

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

October 1, 2023

Contents

1	Additional tables	2
2	Identification of Model Parameters	3
3	Model fit and validation	6

1 Additional tables

Table A.1: Distribution of candidate types in national elections by year

	Type 1	Type 2	Type 3	Type 4	Total	N
2009	46.12	22.05	12.54	19.3	100	3207
2014	44.24	20.06	12.43	23.26	100	3370
Total	45.16	21.03	12.48	21.33	100	6577

Notes: Type 1: educated, Type 2: uneducated, Type 3: Muslim, Type 4: criminal

Table A.2: Distribution of candidate types in national elections by state

	Type 1	Type 2	Type 3	Type 4	Total	N
Andhra Pradesh	50.92	20.94	10.68	17.45	100	487
Assam	55.73	13.02	23.96	7.29	100	192
Bihar	38.17	18.17	13.17	30.49	100	820
Gujarat	40.67	22.33	13.33	23.67	100	300
Haryana	52.34	23.83	3.4	20.43	100	235
Jharkhand	42.29	25.69	9.88	22.13	100	253
Karnataka	48.31	19.85	14.23	17.6	100	534
Kerala	25.66	14.16	38.05	22.12	100	113
Madhya Pradesh	50.34	18.37	9.52	21.77	100	147
Maharashtra	39.6	20.47	15.64	24.3	100	889
Odisha	52.23	16.15	7.22	24.4	100	291
Rajasthan	48.22	27.41	10.15	14.21	100	394
Tamil Nadu	50.39	21.34	7.59	20.68	100	909
Uttar Pradesh	43.76	23.79	11.16	21.29	100	681
West Bengal	41.87	23.19	19.28	15.66	100	332
Total	45.16	21.03	12.48	21.33	100	6577

Notes: Type 1: “educated,” Type 2: “uneducated,” Type 3: “Muslim,” Type 4: “criminal”

Table A.3: Correlation of estimated CPs

		UPA			
		Type 1	Type 2	Type 3	Type 4
NDA	Type 1	0.27	0.14	0.06	-0.29
	Type 2	0.13	0.25	-0.12	-0.08
	Type 3	0.06	-0.08	0.21	-0.16
	Type 4	-0.27	-0.16	-0.07	0.29
<i>Notes:</i>		Type 1: “educated,”	Type 2: “uneducated,”	Type 3:	
		“Muslim,”	Type 4: “criminal”		

2 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 and win probabilities are known to the researcher - these are estimated in a first stage. Let the probability that party $i \in \{1, 2\}$ chooses action $a_i = k$ for $k = 1, \dots, K$ given observable payoff variables \mathbf{z} (i.e., constituency characteristics) be given by $P_i(k, \mathbf{z})$, and write expected winning probability of party i as:

$$w_i^P(k, \mathbf{z}) = E_i[w_i(a_i, a_{-i}, \mathbf{z}) | a_i = k] \quad (1)$$

where the expectation $E_i[w_i(a_i, a_{-i}, \mathbf{z}) | a_i = k]$ is an integration over a_{-i} using player $-i$'s choice probability (see Section 3).

We establish identification in the baseline model with type specific benefit parameters $\mathbf{b} = (b_1, b_2, \dots, b_K)'$ and type specific costs $\mathbf{c} = (c_1, c_2, \dots, c_K)'$ as most of the intuition can be gleaned from this case, and allowing for additional cost parameters as in our full model does not substantially change the identification argument.

Player i 's choice probability satisfies:

$$P_i(k, \mathbf{z}) = \Lambda(b_k \times w_i^P(k, \mathbf{z}) + c_k) \quad (2)$$

where, given our assumption about the error distribution:

$$\Lambda(b_k \times w_i^P(k, \mathbf{z}) + c_k) = \frac{\exp \{b_k \times w_i^P(k, \mathbf{z}) + c_k\}}{\sum_{k'} \exp \{b_k \times w_i^P(k', \mathbf{z}) + c'_k\}} \quad (3)$$

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

Inverting the choice probability gives:

$$\begin{aligned}\Lambda^{-1}(P(k, \mathbf{z})) &= \ln(P(k, \mathbf{z})) - \ln(P(K, \mathbf{z})) \\ &= b_k \times w^P(k, \mathbf{z}) - b_K \times w^P(K, \mathbf{z}) + c_k - c_K\end{aligned}\tag{4}$$

where we have taken type K as the reference type.

Before discussing full identification of the vectors \mathbf{b} and \mathbf{c} , to build intuition let's first consider the case where the preference for winning is common across candidate types: $b_k = b$ for all k . Then we have:

$$\Lambda^{-1}(P(k, \mathbf{z})) = b \times (w^P(k, \mathbf{z}) - w^P(K, \mathbf{z})) + c_k - c_K\tag{5}$$

Define $\Delta_w^P(k, \mathbf{z}) \equiv w^P(k, \mathbf{z}) - w^P(K, \mathbf{z})$. 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* .

Now, consider two values of \mathbf{z} , say $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$. Differencing (5) across these two values:

$$\Lambda^{-1}(P(k, \mathbf{z}^{(1)})) - \Lambda^{-1}(P(k, \mathbf{z}^{(2)})) = b \times (\Delta_w^P(k, \mathbf{z}^{(1)}) - \Delta_w^P(k, \mathbf{z}^{(2)}))\tag{6}$$

or rearranging:

$$b = \frac{\Lambda^{-1}(P(k, \mathbf{z}^{(1)})) - \Lambda^{-1}(P(k, \mathbf{z}^{(2)}))}{\Delta_w^P(k, \mathbf{z}^{(1)}) - \Delta_w^P(k, \mathbf{z}^{(2)})}\tag{7}$$

From (7) it is clear that if the sign of the numerator and denominator are different, b is negative, otherwise b is positive. When are the signs different? Suppose that $\Delta_w^P(k, \mathbf{z}^{(1)}) - \Delta_w^P(k, \mathbf{z}^{(2)}) > 0$, so that in constituencies with characteristics $\mathbf{z}^{(1)}$ type k is relatively more likely to win than in constituencies with characteristics $\mathbf{z}^{(2)}$, and that $\Lambda^{-1}(P(k, \mathbf{z}^{(1)})) < \Lambda^{-1}(P(k, \mathbf{z}^{(2)}))$. Since $\Lambda^{-1}(\cdot)$ is increasing, this implies that

$$P(k, \mathbf{z}^{(1)}) < P(k, \mathbf{z}^{(2)})$$

or in words, that the party is less likely to select type k in constituencies with characteristics $\mathbf{z}^{(1)}$ than in constituencies with characteristics $\mathbf{z}^{(2)}$. So the parameter b is negative if the party tends to not select the candidate type that is relatively likely to win given the constituency characteristics \mathbf{z} .

With the parameter b identified, cost *differences* $c_k - c_K$ are identified as:

$$c_k - c_K = \Lambda^{-1}(P(k, \mathbf{z})) - b \times (w^P(k, \mathbf{z}) - w^P(K, \mathbf{z}))\tag{8}$$

and clearly, the difference $c_k - c_K$ is increasing in the choice probability $P(k, \mathbf{z})$, all else constant.

With this simpler case established, we now move to the case of type specific parameters b_k . Differencing Equation (4) across two values of \mathbf{z} gives:

$$\Lambda^{-1}(P(k, \mathbf{z}^{(1)})) - \Lambda^{-1}(P(k, \mathbf{z}^{(2)})) = b_k \times (w^P(k, \mathbf{z}^{(1)}) - w^P(k, \mathbf{z}^{(2)})) - b_K \times (w^P(K, \mathbf{z}^{(1)}) - w^P(K, \mathbf{z}^{(2)}))$$

Now define:

$$\Delta_w^P(k, \mathbf{z}^{(1,2)}) \equiv w^P(k, \mathbf{z}^{(1)}) - w^P(k, \mathbf{z}^{(2)}), \quad k = 1, 2, \dots, K \quad (9)$$

$$\Delta_\Lambda(k, \mathbf{z}^{(1,2)}) \equiv \Lambda^{-1}(P(k, \mathbf{z}^{(1)})) - \Lambda^{-1}(P(k, \mathbf{z}^{(2)})) \quad (10)$$

We can then re-write the difference in inverted choice probabilities as:

$$\Delta_\Lambda(k, \mathbf{z}^{(1,2)}) = b_k \times \Delta_w^P(k, \mathbf{z}^{(1,2)}) - b_K \times \Delta_w^P(K, \mathbf{z}^{(1,2)}) \quad (11)$$

and isolating for the reference parameter b_K we get:

$$b_K = \frac{b_k \times \Delta_w^P(k, \mathbf{z}^{(1,2)}) - \Delta_\Lambda(k, \mathbf{z}^{(1,2)})}{\Delta_w^P(K, \mathbf{z}^{(1,2)})} \quad (12)$$

This holds at any pair of \mathbf{z} vectors, so we can also write:

$$b_K = \frac{b_k \times \Delta_w^P(k, \mathbf{z}^{(2,3)}) - \Delta_\Lambda(k, \mathbf{z}^{(2,3)})}{\Delta_w^P(K, \mathbf{z}^{(2,3)})} \quad (13)$$

and thus solve for the parameter b_k :

$$b_k = \frac{\Delta_w^P(K, \mathbf{z}^{(1,2)})\Delta_\Lambda(k, \mathbf{z}^{(2,3)}) - \Delta_w^P(K, \mathbf{z}^{(2,3)})\Delta_\Lambda(k, \mathbf{z}^{(1,2)})}{\Delta_w^P(K, \mathbf{z}^{(1,2)})\Delta_w^P(k, \mathbf{z}^{(2,3)}) - \Delta_w^P(K, \mathbf{z}^{(2,3)})\Delta_w^P(k, \mathbf{z}^{(1,2)})} \quad k = 1, \dots, K-1 \quad (14)$$

Again, the parameter b_k is negative when the numerator and denominator have the opposite sign. When do they have the opposite sign? Suppose that $\Delta_w^P(K, \mathbf{z}^{(1,2)}) \simeq \Delta_w^P(K, \mathbf{z}^{(2,3)})$

so that Equation 14 reduces to

$$b_k = \frac{\Delta_\Lambda(k, \mathbf{z}^{(2,3)}) - \Delta_\Lambda(k, \mathbf{z}^{(1,2)})}{\Delta_w^P(k, \mathbf{z}^{(2,3)}) - \Delta_w^P(k, \mathbf{z}^{(1,2)})} \quad (15)$$

In this case, $b_k < 0$ if $\Delta_\Lambda(k, \mathbf{z}^{(2,3)}) < \Delta_\Lambda(k, \mathbf{z}^{(1,2)})$ and $\Delta_w^P(k, \mathbf{z}^{(2,3)}) > \Delta_w^P(k, \mathbf{z}^{(1,2)})$. Intuitively, this roughly can be interpreted to mean that the probability the party selects type k increases less moving from constituency $\mathbf{z}^{(3)}$ to constituency $\mathbf{z}^{(2)}$ than it does moving from constituency $\mathbf{z}^{(2)}$ to constituency $\mathbf{z}^{(1)}$ even though the probability of winning increases *more* moving from constituency $\mathbf{z}^{(3)}$ to constituency $\mathbf{z}^{(2)}$ than it does moving from constituency $\mathbf{z}^{(2)}$ to constituency $\mathbf{z}^{(1)}$.

The parameter on the reference type b_K is also identified by substituting Equation 14 into Equation 12, and cost differences are identified as:

$$c_k - c_K = \Lambda^{-1}(P(k, \mathbf{z})) - b_k \times w^P(k, \mathbf{z}) + b_K \times w^P(K, \mathbf{z}) \quad (16)$$

Note the following interesting features of the identification argument:

1. Variation in \mathbf{z} is crucial for identifying b separately from c_k , and if we allow for type specific b we require more independent values of \mathbf{z} .
2. All type specific values of b are identified, but costs are only identified up to a reference type.

3 Model fit and validation

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

In Table A.4 we present the analogue of Table 10 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 the results Table 9.

Table A.4: Model fit with no cost parameters

Type	UPA actual	UPA predicted	NDA actual	NDA predicted	All actual	All predicted
1	217	130.49	229	147.47	446	277.96
2	24	95.66	43	87.89	67	183.55
3	49	97.68	22	92	71	189.68
4	144	110.17	140	106.64	284	216.81

When parties are restricted to care only about voter preference the model significantly 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).