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

Arvind Magesan
Economics Department, University of Calgary

Andrea Szabó
Economics Department, University of Houston

Gergely Ujhelyi
Economics Department, University of Houston

June 20, 2024

Contents

1 Data and sample

1.1 Construction of the Muslim characteristic

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

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

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

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

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

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

1.2 Sample construction

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

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

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

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

1.3 Missing characteristics

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

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

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

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

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

2 Specification choice for voter preferences

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

Column 3 of Table A.1 shows a random coefficients specification where voter demographics $\mathbf{d}_i$ in equation (6) in the paper are replaced with random variables drawn from a standard normal distribution. We use a separate i.i.d. variable for each of the four candidate characteristics, and estimate this specification using the BLP procedure with the differentiation IVs as instruments. This specification indicates the presence of significant heterogeneity in voters' preference for candidate assets, but not for the other three characteristics.

Table A.1: Specification choice: differentiation IVs and random coefficients

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

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

Table A.2: Specification choice: differentiation IVs and constituency demographics

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

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

3 Identification of model parameters

Here we formally establish identification of the parameters of the model of candidate selection and discuss the intuition of the identification results. Throughout, we assume that choice probabilities, win probabilities, and expected vote shares are known to the researcher - these are estimated in a first stage. Let the probability that party $p \in \{1, 2\}$ chooses action $a_p = k$ for $k = 1, ..., K$ given observable payoff variables $\mathbf{z}$ (i.e., constituency characteristics) be given by $P_p(k, \mathbf{z})$, and write expected winning probability of party p as:

$$w_p^P(k, \mathbf{z}) = E_p[w_p(a_p, a_{-p}, \mathbf{z})|a_p = k] \quad (1)$$

where the expectation $E_p[w_p(a_p, a_{-p}, \mathbf{z})|a_p = k]$ is an integration over a_{-p} using player $-p$'s choice probability (see Section 3). Similarly, write the expected vote share as:

$$s_p^P(k, \mathbf{z}) = E_p[s_p(a_p, a_{-p}, \mathbf{z})|a_p = k] \quad (2)$$

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

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

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

$$P_p(k, \mathbf{z}) = \Lambda\big(b^w \times w_p^P(k, \mathbf{z}) + b^s \times s_p^P(k, \mathbf{z})\big) \tag{3}$$

where, given our assumption about the error distribution:

$$\Lambda\big(b^w \times w_p^P(k, \mathbf{z}) + b^s \times s_p^P(k, \mathbf{z})\big) = \frac{\exp\left\{b^w \times w_p^P(k, \mathbf{z}) + b^s \times s_p^P(k, \mathbf{z})\right\}}{\sum_{k'} \exp\left\{b^w \times w_p^P(k', \mathbf{z}) + b^s \times s_p^P(k', \mathbf{z})\right\}} \tag{4}$$

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

Inverting the choice probability gives:

$$\begin{aligned} \Lambda^{-1}\big(P(k, \mathbf{z})\big) &= \ln\big(P(k, \mathbf{z})\big) - \ln\big(P(K, \mathbf{z})\big) \\ &= b^w \times (w^P(k, \mathbf{z}) - w^P(K, \mathbf{z})) + b^s \times (s^P(k, \mathbf{z}) - s^P(K, \mathbf{z})) \end{aligned} \tag{5}$$

where we have taken type K as the reference type.

Define:

$$\Delta_w^P(k, \mathbf{z}) \equiv w^P(k, \mathbf{z}) - w^P(K, \mathbf{z}) \tag{6}$$

$$\Delta_s^P(k, \mathbf{z}) \equiv s^P(k, \mathbf{z}) - s^P(K, \mathbf{z}) \tag{7}$$

The difference $\Delta_w^P(k, \mathbf{z})$ represents the *increased expected probability of winning when selecting type k relative to the reference type K*, and similarly for $\Delta_s^P(k, \mathbf{z})$.

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

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

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

This is the case if,:

$$\frac{\Delta_w^P(k,\mathbf{z}^{(1)})}{\Delta_w^P(k,\mathbf{z}^{(2)})} \neq \frac{\Delta_s^P(k,\mathbf{z}^{(1)})}{\Delta_s^P(k,\mathbf{z}^{(2)})} \tag{9}$$

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

Solving the system in (8) we get:

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

and:

$$b^s = \frac{\Delta_w^P(k,\mathbf{z}^{(1)})\Lambda^{-1}(P(k,\mathbf{z}^{(2)})) - \Delta_w^P(k,\mathbf{z}^{(2)})\Lambda^{-1}(P(k,\mathbf{z}^{(1)}))}{\Delta_s^P(k,\mathbf{z}^{(2)})\Delta_w^P(k,\mathbf{z}^{(1)}) - \Delta_s^P(k,\mathbf{z}^{(1)})\Delta_w^P(k,\mathbf{z}^{(2)})} \tag{11}$$

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

$$\frac{\Delta_w^P(k,\mathbf{z}^{(1)})}{\Delta_w^P(k,\mathbf{z}^{(2)})} > \frac{\Delta_s^P(k,\mathbf{z}^{(1)})}{\Delta_s^P(k,\mathbf{z}^{(2)})} \tag{12}$$

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

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

$$\frac{\Lambda^{-1}(P(k,\mathbf{z}^{(1)}))}{\Lambda^{-1}(P(k,\mathbf{z}^{(2)}))} > \frac{\Delta_s^P(k,\mathbf{z}^{(1)})}{\Delta_s^P(k,\mathbf{z}^{(2)})} \tag{13}$$

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

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

$$\frac{\Lambda^{-1}(P(k,\mathbf{z}^{(2)}))}{\Lambda^{-1}(P(k,\mathbf{z}^{(1)}))} > \frac{\Delta_w^P(k,\mathbf{z}^{(2)})}{\Delta_w^P(k,\mathbf{z}^{(1)})} \tag{14}$$

The intuition is similar here. We have assumed

$$\frac{\Delta_w^P(k,\mathbf{z}^{(1)})}{\Delta_w^P(k,\mathbf{z}^{(2)})} > \frac{\Delta_s^P(k,\mathbf{z}^{(1)})}{\Delta_s^P(k,\mathbf{z}^{(2)})} \tag{15}$$

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

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

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

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

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

$$\begin{aligned} \Lambda^{-1}\big(P(k,\mathbf{z})\big) &= \ln\big(P(k,\mathbf{z})\big) - \ln\big(P(K,\mathbf{z})\big) \\ &= b^w \times (w^P(k,\mathbf{z}) - w^P(K,\mathbf{z})) + b^s \times (s^P(k,\mathbf{z}) - s^P(K,\mathbf{z})) + c_k - c_K \end{aligned} \tag{16}$$

Now, consider two values of $\mathbf{z}$, say $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$. Differencing (16) across these two values we eliminate the cost parameters, assumed to be independent of $\mathbf{z}$, and get:

$$\begin{aligned} \Lambda^{-1}\big(P(k,\mathbf{z}^{(1)})\big) - \Lambda^{-1}\big(P(k,\mathbf{z}^{(2)})\big) &= b^w \times \big(\Delta_w^P(k,\mathbf{z}^{(1)}) - \Delta_w^P(k,\mathbf{z}^{(2)})\big) \\ &\quad + b^s \times \big(\Delta_s^P(k,\mathbf{z}^{(1)}) - \Delta_s^P(k,\mathbf{z}^{(2)})\big) \end{aligned} \tag{17}$$

or:

$$\Lambda^{-1}\big(P(k,\mathbf{z}^{(1)})\big) - \Lambda^{-1}\big(P(k,\mathbf{z}^{(2)})\big) = b^w \times \Delta_w^P(k,\mathbf{z}^{(1,2)}) + b^s \times \Delta_s^P(k,\mathbf{z}^{(1,2)}) \tag{18}$$

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

$$\begin{bmatrix} \Delta^P_\Lambda(k, \mathbf{z}^{(1,2)}) \\ \Delta^P_\Lambda(k, \mathbf{z}^{(1,3)}) \end{bmatrix} = \begin{bmatrix} \Delta^P_w(k, \mathbf{z}^{(1,2)}) & \Delta^P_s(k, \mathbf{z}^{(1,2)}) \\ \Delta^P_w(k, \mathbf{z}^{(1,3)}) & \Delta^P_s(k, \mathbf{z}^{(1,3)}) \end{bmatrix} \begin{bmatrix} b^w \\ b^s \end{bmatrix} \tag{19}$$

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

$$\frac{\Delta^P_w(k, \mathbf{z}^{(1,2)})}{\Delta^P_w(k, \mathbf{z}^{(1,3)})} \neq \frac{\Delta^P_s(k, \mathbf{z}^{(1,2)})}{\Delta^P_s(k, \mathbf{z}^{(1,3)})} \tag{20}$$

This is the key condition for identification of (b^w, b^s), analogous to the condition in Equation 9 in the simpler case above, except we now need variation in win probability and expected vote share across more values of $\mathbf{z}$ - this is how we identify b^w, b^s separately from the cost difference $c_k - c_K$.

Solving the system in (19) we get expressions analogous to what we had in the simple case with no costs:

$$b^w = \frac{\Delta^P_s(k, \mathbf{z}^{(1,2)})\Delta^P_\Lambda(k, \mathbf{z}^{(1,3)}) - \Delta^P_s(k, \mathbf{z}^{(1,3)})\Delta^P_\Lambda(k, \mathbf{z}^{(1,2)})}{\Delta^P_s(k, \mathbf{z}^{(1,2)})\Delta^P_w(k, \mathbf{z}^{(1,3)}) - \Delta^P_s(k, \mathbf{z}^{(1,3)})\Delta^P_w(k, \mathbf{z}^{(1,2)})} \tag{21}$$

and:

$$b^s = \frac{\Delta^P_w(k, \mathbf{z}^{(1,3)})\Delta^P_\Lambda(k, \mathbf{z}^{(1,2)}) - \Delta^P_w(k, \mathbf{z}^{(1,2)})\Delta^P_\Lambda(k, \mathbf{z}^{(1,3)})}{\Delta^P_s(k, \mathbf{z}^{(1,2)})\Delta^P_w(k, \mathbf{z}^{(1,3)}) - \Delta^P_s(k, \mathbf{z}^{(1,3)})\Delta^P_w(k, \mathbf{z}^{(1,2)})} \tag{22}$$

With b^w and b^s identified we can now obtain an expression for the cost difference $c_k - c_K$ by substituting the identified values into (16) to get:

$$c_k - c_K = \Lambda^{-1}\big(P(k, \mathbf{z})\big) - b^w \times (w^P(k, \mathbf{z}) - w^P(K, \mathbf{z})) - b^s \times (s^P(k, \mathbf{z}) - s^P(K, \mathbf{z})) \tag{23}$$

where b^w, b^s are as defined in (21) and (22), respectively.

4 Candidate wealth and party payoffs

To illustrate the importance of wealth as a determinant of candidate choice, we compute parties' payoff from the candidates chosen in the data. The upper panel of Figure A.4 shows these (normalized) payoffs ordered by constituency and candidate type for the UPA. Lighter colors correspond to higher payoffs. On the lower panel, we present the same information but ordering candidates by wealth, regardless of type. Figure A.5 shows corresponding graphs for the NDA. These graphs reveal that the wealth of the candidate plays an important role in increasing the payoff of political parties.

5 Model fit and validation

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

In Table A.3 we present the analogue of Table 8 but in a model that assumes parties only care about voter preferences (and thus the probability of winning). This is the model estimated in the first column of Table 7.

Table A.3: Model fit with no cost parameters

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

Notes: Number of candidates of each type observed in the data and predicted by the model (average across 100 simulations) using the estimates from column 1 of Table 7. Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively.

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

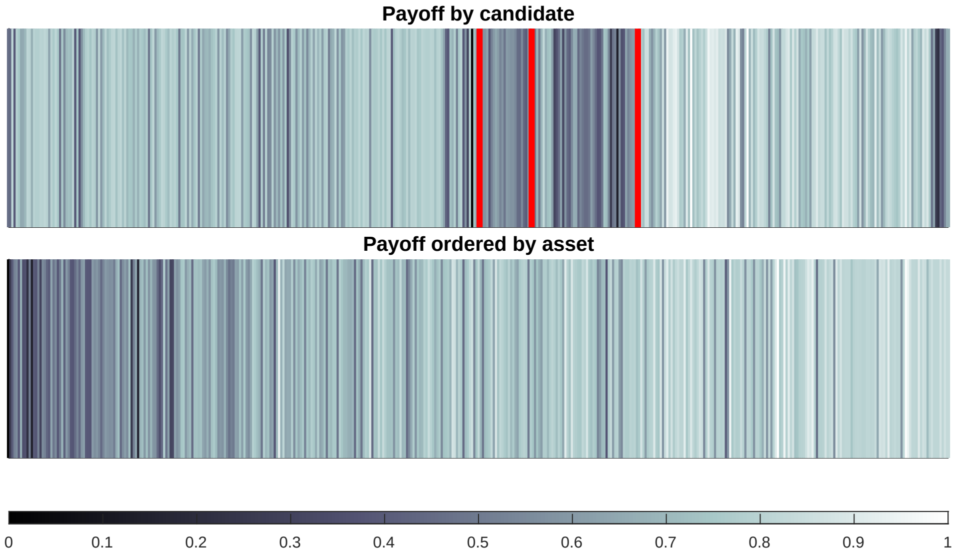

Figure A.4: Candidate wealth and the UPA's payoffs

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

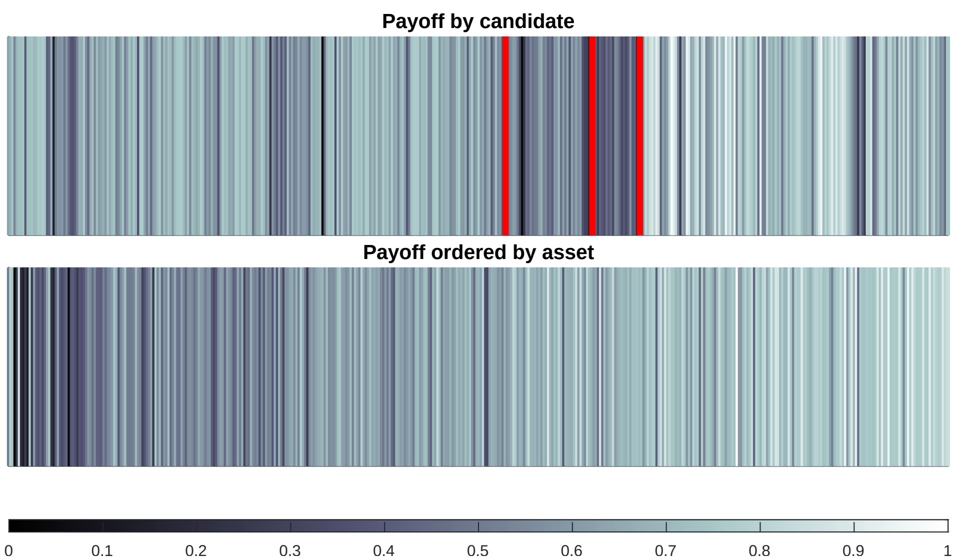

Figure A.5: Candidate wealth and the NDA's payoffs

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

6 Clustering

6.1 Details of k-means clustering

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

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

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

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

where

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

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

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

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

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

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

6.2 Alternative clustering procedure

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

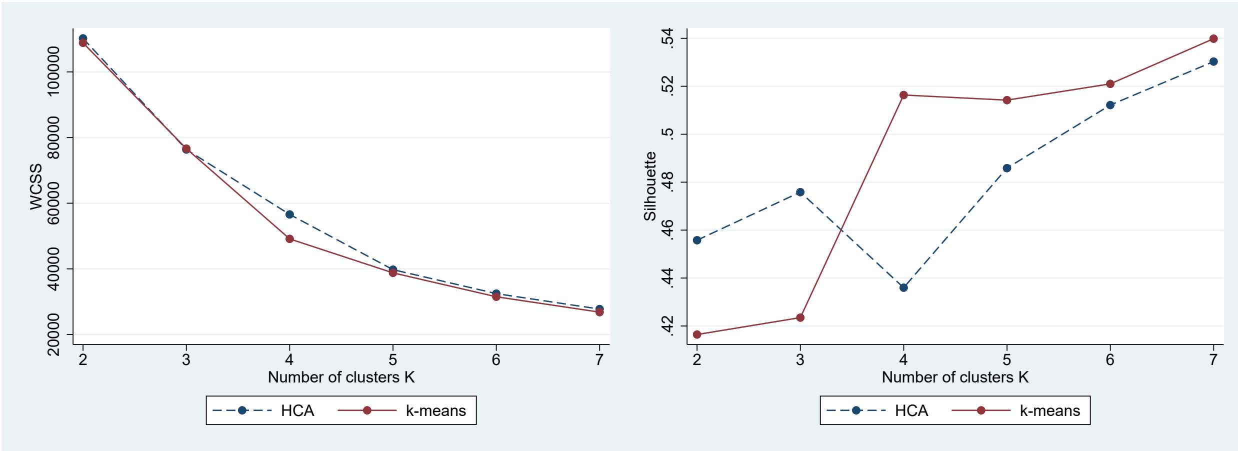

Figure A.6: WCSS and Silhouette: k-means vs HCA

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

First, on the left panel of Figure A.6 we compare the within-cluster sum of squares, WCSS, over a number of possible clusters across the two methods. The k-means approach is uniformly better by this metric (i.e., produces lower WCSS scores).[5]

On the right panel of Figure A.6 we compare the Silhouette score associated with the two approaches across a range of possible numbers of clusters. While HCA performs better at low number of clusters (2 or 3), k-means performs better from 4 clusters on. Moreover, k-means scores with 4 or more clusters dominate any HCA score below 4 clusters. Finally, note that we only ever hit a Silhouette score of 0.5 with k-means and the number of clusters at least 4. In cluster analysis, a Silhouette score of 0.5 is often used as a rule of thumb for"good clustering" (Rousseeuw 1987).

These patterns indicate that k-means is preferable to HCA for our application. We nevertheless computed the types that the HCA algorithm predicts when we set $k = 4$. The types are in Table A.7.

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

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

[5] Note that k-means is based on "centroids" (the average of all points in the cluster), while HCA is not. To generate Figure A.6, we compute the "centroid" of each HCA cluster analogously.

Table A.7: Centroids of candidate types using HCA

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

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

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

7 Criminals as strategic complements

Table A.8: Correlation of estimated choice probabilities

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

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

Table A.8 shows the correlation between the two parties' estimated choice probabilities for the different candidate types. The correlation between the two parties' choice probabilities for criminals is positive, while the correlation between criminals and other types is always negative. This provides suggestive evidence of strategic complementarity between criminal candidates.

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

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

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

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

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

$$\begin{aligned}\frac{\partial P_p(4)}{\partial P_{-p}(4)} &= -\sum_{a=1}^{3}\frac{\exp\left\{\Delta^P(a)\right\}\frac{\partial \Delta^P(a)}{\partial P_{-p}(4)}}{\left(1+\sum_{a=1}^{3}\exp\left\{\Delta^P(a)\right\}\right)^2} \\ &= -\sum_{a=1}^{3} P_p(4)P_p(a)\frac{\partial \Delta^P(a)}{\partial P_{-p}(4)}\end{aligned} \tag{24}$$

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

where

$$\frac{\partial E_P[w_p(a)-w_p(4)]}{\partial P_{-p}(4)} = w_p(a,4) - w_p(4,4) + \sum_{t=1}^{3}\left[w_p(a,t) - w_p(4,t)\right]\frac{\partial P_{-p}(t)}{\partial P_{-p}(4)}, \tag{25}$$

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

Using our parameter estimates, we compute (24) for both parties in every constituency. We do this by assuming that $\frac{\partial P_{-p}(t)}{\partial P_{-p}(4)}$ in (25) is the same for $t = 1,2,3$ (if we instead set $\frac{\partial P_{-p}(t)}{\partial P_{-p}(4)} = -1$ for some t we obtain very similar results). A positive derivative $\frac{\partial P_p(4)}{\partial P_{-p}(4)}$ indicates strategic complementarity between criminal candidates.

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

8 Other robustness checks and counterfactuals

8.1 Robustness

In this section we consider a series of other checks of the robustness of our main estimates in Table 12.

8.1.1 Alternative payoff specifications

First, In Table A.10 we allow for additional non-linearities in party preferences over vote shares (of course, the winning probability itself is a highly non-linear function of vote shares).

As a benchmark, in column 1 we reproduce the estimates from the baseline model. In column 2 we include the expected squared vote share, and in column 3 we add the expected cubic vote share. We find that the linear term is the only one that is always statistically significant. In addition, in column 2 the quadratic term is significant at 10%. Figure A.9 shows the equilibrium choice probabilities implied by each specification. Equilibrium behavior is virtually identical, which further supports our focus on the more parsimonious specification.

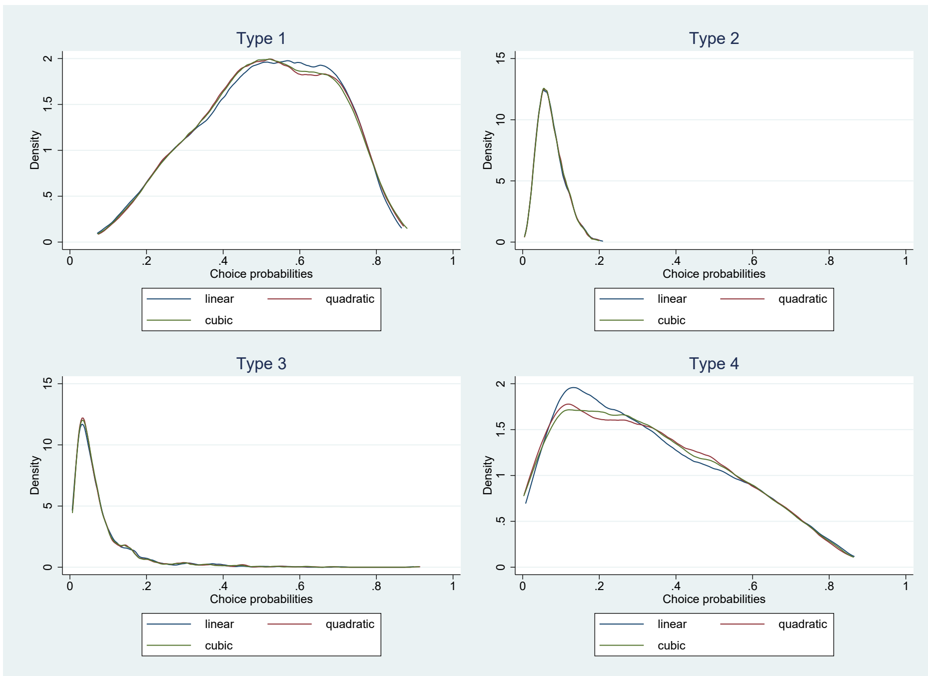

Figure A.9: Distribution of choice probabilities with and without additional non-linearities
Kernel density plots of equilibrium choice probabilities corresponding to the specifications in Table A.10. Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively.

Second, in Table A.11 we replace the expected vote share in parties' objective functions with their expected vote share margin. For winners, this is the difference compared to the runner-up; for losers, it is the difference compared to the winner (which is negative). Qualitatively there is no change in the results. Parties have a preference for winning, but are averse to bigger wins as measured by the vote share margin.

Third, we consider a model where instead of expected share, parties' payoff is affected by the expected number of votes received. This also introduces additional cross-constituency variation depending on the size of the constituency. The estimates are in Table A.12, and they again show no qualitative difference relative to our main results.

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

	(1)	(2)	(3)
b^w	3.24	2.64	2.59
	(1.02)	(1.09)	(1.09)
b^s	-14.15	-9.51	-12.91
	(1.73)	(2.95)	(4.54)
b^s_{sq}	-	- 5.06	5.31
	-	(2.65)	(10.95)
b^s_{cu}	-	-	-10.13
	-	-	(10.53)
$c^0_{2,NDA}$	-1.57	-1.57	-1.57
	(0.27)	(0.27)	(0.27)
$c^0_{3,NDA}$	-2.58	2.49	2.54
	(0.30)	(0.31)	(0.31)
$c^0_{4,NDA}$	0.58	0.43	0.53
	(0.23)	(0.25)	(0.27)
$c^0_{2,UPA}$	-2.09	-2.09	-2.09
	(0.32)	(0.32)	(0.32)
$c^0_{3,UPA}$	-1.93	-1.85	-1.90
	(0.29)	(0.29)	(0.30)
$c^0_{4,UPA}$	0.40	0.27	0.38
	(0.23)	(0.24)	(0.26)
c^{educ}_2	-0.53	-0.53	0.53
	(0.34)	(0.34)	(0.34)
c^{educ}_3	-0.63	-0.62	-0.61
	(0.34)	(0.35)	(0.35)
c^{educ}_4	-0.41	-0.43	-0.42
	(0.23)	(0.22)	(0.22)
c^{crime}_2	-0.03	-0.03	-0.03
	(0.24)	(0.24)	(0.24)
c^{crime}_3	0.36	0.36	0.34
	(0.23)	(0.23)	(0.23)
c^{crime}_4	1.10	1.10	1.10
	(0.15)	(0.15)	(0.15)
c^{asset}_2	0.32	0.32	0.32
	(0.26)	(0.25)	(0.23)
c^{asset}_3	-0.12	-0.11	-0.11
	(0.26)	(0.26)	(0.26)
c^{asset}_4	-0.06	-0.10	-0.09
	(0.16)	(0.17)	(0.17)
c^{Muslim}_2	-0.07	-0.07	-0.08
	(0.41)	(0.41)	(0.41)
c^{Muslim}_3	1.48	1.46	1.46
	(0.29)	(0.29)	(0.29)
c^{Muslim}_4	0.03	0.02	0.03
	(0.24)	(0.24)	(0.24)
Log likelihood	-836.21	-834.16	-833.62

Notes: Estimates of party objective functions allowing for further non-linearities in vote share. Column 1 reproduces column 3 from Table 7. Column 2 adds the expected value of s^2 and column 3 also the expected value of s^3, with corresponding parameters b^s_{sq} and b^s_{cu}. Standard errors in parentheses. Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively. Number of markets: 434.

Table A.11: Party objective function estimates: Expected Margin

	(1)	(2)	(3)
b^w	4.06	2.37	2.24
	(0.74)	(0.85)	(0.90)
b^m	-3.84	-6.11	-9.65
	(0.804)	(0.96)	(1.15)
$c^0_{2,NDA}$			-1.56
			(0.27)
$c^0_{3,NDA}$			-2.60
			(0.30)
$c^0_{4,NDA}$			0.60
			(0.24)
$c^0_{2,UPA}$			-2.08
			(0.32)
$c^0_{3,UPA}$			-1.98
			(0.29)
$c^0_{4,UPA}$			0.47
			(0.23)
c_2^{educ}		-2.00	-0.53
		(0.26)	(0.34)
c_3^{educ}		-2.32	-0.65
		(0.26)	(0.34)
c_4^{educ}		-0.60	-0.38
		(0.20)	(0.22)
c_2^{crime}		-0.47	-0.03
		(0.22)	(0.24)
c_3^{crime}		-0.31	0.37
		(0.21)	(0.22)
c_4^{crime}		1.04	1.08
		(0.14)	(0.15)
c_2^{asset}		-0.74	0.30
		(0.19)	(0.25)
c_3^{asset}		-1.04	-0.04
		(0.20)	(0.26)
c_4^{asset}		0.02	-0.14
		(0.14)	(0.17)
c_2^{Muslim}		0.44	-0.07
		(0.32)	(0.41)
c_3^{Muslim}		1.71	1.45
		(0.27)	(0.28)
c_4^{Muslim}		0.15	0.05
		(0.24)	(0.24)
Log likelihood	-1187.20	-920.22	-835.75

Notes: Estimates of party objective functions in (12) with expected vote share replaced by the expected vote margin (the difference compared to the runner-up for winners, and the difference compared to the winner for losers). Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively. Number of markets: 434.

Table A.12: Party objective function estimates: Expected number of votes

	(1)	(2)	(3)
b^w	3.74	3.29	3.05
	(0.83)	(0.94)	(1.00)
b^t	-0.32	-0.70	-0.98
	(0.09)	(0.11)	(0.12)
$c^0_{2,NDA}$			-1.59
			(0.27)
$c^0_{3,NDA}$			-2.51
			(0.30)
$c^0_{4,NDA}$			0.45
			(0.23)
$c^0_{2,UPA}$			-2.12
			(0.32)
$c^0_{3,UPA}$			-1.84
			(0.29)
$c^0_{4,UPA}$			0.23
			(0.22)
c^{educ}_2		-2.02	-0.53
		(0.26)	(0.34)
c^{educ}_3		-2.31	-0.66
		(0.26)	(0.35)
c^{educ}_4		-0.64	-0.36
		(0.20)	(0.22)
c^{crime}_2		-0.46	-0.03
		(0.22)	(0.24)
c^{crime}_3		-0.31	0.34
		(0.21)	(0.23)
c^{crime}_4		1.07	1.12
		(0.14)	(0.15)
c^{asset}_2		-0.73	0.34
		(0.19)	(0.26)
c^{asset}_3		-1.10	-0.21
		(0.20)	(0.26)
c^{asset}_4		0.16	-0.14
		(0.14)	(0.16)
c^{Muslim}_2		0.44	-0.08
		(0.32)	(0.41)
c^{Muslim}_3		1.75	1.48
		(0.27)	(0.28)
c^{Muslim}_4		0.11	0.03
		(0.24)	(0.24)
Log likelihood	-1187.20	-916.90	-837.36

Notes: Estimates of party objective functions in (12) with expected vote share replaced by the expected number of votes. Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively. Number of markets: 434.

Table A.13: Party objective function estimates: Constituency-level heterogeneity

	Election year 2014	Reservation	Rural population	Unobserved heterogeneity
b^w	2.91	3.24	3.43	3.29
	(1.02)	(1.02)	(1.03)	(1.03)
b^s	-14.43	-13.90	-14.45	-14.00
	(1.73)	(1.74)	(1.75)	(1.75)
c_2^*	-0.76	0.23	0.57	-
	(0.33)	(0.29)	(1.40)	-
c_3^*	0.19	-0.69	-0.92	-
	(0.32)	(0.33)	(1.30)	-
c_4^*	-0.61	-0.37	1.36	-
	(0.21)	(0.20)	(0.87)	-
$c^0_{2,NDA}$	-1.41	-1.66	-2.04	-1.94
	(0.28)	(0.29)	(1.19)	(0.79)
$c^0_{3,NDA}$	-2.67	-2.37	-1.83	-2.08
	(0.32)	(0.31)	(1.11)	(0.95)
$c^0_{4,NDA}$	0.82	0.66	-0.53	1.22
	(0.25)	(0.24)	(0.74)	(0.57)
$c^0_{2,UPA}$	-2.02	-2.20	-2.57	-2.46
	(0.32)	(0.34)	(1.21)	(0.81)
$c^0_{3,UPA}$	-1.99	-1.70	-1.16	-1.40
	(0.29)	(0.30)	(1.12)	(0.57)
$c^0_{4,UPA}$	0.60	0.51	-0.73	1.07
	(0.24)	(0.23)	(0.76)	(0.57)
c_2^{educ}	-0.46	-0.55	-0.56	-0.45
	(0.34)	(0.35)	(0.35)	(0.35)
c_3^{educ}	-0.65	-0.60	-0.60	-0.60
	(0.35)	(0.34)	(0.35)	(0.35)
c_4^{educ}	-0.37	-0.40	-0.46	-0.40
	(0.22)	(0.22)	(0.22)	(0.22)
c_2^{crime}	0.00	-0.02	-0.04	-0.01
	(0.24)	(0.24)	(0.25)	(0.24)
c_3^{crime}	0.35	0.37	0.40	0.37
	(0.23)	(0.23)	(0.24)	(0.23)
c_4^{crime}	1.12	1.08	1.06	1.08
	(0.15)	(0.15)	(0.15)	(0.15)
c_2^{asset}	0.70	0.36	0.36	0.23
	(0.31)	(0.26)	(0.27)	(0.27)
c_3^{asset}	-0.21	-0.19	-0.18	-0.20
	(0.31)	(0.26)	(0.28)	(0.28)
c_4^{asset}	0.24	-0.10	0.02	-0.08
	(0.20)	(0.17)	(0.17)	(0.17)
c_2^{Muslim}	-0.06	-0.04	-0.07	-0.07
	(0.42)	(0.41)	(0.41)	(0.42)
c_3^{Muslim}	1.45	1.39	1.48	1.37
	(0.29)	(0.29)	(0.29)	(0.29)
c_4^{Muslim}	0.00	-0.02	0.05	0.00
	(0.25)	(0.24)	(0.24)	(0.25)
Log likelihood	829.48	-831.80	-834.40	-829.69

Notes: Estimates of party objective functions in (12) allowing for additional heterogeneity. In the first three columns the headings indicate which variable is interacted with candidate types. The parameters on the interactions are denoted c_2^*, c_3^*, c_4^*, respectively. In the fourth column, we include interactions with fixed effects for constituency "types" following Bonhomme, Lamadon and Manresa (2022). Types 1-4 are the Educated, Uneducated, Muslim, and Criminal types, respectively. Number of markets: 434.

8.1.2 Constituency-level heterogeneity

In Table A.13 we study the robustness of our payoff function estimates to various types of observable and unobservable heterogeneity. In the first column we consider the possibility that selection decisions are systematically different across election years (2009 and 2014). The two elections were different - the 2014 election was a breakthrough year for the BJP (and NDA more generally) and marked the end of the INC's near continuous hold on power since the birth of the Indian state. As such, we allow for the possibility that the payoffs to different types of candidates might vary across the two years. We let c_2^*, c_3^*, c_4^* denote the interaction between the type dummies and an indicator for election year 2014. The estimates suggest that parties were less likely to select type 2 and type 4 candidates in 2014 than they were in the 2009 election. However, there is little change in other estimates relative to column 3 of Table 7.

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

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

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

As an alternative we consider the approach of Bonhomme, Lamadon and Manresa (2022), who propose to reduce the dimensionality of the fixed effects by clustering units based on observable characteristics into groups in a first step. Applied to our context, the idea is that constituencies in the same cluster likely have similar *unobservables*. We use kmeans as proposed by Bonhomme, Lamadon and Manresa (2022) to cluster constituencies. The observables we include are the constituency level demographic characteristics in Table 1: Rural population share, Literate population share, ST and SC reservation status, Population with paved roads, Working population and Rural working population. The algorithm identifies four constituency types, and we include dummies for these types (interacted with candidate types) and re-estimate the model in Table A.13.

Controlling for constituency level unobserved heterogeneity yields virtually identical b^w and b^s estimates to our main specification, and causes little change in the cost parameters.

8.2 Additional counterfactual results

In the paper we find that $b^s < 0$, so that parties receive a disutility from candidates with higher expected vote shares (for given winning probability). This result is qualitatively robust across the many specifications we consider. In Table A.14 we present the results of a counterfactual experiment where we set $b^s = 0$ to get a sense of the implications of this finding for different types.

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

Next, we repeat our counterfactual exercise of a criminal ban combined with $b^s = 0$. The resulting choice probabilities are summarized in the last column of Table A.14, and the third column shows the counterfactual results from the paper for comparison. As can be seen, under a ban, parties' behavior is much less dependent on party organization as reflected by the value of b^s.

Table A.14: Different types' probability of being chosen when $b^s = 0$, with and without a criminal ban

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

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

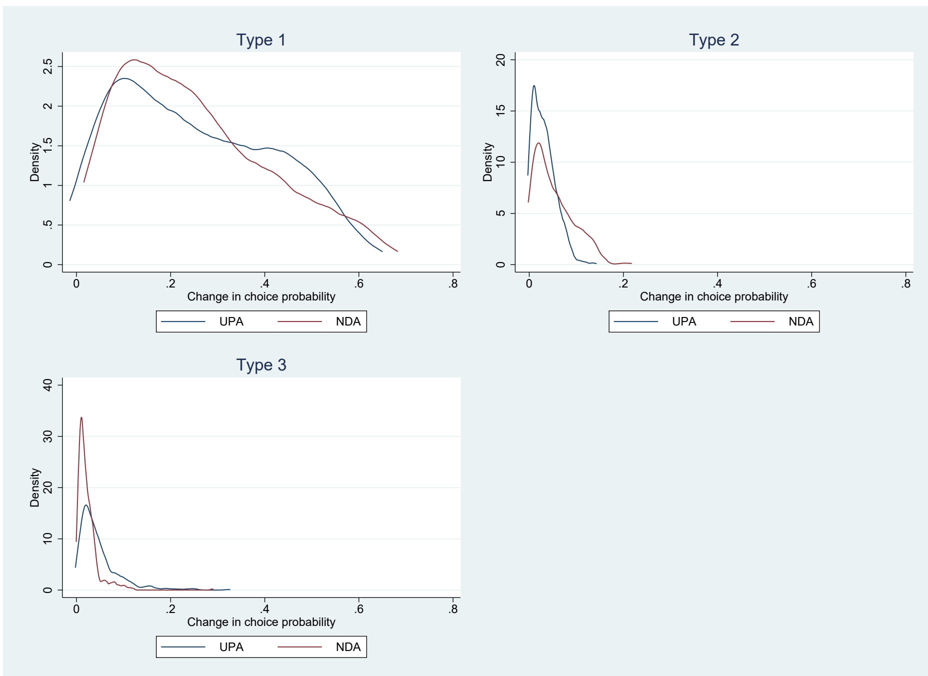

Figure A.15: Distribution of changes in types' choice probabilities after criminal ban

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

References

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

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

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