How Indian Voters Respond to Candidates with Criminal Charges
Evidence from the 2009 Lok Sabha Elections

BHASKAR DUTTA, POONAM GUPTA

This paper uses data from the 2009 Lok Sabha elections to examine the response of voters to candidates who have reported that they have criminal charges against them. Our empirical results show that voters do penalise candidates with criminal charges, but the magnitude of the penalty decreases if there are other candidates in the constituency with criminal charges. The vote shares are positively related to candidate wealth, with the marginal effect being higher for the candidates with criminal charges. Moreover, candidates with criminal charges also have greater wealth.

Bhaskar Dutta (*B.Dutta@warwick.ac.uk)* is with the University of Warwick, the United Kingdom, and Poonam Gupta (*pgupta.nipfp@gmail.com*) is with the National Institute of Public Finance and Policy, Delhi.

1 Introduction

It is now well known that the nexus between Indian politicians and criminals has assumed alarming proportions. Roughly a fourth of the members of the current Lok Sabha face pending criminal charges.[1] A similar situation prevails in the various state assemblies. Many of the members of Parliament or state assemblies have been indicted with serious charges, including murder and rape. Not surprisingly, this has attracted increasing attention in both the media as well as in academic research.

The only legal measure designed to prevent the influx of criminals into Parliament and the state assemblies is the Representation of the People Act, 1951. This Act specifies that candidates will be barred from contesting an election on conviction by a court of Law. The period of disqualification is for six years from the date of conviction, or from the date of release from prison, depending on the severity of the charge. Unfortunately, this law hardly has any bite because of the well-known infirmities in the Indian judicial system. Even when cases are registered, inordinate judicial delay implies that these cases drag on, seemingly indefinitely. And there is also the possibility that governments typically drag their feet when it comes to prosecuting "local elites".

This is why the Election Commission had proposed in 2004 that the Representation of the People Act should be amended to disqualify candidates accused of offences which carry sentences of five years or more as soon *as* a court deems that charges can be framed against the person. However, the Lok Sabha itself would have to pass appropriate legislation to implement the Election Commission's suggestion. Obviously, such legislation is against the interests of a large number of politicians, and so it is not surprising that the Election Commission's proposal has not been implemented.

A landmark judgment of the Supreme Court in 2002 required every candidate contesting state and national elections to submit a legal affidavit disclosing his or her educational qualifications, as well as information about personal wealth, and most importantly their criminal record. The Court also stipulated that wide publicity should be given to the contents of the affidavits so that the electorate can take informed decisions about who to elect to the assemblies and Parliament. Unfortunately, the Supreme Court's order does not seem to have had much impact insofar as the influx of legislators with criminal indictment is concerned.[2]

The continuing entry of large numbers of candidates with criminal records into Indian legislatures raises at least a couple of intriguing questions. First, why do parties nominate such candidates? Given the huge demand for party tickets, the nomination of candidates with criminal records suggests that such candidates must possess some electoral advantage. We discuss some hypotheses which have been suggested to explain this electoral advantage. Second, what is the response of voters to candidates who have reported that they have criminal charges against them?

While the first issue has been discussed in the literature, the second issue has not been scrutinised rigorously. A somewhat cursory look at the data by simply looking at the ratio of winning candidates to the number of contesting candidates amongst the criminal and non-criminal groups suggests that criminal candidates have a higher probability of winning. Perhaps, this has given rise to the feeling that criminals have an electoral advantage. The following from Aidt et al (2011) is representative of the prevailing view: "Criminals, we show, boast an extraordinary electoral advantage in India".

We analyse this phenomenon of electoral advantage of criminals in India using the data on candidates for the 2009 general elections. We use a simple analytical model used in Dutta and Gupta (2012). This model assumes that criminal charges do give rise to some stigma amongst the electorate. This stigma has a negative effect on vote shares since voters are less likely to vote for candidates who have criminal charges levied against them. However, the negative effect of this stigma on a candidate's vote share is lower if there are other candidates in the constituency with criminal charges. Campaigning, the cost of which is borne from candidates' wealth, helps a candidate to increase his or her expected vote share by winning over the "marginal" voter. A criminal candidate gets an additional benefit since he can use the campaigning to convince voters of his innocence, and so reduce the negative effects of the stigma associated with criminal charges. This is plausible since the candidates have not been convicted, but only charged with some criminal offence. We look at a Nash equilibrium of a game in which the only strategic variable is the amount of campaign expenditure. Our regression model described later is derived from the predictions of the theoretical model.

A principal finding is that voters do penalise candidates with criminal charges. That is, all else being equal, the vote share of a candidate with criminal charges is lower than that of ones who do not have any such blemish. Notice that the negative effect of criminal charges on vote shares seems to contradict the prevalent view that candidates with criminal charges – or tainted candidates as we will henceforth call them – have an electoral advantage. However, these tainted candidates are able to overcome this electoral disadvantage because they have greater wealth, and wealth plays a significant role in increasing vote shares. The most plausible channel through which wealth affects vote shares is of course through campaign expenditures, which are likely to be positively related to wealth. The negative effect is also reduced if there are other candidates in the constituency with criminal charges.

Since voters penalise candidates with criminal charges, why do political parties still nominate them when so many candidates without criminal charges fight to get their party's nomination? A plausible explanation starts from the premise that candidates facing the threat of criminal convictions are more keen to contest the elections. Their enthusiasm is easily explained. Apart from the usual benefits which accrue to all successful candidates, candidates with criminal indictments look forward to an additional benefit. In particular, successful candidates (particularly those belonging to parties in the government) can with high probability either use coercion or influence to ensure that the local administration does not pursue the case(s) against them with any vigour.

Moreover, the data suggests that criminal candidates are significantly wealthier than those without criminal charges.[3] Also, they are perhaps willing to contribute a higher fraction of their wealth to the party, or they ask for less resources from the party. This simply reflects the higher price or value that they place on a party ticket. So, criminal candidates generate positive externalities to candidates of their own party since their additional contributions release party funds which can be used in other constituencies. This is a plausible explanation of why parties may nominate candidates with criminal backgrounds even if they are (partially) penalised at the polls.

Several recent papers offer explanations of why parties choose candidates with a dubious background. Banerjee and Pande (2009) start with the observation that voters may have a preference for candidates belonging to their own ethnic group. This implies that a politician belonging to the ethnically dominant group in a constituency may win even if he is of lower *quality*. Banerjee and Pande (2009) assume that parties do want to select candidates of the best quality. However, the quality of candidates available to a party in any constituency is a random variable. They show that an increase in the relative size of the ethnically dominant group or an increase in voters' preferences for candidates belonging to their own group can worsen the quality of the winning candidate. Banerjee and Pande test the predictions of their model by using panel data on politician quality in 102 jurisdictions in Uttar Pradesh.[4]

Of course, the Banerjee-Pande hypothesis does not explain why so many candidates with a criminal background contest elections. But, it does provide at least a partial explanation of why there is an increasing number of *successful* legislators in state assemblies as well as the Lok Sabha with a criminal background.

Vaishnav (2012) studies elections to 28 state assemblies between 2003 and 2009. He finds that personal wealth of candidates is positively associated with criminal status where a candidate is defined to be a criminal if he has been charged with a "serious" crime. The basic result is subjected to a variety of robustness checks. This leads him to offer the same explanation that we have mentioned earlier – parties nominate criminal candidates simply because they contribute larger sums to the party coffers.

Aidt et al (2011) develop an interesting theoretical model where they assume that criminal candidates have some electoral advantage, although parties also incur some reputational cost

in nominating them. They "are agnostic about the sources of this advantage", but speculate that the electoral advantage of criminals could arise because they can intimidate prospective voters of rival parties into staying away from the polls. Notice that this would imply voting turnout should be negatively correlated with number of criminals in a constituency. We show that this is not true in the 2009 Lok Sabha elections.

So, parties face a trade-off between the reputational cost of nominating candidates with criminal charges and their electoral advantage. This trade-off implies that parties would be more willing to incur the reputational cost in constituencies which are likely to witness close contests since the electoral advantage is more attractive in these constituencies. Conversely, a party would be unlikely to field a tainted candidate in a constituency where the party is very likely to win. Similarly, candidates with criminal indictments are more likely to be fielded in constituencies where the cost is lower – for instance, in constituencies where voters are poorly informed about the characteristics of the contesting candidates.[5]

The rest of the paper is organised as follows. The econometric specification and the details on the data and the different data sources used in the paper are described in Section 2. Results from the empirical exercise are discussed in Section 3, and the last section is a concluding section.

2 Data and Econometric Specification

We now describe the data and the econometric specification used in the empirical exercise.

Data

In 2002, the Supreme Court in India decreed that all candidates contesting an election for the Lok Sabha, Rajya Sabha, or state assemblies had to file an affidavit with the Election Commission containing information on their assets (and liabilities), criminal charges and education. We derive the data on these variables directly from the affidavits of the candidates – these are available on the Election Commission's website as well as from a website maintained by the Association for Democratic Reforms (ADR), http://myneta.info.

The data on percentage of votes obtained, age, and gender of the candidates are obtained from the Election Commission's website. Information on candidate incumbency has been gathered using various sources including searching through reports in the newspapers or on various internet sites. We define a party as an *incumbent* in a state if it was in power in the state (or was a major coalition partner), from 2008 up to the elections in 2009. The state-level incumbency information has been put together using the information contained in various articles in the *Economic & Political Weekly* and elsewhere. Appendix A1 (p 51) provides the data sources from where the data on various variables have been obtained, while Appendix A2 (p 51) provides the summary statistics of the variables.

India has 28 states and seven union territories (UTs) in all. Among the UTs, only Delhi has its proper local administration with its own chief minister, while the remaining UTs are administered by the centre. Therefore, we include Delhi as a "state" in our sample while excluding the remaining six UTs from the analysis. We follow Gupta and Panagariya (2012) and exclude the eight north-eastern states since they have a special status with deep involvement of the centre in their development process, as well as the state of Jammu and Kashmir. This leaves us with a total of 20 states including Delhi. These states account for 506 out of the total of 543 parliamentary seats across the country.

Using the data from the affidavits, we define three categories for the education status of the candidates: education up to high school, up to undergraduate level, or with a postgraduate or technical degree, and define different dummies for each one of them. Relative wealth is calculated as the ratio of the wealth of the candidate to the average wealth of the rest of the candidates in the constituency. In the regressions where we exclude all independent candidates, relative wealth of candidate i is defined as the ratio of the candidate's wealth to the average wealth of the other non-independent candidates in the constituency.

Each candidate's affidavit has to contain information on whether the candidate faced any criminal charges, as well as the sections of the Indian Penal Code (IPC) under which the charges if any have been framed. In addition, the candidate has to declare whether he or she has ever been convicted. Thus, in principle, data are available on the number of criminal cases that a candidate faces, the specific sections of the IPC under which the candidate faces these charges and whether the candidate has ever been convicted. The ADR further divides the charges into the charges for serious and non-serious offences, by examining the sections of the IPC under which the candidates face the charges. The conviction rate of candidates facing charges is very low, out of the 1,155 candidates in the 2009 Lok Sabha elections who faced at least one criminal charge, only 15 candidates were convicted.

It is sometimes claimed that the data on criminal charges is misleading since the charges might be initiated by political rivals. Moreover, some of the charges are associated with involvement in political activities. In order to clean the data of such "spurious" charges, we specify a value of one to the criminal dummy only when a candidate faces more than one charge. This adjustment takes care of some obvious cases of frivolous charges or charges arising out of political activities.[6] Henceforth, we will use the term "tainted candidate" to denote a candidate who has two or more criminal charges against them.

Consider now the patterns of criminal charges across candidates, states, and parties, and their correlates with other candidate specific factors for the 20 states that are included in our regression analysis. Table 1 shows that it is the national

Table 1: Candidates with Criminal Cases across Party Types

Party Type	Number of Candidates	Number of Candidates with At Least Two Criminal Cases	% of Candidates with At Least Two Criminal Cases
	I	II	III: (II/1)*100
National parties	1,353	176	11.5
State parties	585	108	15.6
Unrecognised parties	1,790	110	6.2
Independent candidates	3,659	124	3.4

Source: Authors' own calculations using the data mentioned in Appendix A1; data refer to the observations on 20 states included in the regressions.

and recognised state parties which field a substantially higher proportion of tainted candidates. In fact, roughly one in seven candidates fielded by state parties have at least two criminal charges levied against them. The corresponding number for national parties is over one in 10 candidates. This, together with the fact that a substantially higher number of *winning* candidates come from the national and state parties, is a partial explanation of why the win-ratio (the ratio of the number of successful candidates to the number of contesting candidates) is substantially higher for tainted candidates. This is documented in Table 2.

Table 3 shows the distribution of constituencies by the number of candidates who faced at least two charges. On average, about 15 candidates contested the election in each constituency in the 2009 Lok Sabha elections. Despite the large number of candidates, an overwhelming number of constituencies – over 75% – had no tainted candidates. In other words, there was a concentration of tainted candidates in some constituencies. In fact, states like Bihar, Jharkhand and Kerala had a concentration of tainted candidates.

Table 4 shows that on average, tainted candidates were wealthier, more likely to be incumbents and obtained a much larger per cent of the votes. Somewhat surprisingly, the average age and education level of tainted candidates is also higher. Indeed, the differences in averages of these variables for tainted and other candidates are statistically significant at the 1% level.

Econometric Specification

Our main interest is in examining whether the higher win ratio of tainted candidates can be reconciled with the assumption that voters do punish these candidates. The dependent variable in all our regressions is the vote share of each candidate i. Since this takes value between zero and one, we transform the variable by calculating the log odds ratio for vote share of each candidate and estimate the model by ordinary least squares, with heteroskedasticity corrected standard errors. The dependent variable Y_i is thus calculated as $\log\left(\frac{\text{Vote share}_i}{1-\text{vote share}_i}\right)$.

Another constraint that the data imposes is that the vote shares of all candidates add up to one within each constituency. Therefore, in our benchmark regressions, we estimate the regressions either by dropping all the candidates of a large party,[7] or all the candidates belonging to a large coalition such as the United Progressive Alliance (UPA) or National Democratic Alliance (NDA), one at a time. Since the vote shares of these large parties and coalitions are significant (Table 5), the adding-up constraint does not apply any longer.

We first start with a parsimonious specification in which the only explanatory variables are the criminal dummy and *relative wealth*. The reason why relative (and not absolute) wealth is used is straightforward. While the wealth of candidate i himself should have a positive impact, the wealth of other candidates should have a non-positive effect since vote shares add up to one. It therefore makes sense to use the relative wealth of candidate i as an explanatory variable. We also include other candidate characteristics such as the level of education of the candidate, dummy for the incumbent candidates seeking re-election to Lok Sabha, and a dummy for the candidates contesting as members of the state incumbent party(ies) in the regression equation.

The variables of particular interest are the criminal dummy and relative wealth. Tables 6 and 7 (p 47) report the results of the regression exercises. The difference between these two tables is that the latter excludes the set of all independent candidates from the sample. However, there is no difference in the qualitative results. It turns out that the coefficients of both these variables are *positive* and statistically significant. Of course, the fact that the coefficient on the criminal dummy variable is positive seems to corroborate the view that tainted candidates have an electoral advantage. Certainly, it does not suggest that voters attach any stigma to tainted candidates.

Table 2: Distribution of Contesting and Winning Candidates by the Number of Criminal Cases

Number of Criminal Cases	Number of Candidates	Number of Winning Candidates
I	II	III
0	6,551	349
1	607	73
2-4	382	57
5-9	92	16
>10	44	10
Total	7,676	506

Source: Authors' own calculations using the data mentioned in Appendix A1; data refer to the observations on 20 states included in the regressions.

Table 3: Distribution of Candidates with Charges across Constituencies

Number of Candidates with At Least Two Charges	Number of Constituencies
0	206
1	169
2	83
3	25
4	14
5	9

Source: Authors' own calculations using the data mentioned in Appendix A1, data refers to the observations on 20 states included in the regressions.

Table 5: Number of Seats and Vote Shares of Parties

	Parties	Number of Seats	Average Vote Share
National Democratic Alliance	Asom Gana Parishad	6	33.63
	Bharatiya Janata Party	433	25.00
	Indian National Lok Dal	5	28.99
	Janata Dal (United)	55	18.53
	Rashtriya Lok Dal	7	37.66
	Shiromani Akali Dal	10	43.18
	Shiv Sena	47	18.08
	Telangana Rashtra Samithi	9	29.23
United Progressive Alliance	All India Majlis-e-Ittehadul Muslimeen	1	42.14
	All-India Trinamool Congress	35	34.13
	Dravida Munnetra Kazhagam	22	44.89
	Indian National Congress	440	35.06
	Jammu and Kashmir National Conference	3	48.18
	Jharkhand Mukti Morcha	42	5.62
	Kerala Congress(M)	1	50.13
	Muslim League Kerala State Committee	17	9.30
	Nationalist Congress Party	68	16.99
	Republican Party of India (Athawale)	53	0.40
	Viduthalai Chiruthaigal Katchi	3	29.28
Bahujan Samaj Party	Bahujan Samaj Party	500	7.18

Table 4: A Comparison of Variables for Candidates With and Without Criminal Charges (at least two criminal charges)

Criminal Dummy	% Votes	Age	Log Assets (in 1000s)	Education Index	Incumbent (%)
0	5.9	45.7	13.7	2.57	4
1	15.4***	47.2***	15.1***	2.71***	10***
Total	6.59	45.8	13.81	2.58	5

*** indicates that the values are significantly different from those for candidates with one or no charges at 1% level of significance.
Source: Authors' own calculations using the data mentioned in Appendix A1; data refer to the observations on 20 states included in the regressions.

Next, we introduce a new variable which is the interaction of the criminal dummy with the number of *other* tainted candidates in the constituency. The presence of other tainted candidates in the constituency increases the vote share of a tainted candidate – a finding consistent with the theoretical model. The coefficient on the criminal dummy continues to be positive but is no longer statistically significant. So, these regressions do not suggest that voters punish tainted candidates.

However, the data does suggest that tainted candidates are significantly wealthier than the rest. The possibility that wealth may play a differential role for tainted candidates remains open. Since the earlier parsimonious regressions do not incorporate these factors explicitly, we enrich the specification by introducing a new variable which is the interaction of wealth with the criminal dummy.

So, our benchmark regression equation is:
$Y_i = \alpha + \beta_c\ Criminaldummy_i + \beta_{cw}Wealth_i*$
$Criminaldummy_i + \gamma_s\ Number\ of\ Candidates$
$with\ Charges_i*Criminaldummy_i + \beta_w\ Relative$
$Wealth_i + \beta_n\ Incumbency_i + \beta_{ns}State\ Incumbent_i + \beta_e Dummies\ for\ high\ Education\ Status_i$
$+ \gamma Constituency\ Fixed\ Effects + \lambda\ Party$
$Fixed\ Effects + \varepsilon_i$ (11)

In this formulation, β_c measures the response of voters to candidates with a criminal charge. Note that β_{cw} is not quite the marginal effect of wealth on vote shares of the tainted candidates since we use relative wealth as the "uninteracted" variable. It measures the differential impact of wealth on candidates with criminal charges.

In all our regressions we include constituency fixed effects and party fixed effects to control for omitted variables, such as the varying policy platforms of the candidates belonging to different political parties. We also conduct various robustness tests which are reported in detail in the next section.

Our empirical results for the basic regression equations are discussed in detail in the next section.

3 Main Regression Results

In this section, we describe the results of the basic regressions. We have two parallel sets of basic regressions. In the first, we estimate our regressions using the data for all candidates in the 20 states that we have included in our analysis. We then run the same regression on a smaller sample which includes only the candidates affiliated with some political party, thus dropping the observations for "independent" candidates. We drop the independent candidates since the majority of these candidates obtained only negligible vote shares.[8] Almost all the results are invariant with respect to the two samples.

Table 8 (p 48) reports the basic regression results. Column I contains the results for our benchmark specification. In subsequent columns we drop the candidates affiliated with the Indian National Congress (INC), Bharatiya Janata Party (BJP), Bahujan Samaj Party (BSP), UPA and NDA respectively from the sample, in order to avoid the adding up constraint. In Table 9 (p 48) we carry out a similar exercise but after dropping the independent candidates from the data.

The variables we are particularly interested in are the criminal dummy variable, relative wealth, as well as the interaction of

Table 6: Explaining the Vote Share of Candidates – Benchmark Specification
(All Candidates, No Interaction between Wealth and Criminal Dummy)

	I All	II Drop INC	III Drop BJP	IV Drop BSP	V Drop UPA	VI Drop NDA
Criminal dummy	0.132 [1.21]	0.158 [1.43]	0.121 [1.07]	0.128 [1.11]	0.154 [1.35]	0.06 [0.55]
Candidates with charges (among top 4) × criminal dummy	0.278*** [5.09]	0.284*** [4.95]	0.306*** [5.24]	0.287*** [4.92]	0.270*** [4.56]	0.316*** [5.54]
Relative wealth	0.008*** [3.32]	0.013*** [5.04]	0.008*** [3.10]	0.008*** [2.87]	0.014*** [4.61]	0.008*** [3.16]
Education dummy for undergrad degree	0.148*** [4.12]	0.134*** [3.69]	0.135*** [3.65]	0.164*** [4.42]	0.123*** [3.40]	0.122*** [3.30]
Education dummy for masters degree	0.260*** [7.27]	0.252*** [6.91]	0.261*** [7.07]	0.256*** [6.90]	0.240*** [6.52]	0.247*** [6.71]
State incumbent	1.772*** [24.34]	2.072*** [24.57]	1.926*** [20.25]	1.599*** [18.74]	1.976*** [22.20]	1.812*** [18.59]
Incumbent Member of Parliament	0.845*** [9.69]	1.092*** [10.22]	0.946*** [8.87]	0.929*** [10.31]	1.105*** [9.82]	0.931*** [8.39]
Fixed effects for constituencies	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects for parties	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,729	6,338	6,341	6,281	6,145	6,213
Adj R-squared	0.782	0.757	0.763	0.79	0.752	0.762

*,**,*** indicate that the coefficient is significantly different from zero at 10%, 5% and 1% levels of significance, respectively. Robust t statistics are reported in parentheses. The variables are defined in Appendix A1 and in the text. The dependent variable is calculated as $\log\left(\frac{Vote\ share_i}{1 - vote\ share_i}\right)$. In column I we estimate the regression for all the candidates.
In column II -VI we drop candidates belonging to specific parties or coalition groups from the sample of all candidates.

Table 7: Explaining the Vote Share of Candidates – Benchmark Specification
(No Independent Candidates, No Interaction between Wealth and Criminal Dummy)

	I All	II Drop INC	III Drop BJP	IV Drop BSP	V Drop UPA	VI Drop NDA
Criminal dummy	0.218 [1.40]	0.264* [1.66]	0.226 [1.33]	0.208 [1.17]	0.222 [1.30]	0.15 [0.90]
Candidates with charges (among top 4) × criminal dummy	0.229*** [3.10]	0.213*** [2.76]	0.266*** [3.23]	0.244*** [2.91]	0.203** [2.45]	0.280*** [3.46]
Relative wealth	0.007*** [3.23]	0.010*** [3.74]	0.007*** [3.24]	0.008*** [3.15]	0.011*** [3.49]	0.007*** [3.42]
Education dummy for undergrad degree	0.214*** [3.56]	0.191*** [3.01]	0.185*** [2.79]	0.263*** [3.96]	0.166*** [2.58]	0.158** [2.38]
Education dummy for masters degree	0.339*** [6.02]	0.321*** [5.42]	0.359*** [5.77]	0.358*** [5.77]	0.293*** [4.80]	0.332*** [5.29]
State incumbent	1.741*** [24.10]	2.008*** [23.21]	1.860*** [18.08]	1.654*** [18.70]	1.905*** [20.47]	1.748*** [16.69]
Incumbent Member of Parliament	0.783*** [8.89]	1.009*** [9.35]	0.893*** [8.15]	0.859*** [9.28]	1.022*** [8.96]	0.893*** [7.88]
Fixed effects for constituencies	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects for parties	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,629	3,238	3,241	3,181	3,045	3,113
Adj R-squared	0.773	0.767	0.76	0.787	0.769	0.764

*,**,*** Indicate that the coefficient is significantly different from zero at 10%, 5% and 1% levels of significance respectively. Robust t statistics are reported in parentheses. The dependent variable is calculated as $\log\left(\frac{Vote\ share_i}{1 - vote\ share_i}\right)$. In column I we estimate the regression for all the candidates affiliated to some party, thus dropping independent candidates. In subsequent columns, we drop candidates belonging to specified parties or coalitions.

Table 8: Explaining the Vote Share of Candidates – Benchmark Specification (All Candidates)

	I All	II Drop INC	III Drop BJP	IV Drop BSP	V Drop UPA	VI Drop NDA
Criminal dummy	-1.06***	-1.05***	-1.04***	-1.27***	-1.05***	-0.88**
	(2.83)	(2.83)	(2.70)	(3.30)	(2.74)	(2.30)
Candidates with charges (among top 4) × Criminal dummy	0.25*** (4.48)	0.25*** (4.37)	0.29*** (4.71)	0.26*** (4.38)	0.24*** (4.03)	0.29*** (5.10)
Relative wealth	0.007***	0.012***	0.007***	0.007***	0.013***	0.007***
	(3.02)	(4.51)	(2.81)	(2.65)	(4.11)	(2.90)
Wealth log × criminal dummy	0.082***	0.084***	0.081***	0.096***	0.084***	0.066**
	(3.19)	(3.27)	(3.01)	(3.64)	(3.16)	(2.46)
Education dummy for undergrad degree	0.14*** (4.03)	0.13*** (3.63)	0.13*** (3.55)	0.16*** (4.33)	0.12*** (3.35)	0.119*** (3.23)
Education dummy for masters degree	0.26*** (7.16)	0.25*** (6.79)	0.257*** (6.95)	0.25*** (6.74)	0.24*** (6.42)	0.244*** (6.61)
State incumbent	1.76***	2.054***	1.910***	1.599***	1.961***	1.798***
	(24.10)	(24.18)	(19.89)	(18.75)	(21.89)	(18.28)
Incumbent Member of Parliament	1.77*** (24.15)	2.06*** (24.29)	1.91*** (19.99)	1.60*** (18.69)	1.97*** (21.99)	1.80*** (18.38)
Fixed effects for constituencies	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects for parties	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,729	6,338	6,341	6,281	6,145	6,213
Adj R-squared	0.78	0.76	0.764	0.79	0.75	0.76

*,**,*** indicate that the coefficient is significantly different from zero at 10%, 5% and 1% levels of significance respectively. Robust t statistics are reported in parentheses. The regression equation is in equation 8 and the variables are defined in Appendix A1 and in the text. The dependent variable is calculated as $\log\left(\frac{\text{Vote share}_i}{1 - \text{vote share}_i}\right)$. In column I we estimate the regression for all the candidates. In column II-VI we drop candidates belonging to specific parties or coalition groups form the sample of non independent candidates.

Table 9: Explaining the Vote Share of Candidates (Candidates Who Are Affiliated With a Political Party)

	I Party Affiliated	II Drop INC	III Drop BJP	IV Drop BSP	V Drop UPA	VI Drop NDA
Criminal dummy	-0.97*	-1.04**	-0.85*	-1.20**	-1.18**	-0.59
	(1.93)	(2.17)	(1.65)	(2.21)	(2.21)	(1.15)
Candidates with charges (among top 4) × criminal dummy	0.198*** (2.60)	0.17** (2.18)	0.23*** (2.76)	0.21** (2.48)	0.160* (1.88)	0.26*** (3.06)
Relative wealth		0.009*	0.009**	0.012***	0.009*	0.009**
		(1.79)	(2.41)	(3.52)	(1.65)	(2.48)
Wealth log × criminal dummy	0.079**	0.089***	0.07**	0.095***	0.096***	0.05
	(2.37)	(2.79)	(2.13)	(2.64)	(2.71)	(1.48)
Education dummy for undergrad degree	0.23*** (2.76)	0.19*** (2.94)	0.18*** (2.70)	0.25*** (3.83)	0.16** (2.49)	0.15** (2.30)
Education dummy for masters degree	0.35*** (4.37)	0.32*** (5.33)	0.36*** (5.70)	0.36*** (5.70)	0.29*** (4.69)	0.33*** (5.25)
State incumbent	1.74***	2.0***	1.87***	1.64***	1.91***	1.76***
	(23.95)	(23.05)	(17.95)	(18.53)	(20.44)	(16.60)
Incumbent Member of Parliament	0.78*** (8.88)	0.99*** (9.21)	0.89*** (8.08)	0.86*** (9.29)	1.02*** (8.87)	0.89*** (7.80)
Fixed effects for constituencies	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects for parties	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,629	3,227	3,231	3,172	3,037	3,103
Adj R-squared	0.774	0.766	0.76	0.787	0.768	0.763

*,**,*** Indicate that the coefficient is significantly different from zero at 10%, 5% and 1% levels of significance respectively. Robust t statistics are reported in parentheses..The regression equation is in equation 11 and the variables are defined in Appendix A1 and in the text. The dependent variable is calculated as $\log\left(\frac{\text{Vote share}_i}{1 - \text{vote share}_i}\right)$. In column I we estimate the regression for the candidates we estimate the regression for the candidates who are affiliated with a political party; so we drop all independent candidates from the sample. In subsequent columns, we drop candidates belonging to specified parties or coalitions.

the criminal dummy with wealth and with the number of other candidates with criminal charges in the constituency. Table 8 shows our results. The negative coefficient on the criminal dummy shows that tainted candidates lose vote share relative to the others. Relative wealth has a positive effect on vote shares. The coefficient of (log) wealth interacted with the criminal charge dummy is positive, implying that the loss in vote share is smaller for a wealthier candidate. Similarly, the coefficient of the interaction between the number of other tainted candidates with criminal charges and the criminal dummy is positive and significant in the regressions for all the

candidates. This implies that the stigma attached to being a tainted candidate declines if there are other tainted candidates in the constituency.

Among other results, the high education status of the candidates has a positive effect on vote share. We also find that incumbency at the candidate level as well as at the party level in the state increases the vote share of the candidates.[9] Most of these results are robust to the exclusion of independent candidates from the sample. Importantly, the qualitative results hold irrespective of which party or coalition is dropped from the sample in order to take care of the adding-up constraint.

We now report on some robustness checks. Since the primary purpose of the paper is to throw light on voter response to tainted candidates, we conduct a key robustness test by constructing the dummy for criminal charges in an alternative way. This dummy takes value 1 if the candidate faces at least *three* criminal charges (instead of *two* in the earlier specification), and zero otherwise. Construction of the dummy in this way reduces the possibility of labelling a candidate as tainted if the charges against him are politically motivated or perhaps arising from violations of the law while undertaking political activities. The results are qualitatively similar to the ones obtained earlier for most of the variables. The coefficients of the criminal dummy and the interaction between wealth and criminal dummy are somewhat larger than before, thus indicating that the loss of vote share is larger for a candidate who faces three or more charges than for the candidates with at least two charges. For such candidates, additional wealth helps in reducing the stigma by a larger amount as well.

Table 10 (p 49) reports some additional robustness checks. Column III in Table 10 includes the interaction of state incumbent and criminal dummy, while column IV includes age and gender of the candidate in the regressions. Finally, in the last column we include the number of all candidates with criminal charges in a constituency rather than only against the top four candidates by vote share, interacted with the dummy for criminal candidates.

The results show that the coefficients of the main variables of interest – wealth or relative wealth, criminal dummy and the interaction of wealth and criminal dummy, retain their significance. The only variable which loses significance in some of the specifications is the interaction of the number of charges against other candidates with criminal dummy.

Some other robustness tests are reported in Table 11. In column I, relative wealth is calculated as the ratio of the candidate's own wealth to the sum of the wealth of candidates who received at least 3% of the total votes. Similarly, the number of candidates with charges also includes the data for only these candidates. In column II we estimate the regressions using the data only for the constituencies reserved for candidates from the scheduled castes and scheduled tribes. In the last column we estimate the regressions only for the constituencies which are not reserved for the candidates of the scheduled castes or scheduled tribes. Again all our main results hold – the criminal dummy has a negative coefficient, wealth or relative wealth has a positive coefficient, and the interaction of wealth and criminal dummy has a positive coefficient. The coefficient of other candidates with charges is mostly positive, but insignificant in some of the specifications.

Table 10: Explaining the Vote Share of Candidates – Robustness Tests

	I	II	III	IV	V
Criminal dummy			-1.27***	-0.98***	-1.02**
			(3.30)	(2.65)	(2.53)
Criminal dummy (more than 2 cases)	-1.15*	-1.45**			
	(1.87)	(2.15)			
Candidates with charges (among top 4) × criminal dummy (>2 cases)	0.28***	0.195**			
	(3.76)	(2.03)			
Candidates with charges (among top 4) × criminal dummy			0.27***	0.25***	
			(4.84)	(4.43)	
Candidates with charges × criminal dummy					0.037
					(1.22)
Relative wealth	0.008***		0.007***	0.007***	0.007***
	(3.22)		(2.88)	(3.00)	(2.98)
Relative wealth (no independents)		0.008***			
		(2.69)			
Wealth log × criminal dummy			0.099***	0.078***	0.10***
			(3.71)	(3.03)	(3.95)
Wealth log × criminal dummy (>2 cases)	0.089**	0.113***			
	(2.21)	(2.63)			
Education dummy for undergrad degree	0.15***	0.22***	0.14***	0.14***	0.14***
	(4.17)	(3.60)	(4.02)	(4.04)	(4.01)
Education dummy for masters degree	0.26***	0.35***	0.25***	0.25***	0.25***
	(7.25)	(6.09)	(7.10)	(6.92)	(7.10)
State incumbent	1.79***	1.76***	1.85***	1.77***	1.78***
	(24.37)	(24.12)	(24.06)	(24.18)	(24.31)
Incumbent Member of Parliament	0.84***	0.78***	0.83***	0.82***	0.84***
	(9.59)	(8.82)	(9.57)	(9.43)	(9.56)
State incumbent* criminal dummy			-0.52***		
			(3.26)		
Age				0.005***	
				(3.78)	
Gender				0.036	
				(0.69)	
Fixed effects for constituencies	Yes	Yes	Yes	Yes	Yes
Fixed effects for parties	Yes	Yes	Yes	Yes	Yes
Observations	6,729	3,618	6,729	6,728	6,729
Adj R-squared	0.781	0.772	0.783	0.783	0.781

*,**,*** indicate that the coefficient is significantly different from zero at 10%, 5% and 1% levels of significance, respectively. Robust t statistics are reported in parentheses. The regression equation is in equation 11 and the variables are defined in Appendix A1 and in the text. The dependent variable is calculated as $\log\left(\frac{\text{Vote share}_i}{1-\text{vote share}_i}\right)$. In columns I and II we include a new criminal charge dummy, which takes a value 1 only if the candidates face at least three charges. In column I we estimate the regression for all the candidates and in column II we estimate the regression after dropping independent candidates from the regression. In column III we include the interaction of state incumbent and criminal dummy. In column IV we include age and gender of the candidates in the regressions, and in column V, we include the interaction variable of all candidates with charges in the constituency and criminal dummy (rather than the candidates with charges among top four candidates interacted with criminal dummy).

Table 11: Explaining the Vote Share of Candidates – More Robustness Tests

	I Different Reference Group for Relative Wealth	II Only Reserved Constituencies	III Only General Constituencies
Criminal dummy	-1.08***	-1.79*	-0.85**
	(2.86)	(1.87)	(2.01)
Candidates with charges (at least 3% vote share) × criminal dummy	0.08		
	(1.42)		
Candidates with charges (among top 4) × criminal dummy		0.30*	0.25***
		(1.91)	(4.00)
Relative wealth (candidates with at least 3% votes)	0.013***		
	(4.10)		
Relative wealth		0	0.009***
		(0.15)	(4.56)
Wealth log × criminal dummy	0.109***	0.140**	0.064**
	(4.49)	(1.97)	(2.26)
Education dummy for undergrad degree	0.14***	0.19**	0.11**
	(3.98)	(2.53)	(2.54)
Education dummy for masters degree	0.26***	0.39***	0.21***
	(7.14)	(4.83)	(4.95)
State incumbent	1.79***	1.71***	1.76***
	(24.52)	(11.30)	(20.41)
Incumbent Member of Parliament	0.82***	0.66***	0.88***
	(9.37)	(3.24)	(8.62)
Fixed effects for constituencies	Yes	Yes	Yes
Fixed effects for parties	Yes	Yes	Yes
Observations	6,715	2,021	4,708
Adj R-squared	0.781	0.764	0.787

*,**,*** indicate that the coefficient is significantly different from zero at 10%, 5% and 1% levels of significance respectively. Robust t statistics are reported in parentheses. The regression equation is in equation 11 and the variables are defined in Appendix A1 and in the text. The dependent variable is calculated as $\log\left(\frac{\text{Vote share}_i}{1-\text{vote share}_i}\right)$. In column I, relative wealth is calculated with respect to the wealth of the candidates who obtained at least 3% of the vote share. In column II we estimate the regression for only the candidates who contested elections from a constituency reserved for the candidates of scheduled castes or scheduled tribes; in column III we estimate the regression for the unreserved constituencies.

We have conducted two more robustness tests, but do not report the results. In one, we drop one state at a time and estimate our benchmark specification with the rest of the data. All of our results hold with minor variations in the coefficients or the significance levels. This robustness test confirms that our results are not driven by any outlier state. Second we estimate regressions similar to those in Table 10 by eliminating the independent candidates from the sample. The qualitative results remain unchanged.

These results seem to leave very little doubt that voters do punish tainted candidates – this conclusion remains true irrespective of the specification chosen by us, and also remains true when we leave independents out of the regression exercise. However, this raises the obvious question. Why do political parties nominate so many tainted candidates when they have so many other aspiring candidates fighting for a party ticket? As we have mentioned earlier, Aidt et al (2011) construct a theoretical model which assumes that tainted candidates have some electoral advantage which induces political parties to nominate them despite some reputational cost. They do not specify the nature of the electoral advantage, but mention in passing that it could be the power of criminal candidates to intimidate voters who are likely to vote for their rivals. If this were the case, then one would expect voter turnout to be lower the greater is the number of tainted candidates. Table 12 (p 50) negates this hypothesis – the data seem to show no negative relationship between voter turnout and the number of tainted candidates in a constituency.

Table 12: Voter Turnout and the Number of Candidates with Criminal Charges
(Dependent Variable: Percentage of Eligible Voters Who Voted)

	I	II	III	IV	V	VI
Number of candidates with at least two charges	0.079 (0.28)	0.34 (1.21)	0.23 (0.79)			0.20 (0.72)
Total candidates		-0.23*** (4.14)	-0.24*** (4.30)			-0.23*** (4.18)
Number of candidates with at least two charges from a large party				0.861** (2.08)	1.0** (2.38)	
Number of candidates from large parties					-0.579* (1.86)	
Dummy for a constituency reserved for the scheduled caste candidates			-2.25*** (2.91)	-1.25 (1.65)	-1.397* (1.83)	-2.23*** (2.85)
Dummy for a constituency reserved for the scheduled tribe candidates			1.14 (0.94)	2.61** (2.24)	2.74** (2.31)	0.51 (0.41)
Literacy						-0.093** (2.09)
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects for parties	No	No	No	No	No	No
Observations	506	506	506	506	506	506
Adj R-squared	0.78	0.788	0.792	0.784	0.785	0.794

*, **, *** indicate that the coefficient is significantly different from zero at 10%, 5% and 1% levels of significance respectively. Robust t statistics are reported in parentheses. Dependent variable is per cent of eligible voters who voted. Regressions are estimated using linear OLS regressions. A large party refers to a national or a state party. Literacy rate refers to the rate of literacy for each constituency in 2008, the data for which is obtained from Indicus Analytics.

Table 13: Candidate Wealth and Criminal Dummy (Dependent Variable: Candidate Wealth, Log)

	I	II	III	IV	V
Criminal dummy	0.779*** (7.57)	0.781*** (7.58)	0.646*** (6.27)	0.762*** (5.79)	0.714*** (6.06)
Dummy for national party	2.577*** (42.74)	2.072*** (31.15)			1.924*** (22.36)
Dummy for state party	1.882*** (19.82)	1.625*** (17.55)			1.426*** (13.21)
Education dummy for undergrad degree		0.744*** (11.11)	0.648*** (9.79)	0.656*** (5.29)	0.701*** (7.37)
Education dummy for masters degree		1.041*** (16.07)	0.820*** (12.51)	1.020*** (8.63)	1.021*** (11.88)
Incumbent Member of Parliament		1.135*** (11.64)	0.872*** (8.41)	1.097*** (10.54)	1.041*** (10.71)
Fixed effects for constituencies	Yes	Yes	Yes	Yes	Yes
Fixed effects for parties	No	No	Yes	No	No
Observations	7,173	6,733	6,729	2,075	3,629
Adj R-squared	0.253	0.305	0.372	0.208	0.29

*, **, *** Indicate that the coefficient is significantly different from zero at 10%, 5% and 1% levels of significance respectively. Robust t statistics are reported in parentheses. Dependent variable is log wealth of the candidates. Dummy for national party takes a value 1 if the candidate belongs to a national party, and zero otherwise; dummy for a state party takes a value 1 if the candidate belongs to a state party, and zero otherwise. In column IV we estimate regressions only for the candidates of national parties; and in column V regressions are estimated only for candidates who are affiliated with one of the political party, thus dropping the "Independent candidates".

An alternative hypothesis advanced by Vaishnav (2012) is that tainted candidates are wealthier. In fact, he finds empirical support for this hypothesis in his data set which consists of elections in various state assemblies. As Table 13 shows, this seems to be true even in our sample. So, it seems plausible to argue that tainted candidates use their greater wealth to "buy" their tickets. They can use their wealth to campaign more intensively, and perhaps also contribute to party funds. Unfortunately, we have no data on campaign expenditure (other than the self-reported wealth of the candidates) to empirically verify any other hypothesis.[10]

4 Conclusions

Our main empirical results suggest that voters do punish candidates who have criminal charges against them. However, these tainted candidates are able to overcome this electoral disadvantage because they have greater wealth, and wealth plays a significant role in increasing vote shares. The most plausible channel through which wealth affects vote shares is of course through campaign expenditures, which are likely to be positively related to wealth.

There is now a fair body of evidence suggesting that voters who have information about the corruption or non-performance of incumbent politicians do punish the latter. For instance, Ferraz and Finan (2008) use detailed Brazilian electoral and audit data to show that new information about political corruption reduces the probability of re-election for corrupt incumbents. Bobonis et al (2012) find that publicly available pre-election municipal audits significantly reduce the level of corruption in Puerto Rican municipalities.[11] Closer home, Banerjee et al (2011) conclude, on the basis of a field experiment conducted before the Delhi state legislative elections, that voters who had access to information about incumbent performance punished worse performing incumbents and those facing better qualified challengers – these incumbents then received significantly fewer votes.

Our empirical results, along with this body of evidence, suggest that it is important for voters to be better informed about candidate characteristics. The mere requirement that candidates file affidavits with the Election Commission about their characteristics is of limited use if voters do not have access to this information. Perhaps, the Election Commission needs to play a more active role in disseminating this information. The commission must also think seriously about enhancing the existing ceilings on campaign expenditure since practically no candidate or party adheres to the current limits on expenditure. However, the commission must ensure that all candidates adhere to the enhanced (but realistic) ceiling. This will then at least reduce the "wealth advantage" enjoyed by tainted politicians.

NOTES

1 That is, courts have decided that these charges have sufficient credibility for judicial proceedings to be initiated. However, this does not mean that these charges have culminated in convictions.

2 However, the judgment has been of immense help to several researchers who have exploited the information contained in the affidavits. Apart from the present paper, see, for instance, Aidt et al (2011), Chemin (2008), Paul and Vivekananda (2004), Vaishnav (2012).

3 Vaishnav (2012) also finds that potentially criminal candidates have higher wealth.

4 They measure a politician's quality by his record of illegal and corrupt behaviour as identified in a field survey.

5 However, there are some questionable issues with their empirical exercise. On this, see Dutta and Gupta (2012).

6 As robustness checks, we choose alternative specifications where (i) the criminal dummy takes value one if a candidate has three or more criminal charges; or (ii) the number of criminal charges instead of a criminal dummy is used as an explanatory variable.

7 We report results by dropping the candidates of the INC, BJP or BSP, as well as the UPA and NDA coalitions.

8 There were 3,825 independent candidates with an average vote share of about 0.80%. Only 10 independent candidates won in the 2009 election.

9 Gupta and Panagariya (2012) also come to the same conclusion.

10 The recent Supreme Court ruling disqualifying those legislators who are convicted in criminal charges may have a significant effect. In particular, it may deter political parties from nominating candidates with multiple charges against them since they stand to lose a seat if the candidate is convicted.

11 See also Brollo (2011).

REFERENCES

Aidt, Toke, Miriam A Golden and Devesh Tiwari (2011): "Incumbents and Criminals in the Indian National Legislature", Cambridge Working Papers in Economics 1157, *Faculty of Economics*, University of Cambridge, September.

Banerjee, Abhijit V and Rohini Pande (2009): "Parochial Politics: Ethnic Preferences and Political Corruption", CID Working Paper 147, Center for International Development at Harvard University, July.

Banerjee, Abhijit V, Selvan Kumar, Rohini Pande and Felix Su (2011): "Do Informed Voters Make Better Choices? Experimental Evidence from Urban India", mimeo, November.

Bobonis, Gustavo J, Luis R Cámara Fuertes and Rainer Schwabe (2012): "The Dynamic Effects of Information on Political Corruption: Theory and Evidence from Puerto Rico", Documento de Investigación 2012-14, Banco de México, December.

Brollo, Fernanda (2011): "Who Is Punishing Corrupt Politicians – Voters or the Central Government? Evidence from the Brazilian Anti-Corruption Program", Working Paper 336, Innocenzo Gasparini Institute for Economic Research at Università Bocconi, January.

Chemin, Matthieu (2008): "Do Criminal Politicians Reduce Corruption? Evidence from India", CIRPÉE Working Paper 08-25, Department of Economics, University of Quebec at Montreal, September.

Dutta, Bhaskar and Poonam Gupta (2012): "How Indian Voters Respond to Candidates with Criminal Charges: Evidence from the 2009 Lok Sabha Elections", Working Papers 12/109, National Institute of Public Finance and Policy.

Ferraz, Claudio and Frederico Finan (2008): "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes", *Quarterly Journal of Economics*, 123(2): 703-45.

Gupta, Poonam and Arvind Panagariya (2012): "Growth and Election Outcomes in a Developing Country", Working Papers 11/92, National Institute of Public Finance and Policy.

Paul, Samuel and M Vivekananda (2004): "Holding a Mirror to the New Lok Sabha", *Economic & Political Weekly*, 39(45): 4927-34.

Vaishnav, Milan (2012): "The Market for Criminality: Money, Muscle and Elections in India" in *The Merits of Money and "Muscle": Essays on Criminality, Elections and Democracy in India*, PhD Dissertation, Columbia University.

Appendix A1: Description and Data Sources of Variables

Variable	Source	Description
Dependent variable	Election Commission and own calculation	The dependent variable for candidate i is calculated as $\log\left(\dfrac{\text{Vote share}_i}{1-\text{vote share}_i}\right)$
Criminal dummy	Election Commission and Association for Democratic Reforms (ADR)	The dummy takes a value 1 if the candidate has two or more criminal cases against him, and zero otherwise. In robustness tests the dummy takes a value 1 if the candidate has three or more cases against him.
Wealth* criminal dummy	Election Commission	Interaction variable calculated as Log wealth x dummy for criminal charges
Relative wealth	Election Commission and own calculation	Wealth of the candidate/average wealth of all other candidates in the constituency
Candidates with charges	Election Commission and ADR	Number of candidates within the constituency who face criminal cases. In most specifications, as mentioned in the tables, we look at the number of such candidates within top four candidates by vote share, and in robustness tests we include the number of all candidates with charges within the constituency
Candidates with charges* criminal dummy	Election Commission and ADR and own construction	Interaction between number of candidates with charges and criminal charges dummy.
Education: Dummy for undergraduate degree Dummy for masters degree	Election Commission	Dummy for undergraduate degree takes a value 1 if a candidate has education up to undergraduate, and zero otherwise; and dummy for masters degree takes a value 1 for education level higher than undergraduate (or for a technical or professional degree) and zero otherwise.
Age	Election Commission	In years
Gender dummy	Election Commission	Dummy takes a value 1 if the candidate is a female, and 0 otherwise
Incumbent Member of Parliament	Various sources on the web	The dummy takes a value 1 if the candidate was a member of the previous Lok Sabha, and zero otherwise.
State incumbent party	Various sources on the web and different issues of *Economic & Political Weekly*	The dummy takes a value 1 if the candidate belongs to a party which was in power in state government in 2008-09 before the Lok Sabha elections. The state incumbent parties are: Andhra Pradesh, Indian National Congress (INC), TRS; Bihar: JD(U), Bharatiya Janata Party (BJP); Chhattisgarh: BJP; Delhi: INC; Goa: INC, NCP; Gujarat: BJP; Himachal Pradesh: BJP; Haryana: INC; Kerala: CPI(Marxist), CPI; Maharashtra: INC, NCP; Madhya Pradesh: BJP; Odisha: Biju Janata Dal; Punjab: Shiromani Akali Dal, BJP; Rajasthan: BJP; Tamil Nadu: Dravida Munnetra Kazhagam, INC; Uttarakhand: BJP; Uttar Pradesh: Bahujan Samaj Party; West Bengal: CPI (Marxist), RSP; Karnataka: BJP; Jharkhand: JMM, BJP.

Appendix A2: Summary Statistics of Variables

Variable	Observations	Average	Minimum	Maximum
Per cent of votes obtained	7,192	6.82	0.02	78.80
$\log\left(\dfrac{\text{Vote share}_i}{1-\text{vote share}_i}\right)$	7,192	-4.61	-8.52	1.31
Criminal dummy (at least two cases)	7,173	0.07	0	1
Number of candidates with charges (in top four candidates)	7,192	1.22	0	4
Relative wealth	7,192	2.26	0.00	452.22
Wealth (1000s)log	7,192	13.81	0.69	22.57
Education dummy for undergraduate degree	6,749	0.22	0	1
Education dummy for masters degree	6,749	0.27	0	1
Incumbent Member of Parliament	7,192	0.05	0	1
State incumbency party	7,192	0.07	0	1
Age	7,191	45.98	25	88
Gender dummy	7,192	0.07	0	1

* We drop three outliers from the regressions when the relative wealth exceeded 500. Statistics ate given for the data for 20 states that we have used in the paper.