Coethnic Voters and Candidate Choice by Political Parties [*]

Tushar Bharati [†]

August, 2016

Abstract

This paper contributes to the large and growing literature on how voter's ethnic political preferences, defined as a greater preference for the co-ethnic candidates, affect the quality of elected representatives. Unlike previous studies that take the candidate's ethnicity as given, this paper introduces an asymmetric information candidate choice model in which political parties choose candidates of different ethnicity strategically to maximise their vote profits given the ethnic political preference of the voters. Using the observations from parliamentary and assembly elections from two states in India, this paper finds evidence that political parties might indeed be internalizing the ethnic political preferences of the voters when choosing their candidates, especially in low information settings. Elected representatives from constituencies where candidates all ethnicities can run for election (general constituencies), and therefore, diversification along ethnic lines is possible and, perhaps, beneficial, are more likely to have a prior criminal record than the ones from constituencies where only minority ethnicity candidates can contest (reserved constituencies). However, conditional on the criminal involvement of the pool of candidates to choose from, the

differences between general and reserved constituencies disappear, suggesting that election of criminal candidates is in part due to political parties disregarding potential candidates' involvement in crime in order to be able to diversify along ethnic lines.

PRELIMINARY DRAFT: PLEASE DO NOT CITE OR CIRCULATE.

1   Introduction

Ethnic voting and its inefficiencies is an old but resurgent issue in political economy [Key (1949), Dahl (2005), Young (1979), Bates (1983), Horowitz (1985)]. Contrary to Dahl's conjecture, it does not seem to be a transitional phenomenon [McClain and Garcia (1993), Kaufmann (2003), Barreto (2007), Eifert et al. (2010)]. Much has been written about the poor policy outcomes in situations where ethnic voting has, allegedly, been salient. Key (1949) talks about how ethnic preference limited the competition and led to rampant corruption and slowdown of economic development in the southern region of United States (Besley et al. (2005)). It could be a more serious problem in developing countries, where the accountability mechanism is often underdeveloped and discrimination on the basis of ethnicity often more rampant (Vicente and Wantchekon (2009)). Recent studies by Banerjee and Pande (2009), Vaishnav (2010a) and Acharya et al. (2014) have argued that ethnic voting explain a good amount of inefficiency and criminalization plaguing the politics in India even though the electoral competition in India is intense.[1]

Most of these studies look at the impact of ethnic voting, taking as exogenous, the ethnicity of the candidates running for elections in any constituency. While the rise of certain political parties may itself be in response to the ethnic fragmentation of the state (Jaffrelot (2003)), a party's candidate choice for a constituency, even in the short run with a given number of parties contesting against it in the constituency, may be heavily influenced by the ethnic composition of the constituency and ethnic preferences of the constituency's population. By neglecting it, previous work on the topic might have underestimated the inefficiency resulting from preference for co-ethnic political leaders.

---

[1]All of these studies focus on the elections in the state of Uttar Pradesh in India. Uttar Pradesh is the biggest state of India by population and by number of parliamentary and assembly constituencies.

To understand it better, consider the simple situation described in table 1. Two parties, X and Y, are contesting elections from constituency W. Party X has two potential candidates with different ethnicities and different levels of involvement in criminal activies to choose from to run for elections. Candidate Xa is from ethnicity A and has no involvement in criminal activities. Candidate Xb is from ethnicity B and has a fair degree of involvement in criminal activities. The potential candidates from party Y are Yc and Yd, from ethnicity C and D, with low and very high degrees of involvement in criminal activities, respectively. Let us assume that a candidate with high level of involvement in criminal activities makes for a worse political representative. Also, let us assume that the population of constituency W votes ethnically and the ethnicity D is the dominant ethnicity in the constituency, followed by ethnicity B; A is the smallest ethnicity and C the second smallest. If the political parties are interested in maximizing their votes shares but do not internalize the ethnic preferences of the voters and assume that voters prefer high quality candidates, party X would run candidate Xa and party Y would run Yc for office. However, since, by assumption, voters place a higher weight on ethnicity of the candidate than on other characteristics, Yc, the worse of the two candidates, will win the election. This is the inefficiency talked about in the literature up till now. However, if parties internalize the ethnic preference of the voters, party X will run candidate Xb and party Y will run candidate Yd. This will result in a much more perverse outcome than in the first scenario, the candidate with the highest involvement in crime will be elected.

Table 1: Co-ethnic preference and strategic choice of candidates.

Party X	Party Y
Candidate: Xa	Candidate: Yc
Ethnicity: A	Ethnicity:C
Criminal Cases: Lowest	Criminal Cases: Second Lowest
Candidate: Xb	Candidate: Yd
Ethnicity: B	Ethnicity:D
Criminal Cases: Second Highest	Criminal Cases: Highest

In this paper, I model this second possibility of political parties internalizing the co-ethnic preference of voters. I provide empirical evidence to point out that the inefficiency

it causes might be a realistic concern. The literature on how a party chooses its candidate to run for an office is scarce. Field and Siavelis (2011) look at the selection process of candidate by a party in which each member of the party maximizes her own profit and bargains for the position.[2] However, in most situations, all members of the party working towards maximizing the profit of the party appears to be a better assumption.[3] The paper presents a model of selection of candidates by political parties when they internalize the ethnic preferences of the voters and maximize vote counts. The model builds on Hotelling's linear city model (Hotelling (1929)) that has proved to be the workhorse of the literature attempting to explain the ideological and policy stands taken by parties or candidates contesting elections as a way of differentiating themselves to woo voters and follows closely the model of product differentiation from Seim (2006). The implications are tested against assembly and the parliamentary elections data from the states of Uttar Pradesh and Bihar in India. The results suggest that political parties internalize the ethnic preference of voters and choose candidates accordingly for the parliamentary elections. The results are weaker for the assembly elections. Overall, the results are consistent with earlier research on the role of ethnicity as a signal of policy in low information settings. I present suggestive evidence that such strategic behaviour breaks done in settings where the voters are better informed. A policy implication is to ensure that voters have detailed information about the candidates, parties and their political stances.

This paper makes both a theoretical and empirical contribution to the literature on ethnic political preferences. It presents a model where the candidate's ethnicity is endogenously determined and tests how the implications from the model fare at explaining the data. This paper is closely related to recent works by Banerjee and Pande (2009) Vaishnav (2010a), Acharya et al. (2014), Banerjee et al. (2010) and Chauchard (2016) in that it builds on the observation that inefficiencies might arise if voters have a preference for co-ethnic candidates. The present paper contributes to this literature by pointing out that the inefficiencies might be underestimated if we do not take into account that political par-

---

[2]A related strand literature looks at how political parties choose their policy position. The literature builds on the classic works of Hotelling (1929) and Downs (1957). Stokes (1963) provides a good review of the early work on spatial competition models. The candidate choice model presented here belongs to this group with the major difference being the use of ethnicity as the platform of differentiation.

[3]An alternative middle path to take is to assume that individual party member's profit are proportional to party's profit

ties might internalize voters' ethnic preferences and act strategically. It also shows that the selection of a worse candidate in a non-reserved constituency might be the result of the voters in these constituencies facing a worse pool of candidates to begin with.

The paper, perhaps, is closest to Vaishnav (2010b) that argues ethnic identity politics can lead to voters electing criminal candidates. Vaishnav (2010b) argues that their criminal backgrounds make these candidates 'credible' defenders of the ethnic interests in the voters' perception. This creates incentives for political parties to field these candidates. While the points made in the present paper are similar, there are few important ways in which it differs from Vaishnav (2010b). Vaishnav (2010b) does not explicitly model political parties' incentive to run candidates of the largest ethnic population in the constituency alongside competition among political parties. While he does find significant association between whether or not candidates from all ethnicity can contest for elections and the probability that the candidate is indicted, his model does not empirically verify if there was, in fact, higher likelihood to run candidates belonging to the largest ethnic population. Vaishnav (2010b) cannot explain why different parties may choose to act differently in terms of their candidate choice even if their ethnic affliations, if any, were the same and why parties with such ethnic affiliations might choose to run a candidate from a different ethnicity in a constituency with a majority population of the ethnicity they are affiliated to. The model presented in this paper resolves both these questions. In addition, this paper attempts to distinguish between two intertwined ways in which a criminal candidate might come to office - whether voters prefer criminal candidates or whether they are faced with a worse pool of candidates to choose from because parties compromise on quality for a candidate of the 'right' ethnicity. This paper also provides suggestive evidence on how information might play a role in ethnic

The paper is organized as follows: Section II describes the model and its assumptions. Section III characterizes the equilibrium and discusses the empirical analogue for the model. Section IV describes the data and the setting in which the model is tested. Section V presents the results. Section VI looks at back assumptions made in the model and explains why they might be realistic in the present context. Section VII concludes.

2 The Model

Consider a representative democracy which consists of C constituencies, indexed by c. Polling occurs at the constituency level. There are P political parties, indexed by r, all or some of whom may be contesting the election from any particular constituency. The way the parties contest from a constituency is by choosing a candidate from within their pool of potential candidates for that constituency and running her for the office of member of parliament of that constituency.[4] Individuals - citizen voters or political party member - belong to one of the 'ethnicities', general or reserved.[5] In each constituency, each party can choose to run a candidate of either general or a reserved ethnicity. We assume that parties care about maximizing their votes shares and nothing else. In any constituency, each political party chooses to run the candidate that maximizes the votes the political party wins in that constituency. In addition, we assume, *ceteris paribus*, voters in each constituency are more likely to vote for candidates from their own ethnicity than for a candidate from the other ethnicity.

The situation is similar to firms choosing a location within a market to open their stores at, a topic which has been the subject of numerous paper since Hotelling (1929).[6] As in a firm location model, there are two main determinants of which ethnicity candidate the party decides to run in the election. If the parties internalize the co-ethnic preference of the voters, they will want to run a candidate whose ethnicity is same as that of the majority of voters. This is similar to saying that a firm will want to choose a location where the demand for its product is high. However, all parties will want to capitalize on the high voter population of a particular ethnicity. If all parties choose candidates of the same ethnicity, there will be increased competition for the votes from the that ethnic group. The vote bank

---

[4]India, which is the setting for the empirical part of this paper, follows a dual polity. It has a union government at the center and state governments at the periphery. There is a bicameral legislature at the center, consisting of an upper house (Rajya Sabha) and a lower house (Lok Sabha). A few states have a bicameral legislature, with an upper house (Legislative Council) and a lower house (Legislative Assembly) but the majority have only one house (Legislative Assembly). A more detailed description is provided in section IV.

[5]While these categories are closely related to the caste classification, the paper uses 'ethnicity' to maintain consistency with previous literature. We use the term 'reserved' to refer to groups that are officially designated as Scheduled Castes or Scheduled Tribes by the Constitution of India.

[6]The model presented here is similar to the one Seim (2006) presents. Attempts have been made to keep the notation similar to Seim (2006) to facilitate comparison.

from this big ethnic group will, then, be split among the parties. It might be beneficial for a party, under these circumstances, to deviate and run a candidate of a minority ethnicity. This way, even though it might lose its share of votes from the majority ethnic group, it will have all the votes from the minority ethnicity. The gains might outweigh the losses in some scenarios.

Let the decision vector for each party be given by $\mathbf{d_{re}}$ where $d_{re}{}^c$ takes value 1 if the party chooses to run a general candidate for constituency c, and takes a value 0 if the candidate is from the reserved ethnicity. The number of votes that a potential candidate can fetch for its party depends, therefore, on the population of the voter's of his ethnicity, the number of candidates of the same ethnicity as his own competing against him, the number of candidates of other ethnicities competing against him. It also depends on party level variables and constituency level variables. For example, some parties may be more popular than others in the entire state, while others may be more popular in some constituency, and the candidates from these parties will garner more votes. The number of votes won also depend on a host of other candidate specific variables. However, it is, usually, not possible for a party, say party A, to collect this information about all the potential candidates that another party B is considering to run from a particular constituency. More often than not, party A does not even know all the potential candidates being considered by party B, let alone having information on all the characteristics about these potential candidates that matter for the selection procedure. We assume that parties do not have information on characteristics, other than ethnicity, that rival parties look at while deciding the candidate. We come back to this and the other assumptions in Section VI. We use a constituency-party-candidate's ethnicity specific error term, that we call candidate's profitability type, to include all these omitted variables in the vote profit equation that we will specify below.

Assumption 1 (Independent Symmetric Candidate Profitability Type): The profitability type of a candidate of ethnicity e, from party p, in constituency c, ε_{pe}^c is private information to the party and ε's are independently and identically distributed draws from the distribution $F(.)$. This distribution is common knowledge.

The vote profit function for party p, if it runs a candidate of ethnicity e from constituency c, can be written as

$$\Pi_{pe}^c = X_e^c \beta + \xi^c + \mu_p + \tau_p^c - g(\mathbf{n}^c, \Gamma_{.e}^c) + \varepsilon_{pe}^c \qquad (1)$$

X_e^c is the population share of voters of ethnicity e in constituency c. ξ^c are constituency level variables that might affect the vote profit of running a candidate. An example can be degree of co-ethnic preferences of the voters from that constituency. μ_p are party specific variables, like the nationwide popularity of the political party, that might affect the votes a candidate from that political party wins in a particular constituency. τ_p^c are variables that are different across parties within a constituency and different across constituency for the same party. For example, the popularity of the same party could be different across constituency and different parties will have different popularity levels within any constituency. g(.) represents the competition dimension of the game. $\Gamma_{.e}^c$ captures the magnitude of competition that a candidate of ethnicity e in constituency faces from other candidates. $\mathbf{n}^c$ is the vector of the number of other candidates of each ethnicity contesting against a particular candidate.

Assumption 2 (Linear Additive Competition Effect):

$$g(\mathbf{n}^c, \Gamma_{.e}^c) = \Sigma_h \gamma_{he} n_h^c; \quad e, h \in \{\text{General, Reserved}\}$$

Assumption 3 :

(i) *Proximity Competition Effect:* $\gamma_{ee} > \gamma_{ef}$
(ii) *Symmetric Competition Effect:* $\gamma_{fe} = \gamma_{ef}$

$$e, f \in \{\text{General, Reserved}\}$$

Note that while γ_{ee} is expected to be positive, there is no clear prior on whether γ_{ef} should be negative or positive. Given these assumptions, the vote profit function can be rewritten as

$$\Pi_{pe}^c = X_e^c \beta + \xi^c + \mu_p + \tau_p^c - \Sigma_h \gamma_{he} n_h^c + \varepsilon_{pe}^c; \quad e, h \in \{General, Reserved\} \quad (2)$$

However, parties do not have perfect information about the ethnicity of the candidates that the rival's parties will choose. All parties move simultaneously when choosing their candidate. Therefore, they do not know the exact $\mathbf{n}^c$. They form expectations about the

number of other candidates of each ethnicity. Therefore, the vote profit function becomes

$$E(\Pi_{pe}^c) = X_e^c\beta + \xi^c + \mu_p + \tau_p^c - \Sigma_h\gamma_{he}E(n_h^c) + \varepsilon_{pe}^c; \quad e, h \; \epsilon \; \{General, Reserved\} \quad (3)$$

Party p's perception about party r's choice of candidate and, therefore, ethnicity is given by

$$p_{re} = Pr(d_{re}^c = 1|X_e^c, X_f^c, \mu_r, \tau_r^c) = Pr(E(\Pi_{re}^c) \geq E(\Pi_{rf}^c)); \quad e \neq f, r \neq p, \; \forall r \; \epsilon \; P \quad (4)$$

where e denotes general ethnicity and h denotes reserved ethnicity. Therefore,

$$p_{re} = Pr(X_e^c\beta - \Sigma_h\gamma_{he}E(n_h^c) + \varepsilon_{re}^c \; \geq \; X_f^c\beta - \Sigma_h\gamma_{hf}E(n_h^c) + \varepsilon_{rf}^c);$$
$$\forall r \; \epsilon \; P; \; e, f, h \; \epsilon \; \{General, Reserved\} \quad (5)$$

Since party p has the same information for all the other party condidadates, from party p's perspective,

$$p_{re} = p_{qe} = p_e; \quad \forall r, q \; \epsilon \; P; \; r, q \; \neq \; p \quad (6)$$

Given this belief of party p, the expectations that it will have about the number of candidates of any ethnicity h in a particular constituency will be given by

$$E(n_h^c) = (P - 1)p_h \quad (7)$$

where $(P - 1)$ is the number of other parties running a candidate against party p's candidate.

3 Equilibrium and Empirical Analogue

Assumption 4: ε's are i.i.d. draws from standard normal distribution (or extreme value type 1 distribution)[7].

Given this assumption on the distribution of error terms[8], equation (5) can be rewritten

---

[7]Both logit and probit estimation techniques are used and the results are very similar

[8]In fact, the assumption required is weaker than the one made here. The required assumptions are

as

$$p_{re} = p_e^c = Pr(\varepsilon_{re}^c - \varepsilon_{rf}^c \geq -(X_e^c\beta - \gamma_{ee}(P-1)p_e - \gamma_{fe}(P-1)p_f - X_f^c\beta$$
$$+ \gamma_{ef}(P-1)p_e + \gamma_{ff}(P-1)p_f)$$

$$= F((X_e^c - X_f^c)\beta - (\gamma_{ee} - \gamma_{ef})(P-1)(p_e - p_f)) \quad (8)$$

where $F()$ is the c.d.f. of normal or logistic distribution.

The situation can be described as a simultaneous move game where the probability term is the fixed point of a set of simultaneous equations similar to (8) and the solution is $p_{re} = p_e^c$. The outcome for each constituency, therefore, is one where each party does its best given what its rivals are doing. Rewriting X_f^c in terms of X_e^c, the above equilibrium generates the following comparative statics.

$$\frac{\partial p_{re}}{\partial X_e^c} > 0, \frac{\partial p_{re}}{\partial n_e^c} < 0, \quad (9)$$

$$\frac{\partial p_{re}}{\partial n_f^c} \gtreqqless 0, \;\; if \;\; \gamma_{ef} \gtreqqless 0 \quad (10)$$

The probability of choosing a candidate from a particular ethnicity, e, should be positively related to the population of voters of that ethnicity in the constituency. If there are more voters of a ethnicity e, all of whom have co-ethnic preference, it is more profitable to run a candidate of the same ethnicity. It should be negatively related to the number of other candidates with ethnicity e as they add to the competition for votes from the voters of that ethnicity. The impact of an increase in the number of candidates of another ethnicity, f, is not clear and depends on whether it increases or decreases the competition for the votes of ethnicity e. However, the competition effect of an additional candidate of some other ethnicity, if is negative ($\gamma_{ef} \geq 0$), should be less than that of other candidates from

---

discussed in greater detail in Section VI

the same ethnicity ($\frac{\partial p_{re}}{\partial n_f^c} > \frac{\partial p_{re}}{\partial n_e^c}$).

The selection process can be rewritten as

$$G_p^c = \begin{cases} 1 & \text{if } E(\Pi_{p,general}^c) \geq E(\Pi_{p,reserved}^c). \\ 0, & \text{otherwise} \end{cases}$$

or

$$G_p^c = \begin{cases} 1 & \text{if } X_{gen}^c\beta + \Sigma_h\gamma_h E(n_h^c) + \varepsilon_{p,general}^c \geq X_{reserved}^c\beta + \Sigma_h\gamma_h E(n_h^c) + \varepsilon_{p,reserved}^c. \\ 0, & \text{otherwise} \end{cases}$$

(11)

where G_p^c is a dummy variable that takes value one if part p chooses a general candidate in constituency c. We estimate this using a probit (or logit) specification.

4 Data and Setting

4.1 Elections in Uttar Pradesh and Bihar

The model is estimated using the parliamentary election(PE) and assembly elections(AE)[9] data from the states of Uttar Pradesh (UP) and Bihar in India. The states were chosen for two main reasons. First, ethnic rivalries and co-ethnic political preferences are much more salient for these states (Acharya et al. (2014), Vaishnav (2010a), Banerjee and Pande (2009),Chandra (2007), Witsoe (2005)). Second, UP is the most populous state of India and one with the largest number of parliamentary and assembly constituency. Bihar ranks third in terms of population and fourth in terms of number of parliamentary and assembly constituencies. Together, they account for around one fourth of the country's population. More than two-third of the population is engaged in agriculture in both states and the literacy levels are below the country average. While the population in India is diverse, UP and Bihar, arguably, are two of the good representations of the socio-economic and political climate.

---

[9]Parliamentary elections are for the selection of member of lower house(Lok Sabha) of the parliament of India from Uttar Pradesh while assembly elcetions are for the selection of members of the lower (Vidhan Sabha) house of the state.

Elections in the two states are dominated by the Hindus even though Muslim candidates have dominated the political results in a few regions. We do not make any distinctions between different religions for the purpose of this paper.[10] The elections in the state of Uttar Pradesh, both parliamentary and assembly level, have been dominated by four big parties - Bahujan Samaj Party (BSP), Samajwadi Party (SP), Indian National Congress (INC) and Bhartiya Janta Party (BJP). Amongst these, INC and BJP are national parties but constituencies in UP have been the seat for their top candidates in almost all years in the PE. All four parties have some sort of ethnic affiliation. BSP and SP mainly target the reserved ethnicity (lower caste) populations while INC and BJP, allegedly, represent the general ethnicity (upper caste) population. While the parties might claim affiliations to certain caste groups, a clear ethnicity (caste) affiliation is not observed in their actions. This is clear from the recent election results from the state. The Scheduled Caste (SC's) and Scheduled Tribes (ST's) together form around one-fourth of the state's population. The Other Backward Castes (OBC's) form around 40% of the state's population. BSP has explicitly identified itself to be representing the SC's and ST's while the SP has been said to be targeting the OBS votes. However, strict affiliations to the groups could not have produced the results that has been witnessed in the recent past. In the 2007 elections, BSP won 206 out of the 403 seats in the AE, a clear majority while the SP won 97 seats to finish second. In the 2012 AE, however, SP won 224 seats while the BSP bagged 80 seats in total. If voter's co-ethnic preference are taken seriously, this also means that there must have been a good degree of switching of candidate ethnicities by both the parties. In fact, BSP 2007 victory was as a result of the fact that they managed to bag a good amount of general votes by running general (and in particular, brahmin) category candidates even after their historically antagonistic attitude towards the general ethnicity.[11] It is this strategic behaviour that the present model is trying to capture. Bihar is more competitive in terms of the number of major parties running candidates from at least half the total number of constituencies. The state politics is dominated by Rashtriya Janta Dal (RJD), Janta Dal (United), INC and BJP but has considerable presence of BSP, SP, Lok Janshakti Party (LJP), National Congress Party(NCP) and Communist Party of India (Marxist-Leninist)

---

[10]The fragmentation along the religion dimension and it's complex relationship with the politics in the country is interesting in its own right and a potential subject of future research.

[11]Acharya et al. (2014) The BSP took this to an extreme in its 2002 slogan "Thrash the Brahmin, the Bania and the Rajput" (translated from *Tilak, tarazu aur talwar, Inko maro joote chaar*(Jain 1996, p. 215)

Liberation. While LJP enjoys considerable popularity amongst the SC's, RJD is popular amongst the OBC group.

There are 80 parliamentary and 403 assembly constituencies in the state of UP and 40 parliamentary and 243 assembly constituencies in Bihar. Around a fifth of these constituencies are reserved for the SC/ST candidates meaning that no general or OBC candidates can run for office from these constituencies. This is a part of the affirmative action programme in the country. Of course, there is no choice for political parties but to run a candidate from the reserved ethnicity from these constituencies. Therefore, we drop these constituencies from our sample. Next, we also drop the independent candidates from our sample.[12] We do this because an independent candidate has no two ethnicities to choose from. For him, there is no such strategic concern. Of course, there is a first stage to the game presented here where people and parties decide whether to run the elections from a constituency or not. While that may be the topic of future research, we abstract away from that stage for the purpose of this model. We estimate our model separately for the parliamentary and assembly elections but combine the data from the two states.

Data from 2014 and 2009 PE are used. AE elections data come from 2010 elections in Bihar and 2012 elections in UP. We use information from elections after the most recent delimitation of constituencies that started in 2008. The constituency boundaries before the 2008 delimitation had remained unchanged since 1976 and were based on 1971 census population figures. The census units do not fit the assembly and parliamentary boundaries perfectly (See Alam (2010), and the literature it cites.). Therefore, it is difficult to get reliable estimates of constituency specific population by ethnic groups for elections before 2008. The new delimitation was carried out on the basis of 2001 census as per the the Delimitation of Parliamentary and Assembly constituencies order, 2008, and contains information on constituency specific population by three major ethnicity groups. The election results data was obtained from the Election Commission of India's website (http://eci.nic.in/). The elections data, among other information, contain, for each constituency, the ethnicity of the candidates running for office from that constituency and an identifier for the winner from each constituency.

The population composition data was obtained from reports by the Delimitation Commission of India that provide information on the population by ethnic groups for the newly

---

[12]Independent candidates are those who are not affiliated to any party.

delimited constituencies. The delimitation committee followed the 2001 census data while redrawing the new constituency boundaries. The reason I stick to the 2001 census figures used by the Delimitation Commission of India instead to using the 2011 census figures is that, given the lack of fit between census building blocks and delimitation building blocks, there is no easy way to map the 2011 census figures to the assembly and parliamentary constituency boundaries. The candidate characteristic data come from a series of affidavits released as a result of a Public Interest Litigation filed with the Delhi High court. They contain information on age, assets, liabilities, education attainment and criminal cases filed against the candidate.

4.2 Pradhan Mantri Gram Sadak Yojana

The most widely accepted explanation put forth for ethnic preferences is that in situations voters do not have perfect information about the policy stand of the candidates or the candidates cannot make policy commitments, voters use ethnicity as a signal [Banerjee and Pande (2009), Chandra (2007)]. To test for the importance of information in formation of co-ethnic preferences, I utilize information on the Pradhan Mantri Gram Sadak Yojna (PMGSY). PMGSY is a federal madated rural road construction program that started in 2000 and is still underway. As of 2014, the program had costed more than four billion dollars and constructed 75 thousand kilometers of all weather roads in the states of Uttar Pradesh and Bihar combined. The program aims to connect all habitations with a population of at least 500 to the nearest link road via an all-weather road.[13] Aggarwal (2014) and Banerjee and Sachdeva (2015) describe the program in detail and evaluate the impact of the program along different dimensions of development. Aggarwal (2014) finds that the program has substantially increased the adoption of agricultural technologies and has induced teenagers to drop out of school to join the labor force to access expanded labour market opportunities by reduction of time, money and information cost of accessing markets. Banerjee and Sachdeva (2015) find that the program has increased the awareness about government run health care programs, improved health care supply, and increased social interaction within and between villages. Both these studies suggest that the flow of

---

[13]"A habitation is a sub-village level entity, and is defined as 'a cluster of population, whose location does not change over time'" (Aggarwal (2014))

information has improved significantly due to the program. The program, since it makes use of the exogenously determined rule, is exogenous to the process of candidate selection by political parties for each constituencies. Since the program also spelled out clearly that construction was to be prioritized using population based rules, the possibility of some 'favoured' regions receiving the program before others is also ruled out.

The Online Management and Monitoring System (OMMS), a recent initiative of the Government of India, has made all data regarding large public program freely accessible. Data on PMGSY road construction development are available at the habitation level. Along with the status of connectivity during the 2001 census and status as in May 2016, data also list the village, the block, the district, the assembly constituency, and the parliamentary constituency to which the habitat belongs. Unfortunately, the boundaries of the assembly and parliamentary constituencies have changes since then with the delimitation process starting in 2008 under the Delimitation Act of 2002. For this reason, there was a need to match the habitations with the new assembly and parliamentary constituencies. This is a daunting task since the administrative boundaries of the country, that is the delineation of district, sub-district and block level boundaries, do not bear a close relationship to the parliamentary and assembly constituencies' boundaries. While the elections data is available for these electorally demarcated regions, the information on information on socioeconomic indicators are usually available at the level of administrative units of districts and sub-districts. This 'lack of fit between the two vital maps of India', as Alam (2010)puts it, has not only caused 'administrative and political problems' (see Sivaramakrishnan (1997), Sivaramakrishnan (2000), Sivaramakrishnan (2001)), but has also created 'some major obstacles .... for students of public policy, politics and political economy' of India. As a results, there have been numerous attempts at generating socio-economic profiles of parliamentary and assembly constituencies, each with its own weaknesses and pitfalls (See Alam (2010), Bhandari (2009) and the literature cited in Alam (2010) for details.) Alam (2010) also details a new methodology employed by the the Lokniti program of the Centre for the Study of Developing Societies (CSDS). While the data set the method generates it yet to be critically evaluated, it is, by far, one of the most comprehensive attempt at this exercise.

This paper, however, uses slightly different approach, perhaps most similar to Brass (1975). The basic building block for the delimitation exercise in Bihar are 'wards' in the

urban area and community development blocks, or gram panchayats in rural areas. In Uttar Pradesh, these are 'wards' in the urban area and tehsils, kanungo cicles, or patwari circles in rural areas. A gram panchayat or a patwari circle, the smallest unit among the rural units depending on the state we are talking about, or a ward, the smallest urban unit, can consist of one or more habitations.[14] Data on the PMGSY road construction program is available at the habitations level. While the data do contain information on the assembly and parliamentary constituency to which a habitat belongs, this information is available for before the delimitation exercise began in 2008. The habitations, therefore, need to be matched with their new constituency. There are at least two reasons that make it extremely difficult. First, the delimitation commission's order detailing the process, the Delimitation of Parliamentary and Assembly constituencies order, 2008, does not provide with detailed information about the new constituency boundaries at the level of habitations. The information is mostly at the level of wards and blocks or tehsils. Certain details are available at the kunungo or garm panchayat level only in cases where a tehsil or a block is being split among more than one constituency. Second, these gram panchayats, kunungos, blocks or tehsils do not have unique codes that can be used to match the information in the delimitation order with that obtained from OMMS. The fact that there are more than 330 thousand habitations in the two states combined makes a manual matching impossible. For this region, I collapse the information from OMMS at the block level.

I then assign each block, which can contain multiple villages, to a single assembly constituency. While most of the blocks in the PMGSY data set belong entirely to one assembly constituency, some of the blocks are split among more than one assembly constituency. Such occurrences are relatively higher in Uttar Pradesh. For these blocks, I assign them to the assembly constituency which contains, within it, a larger number of gram panchayats and habitations from the block. While this will lead to a wrong assignment of certain gram panchayats to a constituency they do not belong to, a majority blocks in the data set are not split between constituencies. [15] Furthermore, there is no reason to believe that the measurement error in this assignment method is systematically related to the candidate choice by political parties across constituencies. It is, therefore, reasonable

---

[14]While tehsils are larger than commuity development blocks, for the sake of brevity, I will use 'blocks' to refer to both tehsil and community development blocks. Similarly, I will use 'villages' to refer to both patwari circles and gram panchayats.

[15]For example, in Bihar, over 80 percent of the blocks were fully contained in one assembly constituency.

to assume that the coefficient estimates will remain unbiased. Mapping the blocks to the parliamentary constituencies accomplished by matching their assembly constituencies to each assembly constituency's corresponding parliamentary constituencies, the details of which are contained in the Delimitation of Parliamentary and Assembly constituencies order, 2008.

For each assembly and parliamentary constituency, using the information from OMMS, I calculate the percentage of habitations that were connected to the towns via an all-weather road in 2001 and the percentage connected by May of 2016. I use the difference between these two figures, the increase in connectivity between 2001 and 2016, as a proxy for the increase in availability of information within the constituencies. For each assembly (parliamentary) constituency, a dummy variable 'High connect' takes a value of '1' if the particular constituency's increase in connectivity between 2001 and 2016 is more than the median increase in connectivity across assembly (parliamentary) constituencies in the state, and '0' otherwise. If the model fits the data well, interaction of 'High connect' with the main variables of interest in 11 should work towards muting the role of the main variables in the equation.

5   Results

Table 2 presents the results using the 2014 and 2009 PE elections data for the two states. As the model suggests, higher the percentage of general ethnicity voters in the constituency, higher is the probability that a party will field a general candidate The number of other general candidates significantly lowers the probability that a party will choose to field a candidate of general ethnicity. The coefficient for other reserved category candidates variables turns out to be insignificant. The direction, however, suggests that candidates from reserved ethnicity add to the competition that general candidates face. Consistent with the prediction from the model, the magnitude of this coefficient is smaller than that of other general candidates variable's coefficient. The third and fourth column includes only those parties which field candidates from at least two constituencies. It is possible that the party contesting elections from only one constituency is not big enough to have candidates from both ethnicities and, therefore, may be less likely to participate in this

game.[16] Overall, there is evidence in favour of the model at the parliamentary level.

Table 3 presents the results for the assembly results using data from 2010 Bihar AE and 2012 UP AE. While the general ethnicity voter percentage is still significantly and positively related to the probability that a party will chose a general ethnicity candidate, number of other general candidates does not affect the choice even though the sign of the coefficient is consistent with the model's prediction. Surprisingly, the magnitude of other reserved ethnicity candidates variable is not only significant but bigger in magnitude compared to that of other general ethnicity candidates variable. The results for assembly elections seem to be relatively weak but not very perplexing if we focus on the reasons why voters have co-ethnic political preferences. The most widely accepted explanation put forth for ethnic preferences is that in situations voters do not have perfect information about the policy stand of the candidates or the candidates cannot make policy commitments, voters use ethnicity as a signal [Banerjee and Pande (2009), Chandra (2007)].

Ethnicity of a candidate may serves as a signal of policies that are beneficial for and protect the interests of that ethnicity [Vaishnav (2010b). Assembly constituencies are much smaller compared to the parliamentary constituencies. Consequently, the voters might be better informed about the policy stand of the candidate in an assembly constituency due to more frequent interactions with the candidates. Also, alternative measures to influence voter, like election campaigning, might be much more effective at the assembly constituency level due to the higher acquaintance of the candidate with the issues facing that constituency. Heath and Kumar (2012) find that the SC and ST population of the state of Uttar Pradesh was unsatisfied with the performance of BSP government after the 2007 elections. This suggests the possibility that voters might have had greater information about the workings of the state government. Therefore, one needs to account for additional factors when trying to analyse voter preferences in state assemblies. The ethnic political preference may break down to a good extent in such cases. This is consistent with previous literature on why voters vote ethnically Chandra (2007). Also, most of the policies that pertain to ethnic identities are debated on, formulated and legislated at the parliament level and not in the state assemblies. Therefore, the voters might be of the perception that having co-ethnics in the parliament matters more than having co-ethnics in

---

[16]Using larger cutoffs for this size variable does not affect the results.

the state assembly.[17] The political parties might internalize this information as well when running a candidate for assembly elections. Assembly constituencies, therefore, will have a higher degree of presence of cross-cutting cleavages - dimensions of identity or interest along which members of the same ethnic group may have diverse allegiances.

To test for the importance of information, we interact the dummy variable 'High connect', denoting high increase in road connectivity of the constituency between 2001 and 2016, described in 4.2, with the main variables of our interest - percentage of general ethnicity voters, number of other general candidates and the number of other reserved category candidates. The results are presented in 4. Column (1) presents the results for the parliamentary elections. For the regions with low increase in connectivity, the coefficients are in the expected direction. The coefficient for other general candidates variables drops below that for other reserved candidates variables even though the former is not significant. Interestingly, the coefficient for interaction of high connectivity dummy with the main variables suggest that the model is far less representative of the candidate selection procedure from these regions. The interaction of 'High connect' with percentage of general voters variable and with other reserved candidates variable have coefficients with signs opposite of their independent effects. If PMGSY indeed increased the extent to which the voters in the constituencies had access to relevant information, the results suggest that political parties do not utilize strategies that rely on co-ethnic voting to choose candidates in regions where voters are more informed. It could be, as others have argued, because of a break down of the co-ethnic preferences in these regions or could simply be because political parties start believing that voters, on an average, do not vote ethnically in these regions. While there is no way to test the competing hypotheses, the former seems more likely *a priori*. For assembly elections, presented in column (2), the independent coefficients are in the same direction and of a similar magnitude as those in 3. The interactions, however, are close to zero in magnitude and insignificant. This is consistent with the results from 3 that suggests parties, to woo voters, do not rely as much on choosing co-ethnic candidates in assembly elections as they do in parliamentary elections. As explained earlier, assembly constituency voters might be better informed about the policy stand of the political parties

---

[17] A similar argument has been made in Jaffrelot (2003) and Chandra (2007). They argue that the political salience of caste identities become more pronounced in contexts where affirmative action policies are involved. Historically, in post independence India, question of seat reservations and quotas for people to redress caste discrimination are debated at the parliamentary level.

and candidates and may not need to use ethnicity of candidates as signals.

We now try and distinguish between the inefficiency arising from voter's voting ethnically and from political parties' belief that voters vote ethnically. To do that we use information on candidate characteristics from PE 2014 and 2009 for both states, from AE 2012 for UP and from AE 2010 for Bihar. We use the number of criminal cases ever filed against the candidate as our indicator of quality of the candidate. Tables 5 and 6 present the results. Columns (1) from the tables show that constituencies reserved for reserved ethnicity candidates tend to have a candidates with lower criminal cases filed against them running for office during these elections. Vaishnav (2010a), too, finds that candidates are more likely to have criminal records in general constituencies. Such an analysis, however, does not distinguish between the two reasons that can lead to, on an average, perverse quality of elected representatives in unreserved constituencies. This can happen if the voters from unreserved constituencies vote ethnically disregarding the criminal record of the candidates or if the pool of candidates they choose from is worse than the pool from which the voters in reserved constituencies choose from. Columns (2) from the tables show that if one controls for the proportion of criminal candidates, reserved constituencies are not very different from general constituencies in their chances of electing a criminal candidate. While the coefficients on 'reserved constituency dummy' suggest that reserved constituencies are still less likely than the unreserved ones to select criminal candidates, the magnitudes are now smaller, especially for the parliamentary elections where, as per our results, co-ethnic voting might be more of a problem. The proportion of criminal candidates, however, is a strong predictor of whether the elected candidate has a criminal record or not. This is indicative of the fact that the higher on average criminal records of the elected representatives in unreserved constituencies is because of parties running worse candidates. Due to their beliefs about the voter's co-ethnic preference, they might ignore the potential candidate's criminal record as they expect the voters to do that as well. This second kind of inefficiency pointed to in this paper, therefore, seems to be more important than the one pointed out in previous literature.

6 Revisiting the assumptions

This sections looks at the assumptions made in the model and during the estimation. The assumption that is at the very base of the model is that all that political parties care about is the aggregate amount of vote that they win in a constituency. While this assumption is readily accepted in the literature, one can argue that parties like BSP and SP have strong ethnic affiliations and might choose candidates in accordance with that affiliation regardless of the vote profits from that candidates in order to better represent the ethnic interests. There are at least three arguments that can be made against it. First is the evidence from 2007 UP assembly elections where the SP chose to run many candidates from the general category even though it is known for its strong affiliation to the SC's and antagonistic outlook towards the general candidates. Second, for these parties which have strong affiliations, the focus seem to be on having representatives for the ethnicity in the government and not on having ethnic representative in each constituency. A party is open to running a candidate from an ethnicity it is not affiliated to as long as the majority of candidate it gets to form a ministry with are from the ethnicity it is affiliated to; so the ethnic affiliations might not express themselves when considering a constituency in separation. Third, ethnic affiliations can be and are incorporated in this model. Consider this hypothetical case where three parties are contesting the election from a particular constituency. Out of the three, one party has very strong ethnic affiliation. In that case, the other two parties may be quite certain that the first party will run a candidate of ethnicity it is affiliated to. The game is then played between the other two parties where both parties incorporate this decision of the first party in their evaluation of the costs and benefits of running a candidate of a particular ethnicity. The model works as long as there is at least one player acting strategically. The higher the number of parties with ethnic affiliations, the lower the significance of the coefficients from the model.

One could argue that the assumption of asymmetric information is not as realistic and that political parties might have all the relevant information about all the potential candidates that every rival party is considering to run from a particular constituency. This is highly unlikely. Most often, parties do not have very good information about the identity of all the potential candidates that a rival party is considering, let alone having detailed information about them. But suppose some party did have this information about even

one rival party. This party will be able to calculate exactly what candidate the rival party will be running from this constituency. Again, this party will start treating the rival party's choice as given and act strategically against all other rival parties. Such a change will be subsumed in the constituency fixed effect (constituency specific characteristics) in our model. The model holds as long as there is at least one party without complete information about all the candidates of at least one rival party.

A more plausible situation is that all parties might know exactly what candidate a particular party is going to run from a constituency. For example, in our context, it is known to everyone that parliamentary elections from Amethi for INC are almost always contested by the most influential member of the Gandhi family. In such a scenario, all the other parties might make expectations about the actions of all parties except INC and the game is played amongst them. Such a change will also be accounted for in our model due to the inclusion of constituency fixed effects.

Another assumption that is being made is about the distribution of i.i.d. error terms. An argument that can be made against the model is that parties might be drawing their general and reserved candidates with distributions that have very different means across parties within a constituency and across ethnicities within a constituency within a particular party. The reason why this will not necessarily violate the identical error distribution can be better explained using an example. Assume that there are two parties, A and B, contesting elections from a particular constituency. Let us assume that the mean of the distribution from which party A chooses its reserved category candidate is $x + y$ and the mean of the distribution it chooses its general category candidate is $x + y + z$. The means of the distributions from which party B chooses its reserved and general candidates are x and $x + y + z + w$, respectively. Note that the difference in means of the two distributions for part A and B are z and $w + z$, respectively. However, once we include the party-constituency fixed effect, the difference in means of the remaining errors, the ε's is z for both the parties. The only assumption that is being required for the estimation is that this difference z follows a normal (or logistic) distribution across all constituencies. Results are clustered at the level of constituency to account for within constituency correlation.

Assumptions 2 and 3 have been made to simplify the model and limit the number of parameters to be estimated. The empirical results should not be very different qualitatively even if other alternative assumptions are made.

7 Conclusion

Banerjee and Pande (2009) argue that ethnic biases may result in selection of lower quality politicians. Vaishnav (2010a) shows that parties are less likely to put up candidates with criminal backgrounds in the reserved constituencies. The results here are consistent with both these findings. In constituencies which are reserved for SC/ST candidates, diversification along the ethnicity dimension is not possible. Therefore, the competition is more intense. As a result, parties ensure that the candidates they run are not 'tainted'. Voters, too, pay attention to these other dimensions in these constituencies. In constituencies which are not reserved, parties try to exploit these co-ethnic preferences and, in the process, may be willing to compromise on other dimensions, like quality. While reserving all constituency for one or the other ethnicity might work, it is not realistic and may work towards making these ethnic rivalries more pronounced. To address this inferior candidate selection, we will need to address the co-ethnic political preference. One way in which that can be done is by providing the voters with more information about the candidate. If, as argued by Chandra (2007), ethnicity works as signals, this will weaken the co-ethnic preferences. There is some indication of that in the AE results presented here.

While this model seems to fit the data well, it is by no means an accurate or complete description description of the complex selection procedure for the candidates that a party uses. The selection process is an immensely complicated and a more comprehensive model should allow for other forces, like religion, incumbency, etc., to play a role. This paper abstracts away from those dimensions and attempts to model the candidate selection process in a simple way possible while still managing to explain certain stylized facts about elections in India. Moreover, the selection process described by the model is more of a second stage to a selection of parties into constituency, which is of great interest and should be the topic of future research. Also, the paper does not take into account time varying factors that may influence the candidate choice.[18] It is possible, for example, that the ethnic preference of the voters evolve in a certain manner over time. A more sophisticated model should attempt to account for these.

---

[18] A large number of factor besides information could affect co-ethnic preferences and candidate choice over time.

References

Acharya, A., Roemer, J. E., and Somanathan, R. (2014). Caste, corruption and political competition in India.

Aggarwal, S. (2014). Do rural roads create pathways out of poverty? evidence from india. *University of California, Santa Cruz, unpublished*.

Alam, M. S. (2010). On matching census tracts and electoral boundaries: The bottom-up aggregation approach. *Economic and Political Weekly*, pages 64–72.

Banerjee, A., Green, D., Green, J., and Pande, R. (2010). Can voters be primed to choose better legislators? experimental evidence from rural india. In *Presented and the Political Economics Seminar, Stanford University*.

Banerjee, A. and Pande, R. (2009). Parochial politics: Ethnic preferences and politician corruption. *mimeo*.

Banerjee, R. and Sachdeva, A. (2015). Pathways to preventive health, evidence from india's rural road program. *USC-INET Research Paper*, (15-19).

Barreto, M. A. (2007). I sí se puede! latino candidates and the mobilization of latino voters. *American Political Science Review*, 101(03):425–441.

Bates, R. H. (1983). Modernization, ethnic competition, and the rationality of politics in contemporary africa. *State versus ethnic claims: African policy dilemmas*, 152:171.

Besley, T., Persson, T., and Sturm, D. (2005). Political competition and economic performance: Theory and evidence from the united states. Technical report, National Bureau of Economic Research.

Bhandari, L. (2009). Socio-economic performance of constituencies: A response [with reply]. *Economic and Political Weekly*, pages 61–63.

Brass, P. R. (1975). Ethnic cleavages and the punjab party system, 1952-1972. *Studies in Electoral Politics in the Indian States: The Impact of Modernisation*, 4:3–62.

Chandra, K. (2007). *Why ethnic parties succeed: Patronage and ethnic head counts in India*. Cambridge University Press.

Chauchard, S. (2016). Unpacking ethnic preferences theory and micro-level evidence from north india. *Comparative Political Studies*, 49(2):253–284.

Dahl, R. A. (2005). *Who governs?: Democracy and power in an American city*. Yale University Press.

Downs, A. (1957). An economic theory of political action in a democracy. *The Journal of Political Economy*, pages 135–150.

Eifert, B., Miguel, E., and Posner, D. N. (2010). Political competition and ethnic identification in africa. *American Journal of Political Science*, 54(2):494–510.

Field, B. N. and Siavelis, P. M. (2011). Endogenizing legislative candidate selection procedures in nascent democracies: evidence from spain and chile. *Democratization*, 18(3):797–822.

Heath, O. and Kumar, S. (2012). Why did dalits desert the Bahujan Samaj Party in Uttar Pradesh? *Economic and Political Weekly*, 47(28).

Horowitz, D. L. (1985). *Ethnic groups in conflict*. Univ of California Press.

Hotelling, H. (1929). An empirical model of firm entry with endogenous product-type choices. *Economic Journal*, 39:41–57.

Jaffrelot, C. (2003). *India's silent revolution: the rise of the lower castes in North India*. Orient Blackswan.

Kaufmann, K. M. (2003). Black and latino voters in denver: Responses to each other's political leadership. *Political Science Quarterly*, 118(1):107–125.

Key, V. (1949). Southern politics in state and nation.

McClain, P. D. and Garcia, J. A. (1993). Expanding disciplinary boundaries: Black, latino, and racial minority group politics in political science. *Political science: The state of the discipline II*, 2:247.

Seim, K. (2006). Stability in competition. *The RAND Journal of Economics*, 37(3):619–640.

Sivaramakrishnan, K. (1997). Under-franchise in urban areas: Freeze on delimitation of constituencies and resultant disparities. *Economic and Political Weekly*, pages 3275–3281.

Sivaramakrishnan, K. (2000). North-south divide and delimitation blues. *Economic and Political Weekly*, pages 3093–3097.

Sivaramakrishnan, K. (2001). Constituencies delimitation: Deep freeze again? *Economic and Political Weekly*, pages 4694–4696.

Stokes, D. E. (1963). Spatial models of party competition. *The American Political Science Review*, pages 368–377.

Vaishnav, M. (2010a). The market for criminality: Money, muscle and elections in India. *Unpublished paper*.

Vaishnav, M. (2010b). The market for criminality: Money, muscle and elections in india. *Unpublished paper*.

Vicente, P. C. and Wantchekon, L. (2009). Clientelism and vote buying: lessons from field experiments in african elections. *Oxford Review of Economic Policy*, 25(2):292–305.

Witsoe, J. (2005). *Democracy against the state: the politics of caste empowerment in Bihar*. PhD thesis, University of Cambridge.

Young, C. (1979). *The politics of cultural pluralism*. Univ of Wisconsin Press.

Table 2: Parliamentary Elections

VARIABLES	(1) (PE: Probit) General candidate selected	(2) (PE: Logit) General candidate selected	(3) (PE ($\geq$ 2):Probit) General candidate selected	(4) ((PE ($\geq$ 2): Logit) General candidate selected
General Voters (%)	0.0256***	0.0502***	0.0280***	0.0550***
	(0.0086)	(0.0168)	(0.0094)	(0.0182)
Other general candidates	-0.0407***	-0.0797***	-0.0318*	-0.0635**
	(0.0154)	(0.0306)	(0.0163)	(0.0321)
Other reserved candidates	-0.0266	-0.0574	-0.0145	-0.0335
	(0.0504)	(0.0993)	(0.0555)	(0.1116)
Constant	-0.3424	-0.9823	-0.5904	-1.4720
	(0.7011)	(1.3630)	(0.8083)	(1.5345)
Observations	2,071	2,071	1,908	1,908
YEAR FE	YES	YES	YES	YES
STATE FE	YES	YES	YES	YES

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. The standard errors are clustered at parliamentary constituency level.

27

Table 3: Assembly Elections

VARIABLES	(1) AE (All): Probit General candidate selected	(2) AE (All): Logit General candidate selected	(3) (AE $\geq$ 2): Probit General candidate selected	(4) ((AE $\geq$ 2): Logit General candidate selected
General Voters (%)	0.0207***	0.0427***	0.0207***	0.0428***
	(0.0039)	(0.0078)	(0.0039)	(0.0079)
Other general candidates	-0.0088	-0.0174	-0.0086	-0.0171
	(0.0075)	(0.0147)	(0.0075)	(0.0148)
Other reserved candidates	-0.0839***	-0.1617***	-0.0862***	-0.1666***
	(0.0259)	(0.0502)	(0.0262)	(0.0508)
Constant	0.0452	-0.3643	0.0478	-0.3728
	(0.3142)	(0.6309)	(0.3164)	(0.6349)
Observations	6,117	6,117	6,053	6,053
STATE FE	YES	YES	YES	YES

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Robust standard errors in parentheses. The standard errors are clustered at assembly constituency level.

28

Table 4: PMGSY and Candidate Selection

VARIABLES	(PE) General candidate selected	(AE) General candidate selected
General Voters (%)	0.0403***	0.0281***
	(0.0127)	(0.0062)
Other general candidates	-0.0323	-0.0107
	(0.0262)	(0.0113)
Other reserved candidates	-0.0786*	-0.0711*
	(0.0409)	(0.0380)
High Connect	3.5687**	-0.1074
	(1.5184)	(0.8011)
General Voters * High Connect (%)	-0.0413**	0.0008
	(0.0187)	(0.0102)
Other general candidates * High Connect	-0.0579	0.0005
	(0.0412)	(0.0146)
Other reserved candidates * High Connect	0.3758***	0.0478
	(0.1390)	(0.0625)
Constant	-1.5579	-0.5609
	(1.0356)	(0.4969)
Observations	2,002	5,326
YEAR FE	YES	NO
STATE FE	YES	YES

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. The standard errors are clustered at the constituency level.

Table 5: Criminal candidates in PE

VARIABLES	(1) Criminal cases against candidate	(2) Criminal cases against winner
Reserved constituency dummy	-0.5046***	-0.2663
	(0.1161)	(0.6463)
Proportion of criminal candidates		5.2378**
		(2.1388)
Constant	0.7502***	1.8291
	(0.2382)	(1.5422)
Observations	2,386	237
R-squared	0.1802	0.1160
STATE FE	YES	YES
YEAR FE	YES	YES

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. The standard errors are clustered at parliamentary constituency level.

Table 6: Criminal candidates in AE

VARIABLES	(1) Criminal cases against candidates	(2) Criminal cases against winner
Reserved constituency dummy	-0.3946***	-0.3883
	(0.0392)	(0.2931)
Proportion of criminal candidates		2.1978**
		(0.8639)
Constant	1.0098***	-0.2619
	(0.0988)	(0.4959)
Observations	8,979	625
R-squared	0.0764	0.2044
STATE FE	YES	YES

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. The standard errors are clustered at assembly constituency level.