidate’s relative held was local politics or if the relative’s only activity in politics was a failed bid for an MLA seat and 0 otherwise. *Family: MP/MLA* takes a value of 1 if a candidate previously had family members whose highest office was at the state- or national-level but is not part of one of the state’s major political dynasties and 0 otherwise. Finally, *Family: Dynasty* is coded as 1 if a candidate belongs to one of the state’s major political dynasties, defined as being related to Bansi Lal, Bhajan Lal,

¹⁸ Chandra and Umaira, “India’s Democratic Dynasties.”

Bhupinder Singh Hooda, Devi Lal, or Rao Birender Singh.¹⁹ I only treat family members in politics as relevant for the construction of these variables if the family member preceded the candidate in politics. Family members who enter politics after the candidate do not count.

Work. Occupational data were collected based on findings in the broader political science literature on candidate quality that suggest that a candidate's occupation can signal competence and therefore influence voting behavior.²⁰ Based on the experience of many consolidated democracies, one might expect that voters would favor candidates with high status occupations in the professions, since these occupations may be signals of competence. In the Indian context, I am agnostic about this prediction. Rather, based on recent work documenting a link between candidate wealth and electoral success,²¹ I expect that candidates with lucrative occupations in business are more likely to do well in elections. Such candidates are likely to have ample campaign resources useful not only for traditional campaign activities, but also for vote-buying. They are also more likely to be sources of local employment, thereby not only capable of distributing their own largesse but also of providing livelihoods to constituents. Given the large number of occupations cited by respondents, I use occupation data to create three variables. *Business* is coded as 1 if a candidate has an occupation related in some way to commerce, whether large or small and 0 otherwise. *Professional* is coded as 1 if the candidate is a lawyer, doctor, engineer, teacher, state civil servant or other white collar job requiring high levels of education and 0 otherwise. Finally, *Agriculture* is coded as 1 if the candidate is involved in some

¹⁹ Each was a chief minister of the state, and together they have governed Haryana for most of its existence. Several former chief ministers are not counted as the founders of political dynasties because they never engendered strong public followings (i.e., Bhagwat Dayal Sharma, Banarsi Das Gupta, and Hukum Singh).

²⁰ See Monica L. McDermott, "Candidate Occupations and Voter Information Shortcuts," *Journal of Politics* (February 2005), 201-219.

²¹ Vaishnav, "The Market for Criminality."

way with agriculture and 0 otherwise. These categories are not mutually exclusive; some candidates are, for instance, coded as 1 for both *Professional* and *Business* if, for example, they are lawyers but also have a family business. Some candidates—those described by respondents as laborers or housewives, for instance—are not coded as having any of the above types of work.

Gender. I also have a variable for gender, which is provided by the Election Commission of India. The variable is coded 1 if a candidate is male and 0 if she is female. I do not have a strong expectation about gender, as gender never arose in my fieldwork as a relevant consideration in state-level elections that (dis)advantages some candidates over others. Furthermore, work on gender in the comparative context suggests that the dearth of women in elected office stems more from women's failure to run for office than from the electorate's preference for male candidates.²² Nevertheless, given persistent gender hierarchies in India, I include gender in the event that male candidates may be more successful than female ones.

2. The Candidate Pool

What does the pool of candidates look like? Description of the pool of candidates is itself a valuable contribution to the study of Indian elections. While much research has documented the characteristics of legislators, almost no research has sought to describe the broader pool of candidates seeking office. Figure 1 graphically presents the frequency of candidate characteristics, indicating the share of candidates in the dataset who possess a particular characteristic. The dark gray bars represent all 2,032 candidates. The light gray bars represent the 963 candidates from major parties, defined in the figure as those winning more than 10% of total

²² See Jennifer L. Lawless and Richard L. Fox, *It Takes a Candidate: Why Women Don't Run for Office* (New York, Cambridge University Press, 2005);

vote in an election.²³ One immediately striking feature of Figure 1 is that with the exception of *Previous MLA/MP*, *Incumbent*, and *Previous Minister*, the differences between the full sample and the pool of candidates fielded by the largest parties are not all that large.

[Figure 1 about here]

Focusing on the full sample, several other important trends stand out. Most candidates are male (93%), from numerically large castes (72%), and natives of the constituency from which they contest (73%). Nearly 92% are from the same administrative district in which their assembly constituency is located (not shown in Figure 1), suggesting that about 20% are not natives of their constituency but are natives of nearby places. Substantially more candidates are engaged in business (48%) than in agriculture (22%) or the professions (24%). All of these trends, if not the exact percentages, are also true among major-party candidates.

Some form of previous political experience is common as nearly half of all candidates in the sample had some prior political experience. About 33% of the pool consisted of former MLAs or MPs, while about 23% had previously held positions in local government. There is, however, little overlap between those two groups, as only about 8% had experience in both local politics and as an MLA or MP. Not surprisingly, the share of candidates who were sitting incumbents (15%), former ministers in the state government (12%) or former chief ministers (1%) is considerably smaller than the overall number of candidates who were, at some point, previously MPs or MLA. Finally, with regard to family in politics, around a quarter of candidates had family precede them in politics. About 8% had family who had been engaged in local politics, but no higher; another 13% had had family in state or national politics; and 2% came from one of the state's major political dynasties. For all variables related to experience and

²³ The figure therefore includes Congress and INLD candidates from all elections as well as Haryana Vikas Party (HVP) candidates in 1991 and 1996, but not 2000.

family in politics, with the exception of *Local Government*, the shares for candidates from major parties are higher than for the full sample, though not always dramatically so.

It is worth remembering that the percentages represented by the dark gray bars in Figure 1 refer only to candidates in the dataset, not to the entire candidate pool from 1991 through 2009. The gender, caste, or occupational composition of the entire candidate pool may be little different from that of the candidates in the dataset since individuals with these characteristics are abundant in the population as a whole and may be similarly well represented among electorally irrelevant candidates. However, given the limited number of individuals with previous experience in politics or family in politics, the share of candidates with these latter attributes would almost certainly decline if it were possible to gather data all candidates, including ones winning miniscule vote shares. Thus, these data cannot tell us definitively what the entire candidate pool in Haryana looked like during this period. However, the data do provide a very clear sense as to the frequency of various characteristics among those candidates who were either 1) successful in winning the nomination from a medium-sized or large political party or 2) modestly successful in at least one election, irrespective of party label.

3. Who Wins Votes?

Having described the candidate pool, I now turn to the question of whether candidates with certain characteristics tend to win larger voter shares. To estimate the relationship between a candidate's characteristics and her vote share, I estimate a series of regressions using ordinary least squares (OLS), with standard errors clustered by the electoral constituency. In these models,

the dependent variable is *Candidate Vote*, or a candidate's vote share in an election.²⁴ My main independent variables of interest are the characteristics described in Section 1 and included in Figure 1. In addition, I need to include some measure capturing the expected level of popular support that a candidate can expect based on her party label. Because a candidate may win many of her votes thanks to her party label rather than her own attributes, it is important to take account, in some way, of her party label. If certain types of candidates are systematically recruited into certain political parties, then the observed relationships between attributes and vote share may be driven more by the candidate's party label and voters' preferences for a specific party than by the candidate's characteristics. To account for the portion of a candidate's vote share attributable to her party label, I construct three different variables related to party.

The first is *State Party Vote*, which is the average vote share won by all of the candidates from a candidate's party in a given election.²⁵ This variable differs from the party's overall vote share in the election. If, say, a party fields candidates in only four of 90 constituencies, then the party's overall vote share is necessarily very small. But, if the average vote share won by those four candidates is 40%, then the value entered for *Single Party Vote* is 40% for any candidate from that party in that election. This variable captures broad differences in overall levels of

²⁴ In analysis not included here, I replicate all of the analyses described below with *Candidate Winner* as the dependent variable. This variable takes a 1 if a candidate won her seat and 0 if she lost. The results are broadly similar, and almost all coefficients in this analysis have the same sign as those in the analysis presented here. I do not focus on the analysis of *Candidate Winner* because the dependent variable itself is quite problematic. Given that some constituencies feature only two credible candidates while others feature four or five, the correlation between a candidate's vote share and winning her seat is only 0.70. There are many losing candidates who won greater vote shares than many winning candidates. For instance, 186 losing candidacies (12% of the total in the dataset) won more than 30% of the vote, while 61 winning candidacies in the dataset (14%) won less than 30%. Identifying characteristics associated with winning is not particularly meaningful when candidates can win with vote shares ranging anywhere from 16% to 80%.

²⁵ For all party related variables, I treat independent candidates as part of a common party label.

support for candidates from different parties. However, *State Party Vote* does not take account of the fact that parties may have relatively heterogeneous support bases even within a state. The next two variables take account of variation in partisan support within a state.

Region Party Vote is the vote share won by the other members of the candidate's party within the region of the state in which her constituency is located. I group electoral districts into seven regions. The seven regions are based loosely on the seven administrative divisions into which Haryana was divided in 1966, when it became a state.²⁶ When calculating the average vote share of candidates from the party, I exclude the candidate in question. Since the calculations for *Region Party Vote* rely on a small number of candidates, a single candidate with a very large or small vote share could significantly alter the average. And since I am interested in predicting a candidate's vote share, that very same vote share should not heavily influence one of the predictors. Thus, if a region has nine constituencies, and Candidate X's party fields candidates in all nine constituencies, then the value of *Region Party Vote* for Candidate X is the average vote share of the other eight candidates in the region. The quantities in *Region Party Vote* therefore vary for each candidate, even within the same region.²⁷

As an alternative to *Region Party Vote*, I also calculate *District Party Vote*, which is calculated in an identical manner as *Region Party Vote* except that it averages across an administrative district rather than a region. Whereas I use seven regions for *Region Party Vote*, I

²⁶ The seven regions are grouped as follows. The first region includes constituencies in the administrative districts of Ambala, Panchkula, and Yamunanagar; the second includes Faridabad, Gurgaon, Mewat, and Palwal districts; the third includes Jhajjar, Rohtak, and Sonapat districts; the fourth includes Jind and Kaithal districts; the fifth includes Karnal, Kurukshetra, and Panipat districts; the sixth included Fatehabad, Hisar, and Sirsa districts; and the seventh includes Bhiwani, Mahendragarh, and Rewari districts.

²⁷ For *State Party Vote* I do not bother excluding a candidate's own vote share because the averages are, in most cases, based on a sufficiently large number of observations that the inclusion or exclusion of any one candidate would not substantively alter the averages.

use Haryana's 21 administrative districts for *District Party Vote*. One potential advantage of this measure is that it can capture fairly fine-grained variations in levels of party support within the state. One disadvantage is that it relies on the average across a very small number of observations. Since many districts include only three or four constituencies and the variable is calculated so as to exclude the candidate whose vote share is being predicted, the variable often relies on an average across only two or three candidates. If one of these candidates performs uncharacteristically well or poorly, then the value for this variable may be unusually large or small, capturing idiosyncratic factors more than a party's baseline level of support in an area.

I begin by first estimating a series of regressions, estimating the relationship between each of the individual attributes described in Section 1 and a candidate's vote share. Each of these models includes only two independent variables, an attribute and *Region Party Vote*. The coefficients resulting from these regressions are plotted in Figure 2 as gray dots. The placement of the dot along the horizontal axis indicates the size of the estimated coefficient. None of these models take account of the fact that a candidate may possess many of these attributes. Therefore, I next estimate a model in which I include all of the individual attributes as well as *Region Party Vote*. The coefficients from this model are plotted in Figure 2 as black dots. The bars represent the 95% confidence intervals. For the most part, the size of the regression coefficients in the model with all attributes are in the same direction as those in the models with only a single attribute. The coefficients in the full model are often somewhat smaller, though many are of roughly similar magnitude. With a few exceptions, the coefficients in Figure 2 suggest that most of the characteristics on which I have data are, as expected, associated with greater candidate vote shares.

[Figure 2 about here]

With respect to caste, in the full model, the estimated coefficient is positive; however, it is relatively small and does not reach conventional levels of statistical significance. The coefficient for natives of their assembly constituency (*AC Native*) is, in the full model, of a relatively modest magnitude (0.01), but statistically significant. For the variables dealing with previous political experience, all reach conventional levels of statistical significance except *Incumbent*. The magnitudes on *Local Government* and *Previous Minister* are relatively modest in size (0.02), especially as compared to *Previous MLA/MP* (0.06) and *Previous CM* (0.16), which are the two largest coefficients. Interestingly, when controlling for no other attributes, the coefficient on *Incumbency* is sizable, positive, and statistically significant. However, once other variables are included, the estimated coefficient is very small, negative, and statistically insignificant. Taken together with the coefficient on *Previous MLA/MP*, the findings for *Incumbency* suggest that, net of the advantage associated with previously holding high-level legislative office at some point, being the sitting legislator offers no electoral advantage. In a context of relative economic scarcity, the fact that incumbency is not associated with electoral advantage may reflect voters' frustrations with those who are in power and perceived as having failed to improve economic conditions in their constituencies.

The coefficients on all of the variables related to family in politics are of a similar size, 0.05 for *Family: Dynasty* and 0.03 for *Family: Local* as well as *Family: MLA/MP*, though the first is only significant at the 10% level. *Business* is the only occupational variable that is associated with a candidate's vote share. *Agriculture* and *Professional* are not statistically significant. Interestingly, the latter two are both negative. Finally, the coefficient on *Male* is very small and statistically insignificant.

Most of the results presented in Figure 2 are robust to alternative specifications of the party variable. I replicate the full model with all candidate attributes in Table 2 using *State Party Vote* (Model 1) and *District Party Vote* (Model 3) in lieu of *Region Party Vote* (Model 2).²⁸ For *Local Government*, *Previous MLA/MP*, *Previous CM*, *Family: Local*, and *Family: MLA/MP*, the coefficients are statistically significant in all models and of roughly similar magnitude. Similarly, for *Incumbent*, *Agriculture*, *Professional*, and *Male*, the results are also similar in that none of the coefficients are statistically significant. For the occupational categories, the sizes of the coefficients are very stable. However, the remaining variables do not reach conventional levels of statistical significance in all models. Nevertheless, with the exception of *Previous Minister*, the coefficients and standard errors across all specifications do not fluctuate wildly, even if in some of the models the variables just miss conventional levels of statistical significance. I take the relative stability of the results across models and the statistically significant relationships in two of the three models as tentative evidence of a relationship between these variables—*AC Native*, *Family: Dynasty*, and *Business*—and *Candidate Vote*. Naturally, of course, these findings should be treated with caution. For *Large Caste*, the relatively small coefficient is statistically significant only in Model 3, though the p-value never exceeds 0.135 in any of the models in Table 2, indicating that it may be wrong to firmly conclude that no relationship exists between *Large Caste* and *Candidate Vote*. Finally, I set little stock in the estimate of *Previous Minister* in Figure 1, given that the coefficient is very different than in the other two specifications in Table 2.

[Table 2 about here]

²⁸ The results in Table 2 are virtually unchanged if I include dummy variables indicating whether the candidate comes from the incumbent state or national government at the time of the election.

In sum, the results from this analysis suggest a robust association between several candidate characteristics and a candidate's vote share. To illustrate the substantive significance of these findings, the coefficients represented by the black dots in Figure 2 (and in Model 2 of Table 2) imply that a candidate whose fellow party members win an average of 17.9% of the vote in the other electoral districts in his region of Haryana (the mean for the variable *Region Party Vote*) but who possesses none of the characteristics included in the model can expect to win 14.8% of the vote in her race. However, if this same candidate is a former MLA (but not a sitting incumbent), previously held a position in local politics, and has a relative who was earlier an MLA, she can expect to win 26.5%, which represents an increase of nearly 12%. To put this increase into further context, whereas no one in the dataset won their seat with less than 14.8% of the vote, 32 candidates won their seats with less than 26.5%.

4. Discussion

The previous section showed that candidates who were previously an MLA or MP, were previously a chief minister or had family members precede them in politics who were MLAs or MPs tend to win significantly larger vote shares than other candidates. Additionally, candidates who are natives of their constituency, previously held local political office, had family who were involved in local politics, are members of the state's major political dynasties, and have occupations in business also tend to win larger vote shares. Gender, incumbency, agriculture or professional occupations, and, for the most part, caste, are not associated with greater vote shares. Indeed, the results suggest that, if anything, incumbent MLAs and candidates with occupations in agriculture and the professions may do worse in elections. In this final section, I discuss some of the implications of these findings. In particular, I touch on the surprising

unimportance of caste in my analysis, the generalizability of the findings across India, and what these results can and cannot say about the importance of candidates in shaping election outcomes.

First, one aspect of these findings that may seem surprising is the apparent unimportance of caste. In two of the models in Table 2, the coefficient on *Large Caste* is statistically insignificant, and in all models the size of the coefficient is fairly small. Given the primacy placed on caste in many studies of electoral politics in India, the findings from these data may, at first, come as a surprise. However, upon further reflection, a substantively large association between membership in a large caste and a candidate's vote share might actually be the more surprising result. For one, my measurement of caste is extremely blunt. Ideally, I would want to know the share of a constituency's population that belongs to a candidate's caste to more precisely measure the extent to which a candidate's membership in a numerically large caste is associated with a greater vote share. In the absence of a caste census, simply dichotomizing candidates into members of large castes and small castes—assessments made on the basis of respondents' estimates of the caste population—is an admittedly crude measure. A more refined measure might reveal a stronger, more substantively significant association between a candidate's caste identity and her vote share.

However, another possibility is that observing a clear association between caste and vote share is difficult precisely because of caste's importance. The importance of caste may, in many cases, prevent parties from fielding candidates from numerically small castes that they might otherwise nominate. Indeed, in many districts, parties field candidates from the same numerically large caste. If, for example, all parties field a candidate from the Jat caste in a constituency where Jats constitute nearly half of all voters—as they do in some parts of Haryana—then caste is no

longer as salient a candidate characteristic. In such a constituency where all candidates are Jats, being a member of this large caste should not constitute much of an electoral advantage. If all candidates are from the same caste, then naturally some candidates from this large caste will fare very poorly. In short, it is possible that the salience of caste produces situations in which its importance is difficult to observe. Furthermore, in reserved constituencies, parties are constrained in whom they can nominate. Although one or two Scheduled Caste *jatis* are, in many constituencies, among the larger castes, reservation ensures that members of some of the constituency's largest castes cannot compete. The fact that I find little association between a candidate's membership in a large caste and her vote share should not, therefore, be understood as strong evidence that caste does not matter.

Second, to what extent are these findings generalizable across India? As noted in Section 1, no state in India is representative of the country as a whole. However, Haryana, like some other states, is not especially unrepresentative. While I obviously cannot rule out the possibility that the findings presented here are unique to Haryana, I see no obvious reason to assume that this is the case. In particular, it is not clear why candidate vote shares would correlate with one set of candidate characteristics in Haryana but not in other states—that is, why, for example, former MLA and MPs would do better in elections in Haryana but not in another state. Perhaps the greater threat to generalizability lies in whether any relationship at all exists between candidate characteristics and vote shares. One possibility is that, in some Indian states, voters are so overwhelmed by considerations related to party that candidate characteristics in no way correlate with vote shares. If parties in Haryana were especially young or fluid or devoid of ties to large ethnic groups, then such a concern would be particularly relevant, as one could argue that the state's parties are especially weak. However, Congress and the BJP (or its predecessor,

the Bharatiya Jana Sangh) have been important political players since Haryana became a state. Although the INLD has repeatedly changed names, it has existed in some form since the 1980s, consistently under the leadership of Devi Lal or his progeny. And, while BJP has long been associated with Punjabi migrants from Pakistan, the INLD has long been seen as a party closely aligned with the Jat caste. As such, Haryana does not appear to be state in which voters are likely to be especially blind to a candidate's party label.

Third, what can this study say and what can it not say about the relationship between candidate characteristics and election outcomes? Importantly, this study speaks only to certain kinds of attributes, those that are objective and widely known. The findings presented here do not address less objective candidate characteristics, such as honesty, responsiveness, or accessibility. Such characteristics may matter, but they can only be measured by tapping voter's opinions on a large scale and are also vulnerable to being shaped largely by voters' partisan preferences.²⁹ In addition, I focus only on those attributes that are widely known and therefore more likely to shape voters' decisions. Other research has addressed the association between less widely known characteristics, such as financial assets and pending criminal cases, that candidates have been forced to reveal to the Election Commission of India since the early 2000s.³⁰ Thus, even though this study is not exhaustive in that it does not address every candidate characteristic that could plausibly influence a candidate's performance, it attends to those characteristics on which all people can agree and that are most likely to influence the decisions of voters because of their relatively public nature.

²⁹ In other words, Congress supporters may be more likely to characterize a Congress candidate as honest and accessible simply because of her partisan label.

³⁰ Vaishnav, "The Market for Criminality." These data are not available for three out of the five elections I consider.

Finally, the ambitions of this study are intentionally descriptive; I aim to describe whether certain types of candidates do better in elections than others. I show that certain kinds of candidates fare better in elections than others and identify those characteristics associated with the greatest electoral advantage. Thus, this study marks an important step forward in trying to understand electoral outcomes in India and the role that candidates play in shaping those outcomes. In doing so, this article opens up important new avenues for future research. Having identified those characteristics associated with greater candidate vote shares, future studies can focus more squarely on causal identification. The associations presented in this study may potentially constitute causal relationships. For instance, it is more than plausible that previously holding the office of MLA or MP causes candidates to win more votes than they otherwise would. However, establishing a causal relationship in the context of electoral outcomes is challenging. Former MLA and MP candidates may systematically possess some other characteristic that is, in fact, the cause of their improved electoral success. Nevertheless, by establishing what characteristics seem to matter and which do not, this study lays the foundation for future work in this area. It also lays the groundwork for future studies focused on mechanisms—why certain characteristics should lead to greater electoral success. In discussing why I focus on certain characteristics in this study, I outlined reasons why we should expect these characteristics to be associated with greater candidate vote shares. But, the question of mechanisms remains an open one ripe for future research.

Table 1. Distribution of Observations by Party

| Party (Abbreviation) | Years in Dataset | Observations | Average Vote |
|--|-------------------------|---------------------|---------------------|
| Bahujan Samaj Party (BSP) | All | 241 | 7.3% |
| Bharatiya Janata Party (BJP) | All | 299 | 13.2% |
| Haryana Janhit Congress (Bhajan Lal) (HJC) | 2009 | 75 | 8.8% |
| Haryana Vikas Party (HVP) | 1991-2000 | 177 | 20.6% |
| Indian National Congress (INC) | All | 444 | 32.5% |
| Indian National Lok Dal (INLD) | All | 403 | 27.7% |
| Janata Dal (JD) | 1991, 1996 | 40 | 11.6% |
| Other parties | All | 85 | 9.6% |
| Independents | All | 268 | 16.4% |
| <i>Total</i> | <i>All</i> | <i>2,032</i> | <i>20.3%</i> |

Figure 1. Characteristics of the Candidate Pool, Haryana 1991-2010

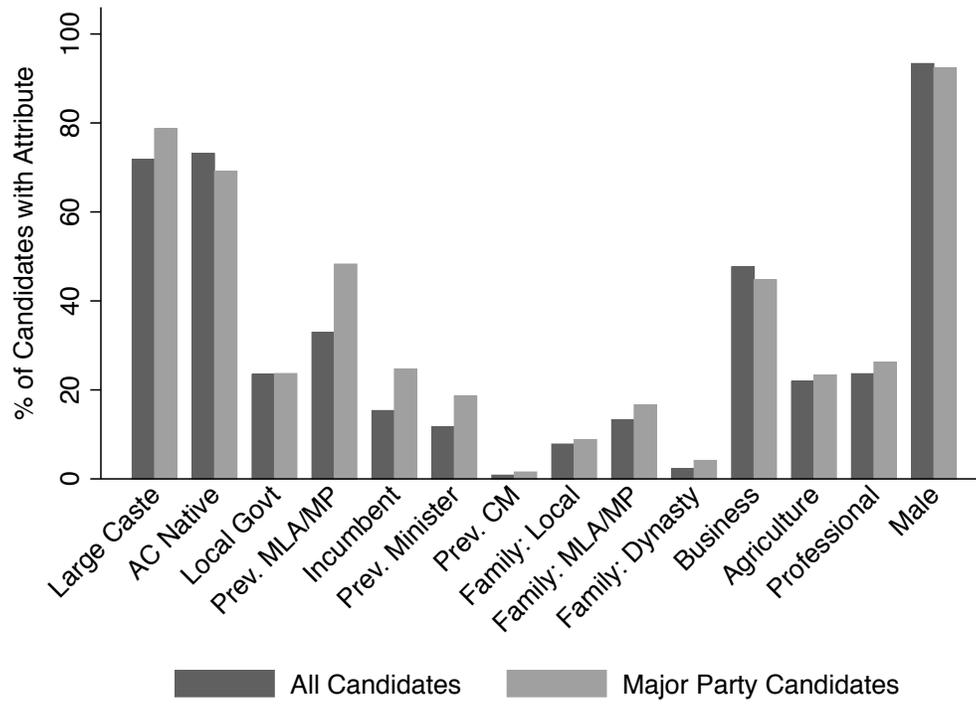


Figure 2. Coefficients from OLS Regressions Predicting Vote Choice

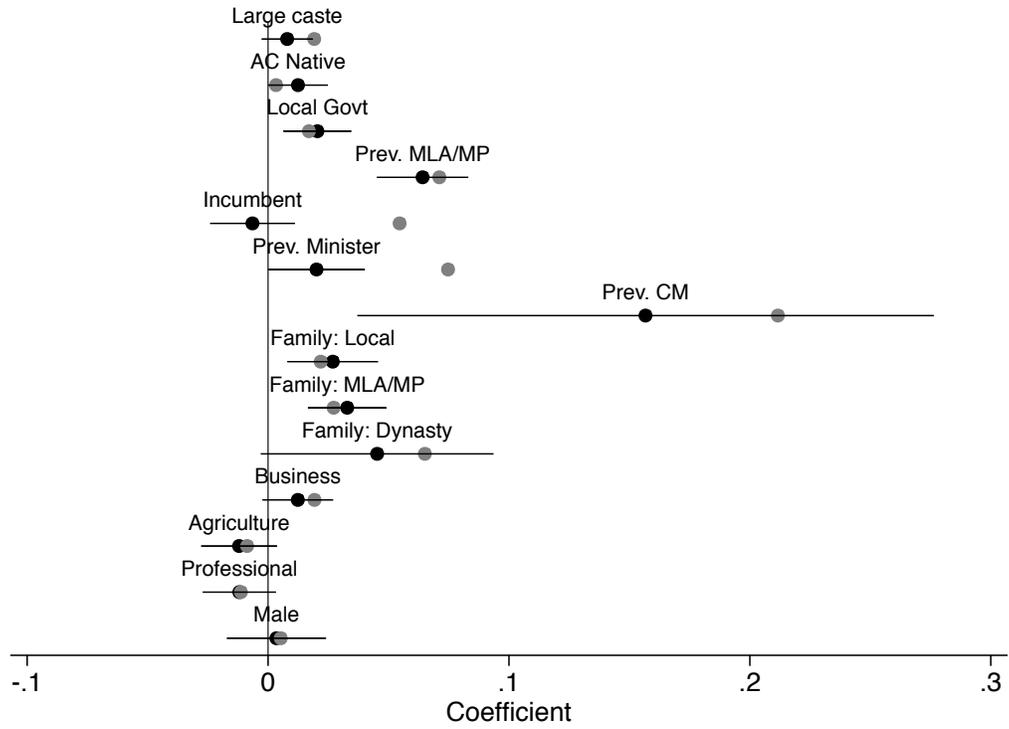


Table 2. Regressions with Various Controls for Party

| | Dependent variable: Candidate Vote | | |
|---------------------|---------------------------------------|-------------------|-------------------|
| | (1) | (2) | (3) |
| State Party Vote | .691*** (.029) | | |
| Region Party Vote | | .673** (.029) | |
| District Party Vote | | | .612*** (.028) |
| Large caste | .008 (.005) | .008 (.005) | .010* (.006) |
| AC Native | .016** (.006) | .012** (.006) | .008 (.006) |
| Local Government | .024** (.007) | .020*** (.007) | .020*** (.008) |
| Previous MLA/MP | .064*** (.009) | .064*** (.009) | .065** (.009) |
| Incumbent | -.002 (.009) | -.006 (.009) | -.008 (.009) |
| Previous Minister | .012 (.010) | .020** (.010) | .012 (.010) |
| Previous CM | .191*** (.059) | .157** (.060) | .149** (.061) |
| Family: Local | .029*** (.009) | .027*** (.009) | .022** (.010) |
| Family: MLA/MP | .034*** (.008) | .033*** (.008) | .033*** (.008) |
| Family: Dynasty | .051** (.022) | .045* (.024) | .031 (.022) |
| Business | .010 (.007) | .012* (.007) | .015* (.008) |
| Agriculture | -.012 (.008) | -.012 (.008) | -.006 (.008) |
| Professional | -.010 (.008) | -.012 (.008) | -.008 (.007) |
| Male | .007 (.010) | .003 (.010) | -.000 (.010) |
| Constant | .020 (.014) | .028* (.013) | .044*** (.013) |
| N | 2032 | 2032 | 2032 |
| R-squared | .5066 | .5197 | .5063 |