Criminality and Incumbency of Candidates to the National Legislature in India[1]

Toke Aidt
University of Cambridge
`toke.aidt@econ.cam.ac.uk`
and

Miriam A. Golden
University of California at Los Angeles
`golden@ucla.edu`
and

Devesh Tiwari
University of California at San Diego
`dtiwari@ucsd.edu`

August 31, 2010
Version 3.0. Reports preliminary results. Comments welcome.

Abstract

Utilizing new data on criminal charges against candidates to India's Fourteenth and Fifteenth Lok Sabha elections, we study the conditions that resulted in approximately a quarter of those elected to each legislature facing or having previously faced criminal charges. We show that Indian political parties are more likely to select criminals to run as candidates in electoral districts with lower levels of literacy and when parties face greater electoral uncertainty. The first finding resonates with a large literature that documents that poorly informed voters are more vulnerable to manipulation by unscrupulous politicians. The latter finding runs contrary to standard democratic theory, which claims that more electoral competition improves democratic accountability. We provide a formal model to gain insight into this surprising result. Subsidiary results show that the well-known incumbency disadvantage characterizing India legislative elections stems from the extraordinary performance of criminal candidates.

1 Introduction

Why would a political party in a competitive democratic system recruit known criminals to run for national public office? There are multiple puzzling aspects to this phenomenon. Not only is it puzzling that a party would stand a known criminal, since this ought to prove electorally disadvantageous, it is puzzling that voters, instead of repudiating a criminal candidate for public office, would elect him. And it is genuinely astonishing that this would occur in fully a quarter of nearly 550 single-member legislative districts not once but twice in a row in a highly competitive multiparty system in a well-established, long-standing democratic polity.

We analyze data on the 2004 and 2009 Indian national elections, the first and second legislative elections in which candidates to the national parliament, the Lok Sabha, were required by a 2003 Supreme Court ruling to file sworn affidavits that included, among other pieces of information, declarations of criminal records. Candidates also had to disclose whether they currently faced criminal proceedings.[1] Our dataset contains information drawn from all the affidavits for the more than $12,000$ legislative candidates in these two elections. Empirically, our goal is to isolate the main electoral, demographic, partisan, or other factors that impel parties to select criminals as candidates. Theoretically, we seek to shed light on how electoral democracies generate situations in which accountability is so seriously compromised.

Substantively, who wins an election is of greater interest than who runs for office. If criminals are listed on the ballot but win in only a handful of districts, the phenomenon would be of only marginal interest or political importance. But in fact, the reverse is the case. As the data reported in Table 1 shows, in both 2004 and 2009 candidates whose affidavits reveal them as criminals have a much larger likelihood of winning than non-criminals. In 2004, more than a quarter of criminals won their seat compared with a success rate of only 8 percent for non-criminal candidates. The 2009 elections were much more competitive — the total number of candidates rose 50 percent over 2004 — but even so criminals won 14 percent of the time compared with an election rate of 6 percent for those who did not report criminality on their affidavits. In both elections, criminals were two to three times as likely to win as non-criminals. This apparent electoral advantage makes criminals obviously attractive to parties to secure as candidates. Although in most circumstances we expect criminality to constitute an electoral hindrance, in contemporary India the reverse appears to be the case.

[Table 1 about here]

Our main empirical results of theoretical interest are that in 2004 and again in 2009 criminals are more likely to appear on the ballot in electoral districts with a larger proportion of illiterates in the electorate and where the electoral contest is more competitive. We proceed in four steps to document and analyze these findings. First, we briefly review literature related to our study. Second, using formal methods we model the decision by a political party to select a criminal candidate. The model generates hypotheses to investigate empirically. We then turn to empirical estimations,where we proceed using two statistical methods. Conceptualizing our dependent variable as the likelihood that a criminal candidate appears on the ballot in an electoral district, we first use logistic regression to study the problem. We then study the percentage of criminal candidates on the ballot in each district, for which we use OLS. The latter strategy allows us to pull out and discuss some quantities of interest that are not easily available with logistic regression. A fourth section uses a regression discontinuity design to investigate selection effects. We do this to rule out the possibility that criminals appear on the ballot where a party would have won the election anyway. In addition, the regression discontinuity analysis allows us to document that the well-known but relatively unusual electoral disadvantage faced by incumbents in India is due to the frequency with which non-criminal incumbents square off against criminal challengers. Criminals, we show, boast an extraordinary electoral advantage in India. We conclude with some ideas for understanding why this might be the case.

---

[1]Because the penalty for non-compliance or furnishing inaccurate information was to debar the candidate from standing and to make him liable to prosecution — and because the affidavits were released to the media and the opposition — strong incentives were in place encouraging accuracy in the material furnished to the Election Commission.

Our study has two theoretically unexpected and normatively troubling implications. First, as we document in the next section, a number of studies show that once information about political corruption is released and disseminated to voters, they will use that information to elect honest challengers or, what comes to the same thing, public officials, anticipating voter backlash, will substantially reduce the frequency of corrupt activities. We observe voters successfully using new information in settings as diverse as Uganda, Brazil, and Italy to obtain less corruption in specific domains. On the basis of these studies, we expect that the release of public information about criminality among the Indian political class will reduce the number of criminals elected — if not immediately then certainly by the second election in which candidates file affidavits reporting their legal status. Instead, we observe a nearly uniform persistence in the rate of criminality among those elected. This suggests that information alone is inadequate in reining in corruption and that prior studies must have omitted important variables from consideration. Other research that we summarize below indicates that ethnic or programmatic attachments to a party may dissuade voters from turning against that party even if it its leaders engage in corrupt or criminal behavior.

Second, a number of papers have documented an incumbency disadvantage for legislators at the federal and state levels in India. Our results show that the Indian incumbency disadvantage, which until now has lacked explanation, is linked to the high rate of criminality among candidates. Incumbency disadvantage occurs because of the extraordinary political advantage of criminals against even experienced politicians. We thus appear to be witnessing the nearly wholesale take-over of large chunks of an established democratic political system by persons with ties to organized crime. Moreover, this is occurring even as Indian politics becomes increasingly competitive. This too suggests that democratic accountability can encounter unexpected obstacles that have not been previously identified or well understood.

2 Related Literature

Our paper builds on three prior studies that use the same affidavit information that we do, on studies of electoral responses to political corruption, on studies of incumbency disadvantage in India, and on studies of political competition and accountability. We briefly review these various classes of studies.

We have identified three other studies that analyze candidate affidavits (Chemin (2008); Banerjee and Pande (N.d.); Viashnav (2010)) although for somewhat different purposes or, in the cases of Banerjee and Pande (N.d.) and Viashnav (2010), at levels of government below the national. Chemin (2008) studies local outcomes after 2004 and reports that where criminals are elected into public office, bureaucratic corruption is reduced but poverty rates increase. These findings are consistent with a theoretical framework in which criminals enter and capture the polity; such a view would expect to see organized crime reduce evidence of unorganized criminality and also encourage greater economic inequality once it gained political power. Banerjee and Pande (N.d.) examine the 2004 affidavits in electoral districts in the state of Uttar Pradesh, a state with a high level of political corruption, and find that winners are more likely to be criminals in districts with more low-caste voters. The broad lines of the argument is that corruption and criminality among politicians in Uttar Pradesh has been due to the rise of low-caste and ethnic voting and that ethnic party politics creates incentives for political corruption. These findings are not inconsistent with ours; we find that areas with more illiterate voters (which are likely to coincide with areas with more low-caste voters) also see more criminals on the ballot. This is also broadly consistent with studies, such as Wilkinson (2007), that contend that poor Indians are more susceptible to patronage appeals. The study closest to ours, finally, is Viashnav (2010), who analyzes the affidavit information of party-affiliated candidates to state elections in India in a political selection framework. His main argument is that criminal candidates are attractive to political parties because they are self-financing, presumably because of the financial returns that their criminal activities have brought them. However, Vaishnav finds no evidence that electoral competitiveness increases the likelihood that a criminal appears on the ballot. We discuss why his results differ from ours in

this respect later.

A number of studies have analyzed how voters respond to revelations of corruption on the part of public officials, including Reinikka and Svensson (2005); Ferraz and Finan (2008); Bobonis and Cámara Fuertes (2009); Brollo (2009); Chang, Golden and Hill (2010); Banerjee et al. (2009). These studies find that revelations of corruption by elected officials result in electoral retribution under certain conditions — for instance, when information about corrupt activities is disseminated widely to voters. When it is not, elected officials with records of malfeasance may be repeatedly reelected.

Our study's findings accord only partially with this view. We show that criminals appear on the ballot less frequently in districts with a larger proportion of literate voters. This is consistent with an informational view of voter behavior vis-a-vis political corruption. It suggests that parties anticipate that literate and better informed voters are more likely to reject criminal candidates and hence do not permit them on the ballot where such voters predominate. However, the data we presented in Table 1 also shows that despite the considerable public outcry that occurred after the 2004 election of more than a hundred and a quarter known criminals to the Lok Sabha, a nearly identical number of criminals secured election to the same legislative institution five years later. Even in 2004, information from the affidavits about which candidates faced criminal indictments had been well publicized. We infer that in India illiterate voters are not effective in repudiating criminal candidates; an alternative interpretation is that Indian political parties are able to subject illiterate voters to criminal representatives. One possible explanation for this, put forward by Banerjee et al. (2009) and building on Banerjee and Pande (N.d.), is that caste voting locks voters into supporting criminal candidates. The experiments reported in Banerjee et al. (2009) show that the information that must be disseminated to curtail electoral support for criminal candidates does not pertain to their status as criminals but rather must explicitly and actively discourage caste voting. This is an obviously lively and open debate to which our study hopes to make a modest contribution.

Various studies show that in India, unlike many other political systems, federal and state legislators face an incumbency disadvantage (Linden (2004); Nooruddin and Chhibber (2007); Uppal (2007); Chakrabarti, Gangopadhyay and Krishnan (N.d.)). This disadvantage is apparently recent and it overlaps in part with the growing turnover of legislative representatives in India. Nooruddin and Chhibber (2007), for instance, show that electorally more volatile states in India also exhibit larger anti-incumbent swings; likewise, Wilkinson (2007) contends that party volatility, electoral turnover, and clientelism have risen together in India since the late 1960s. Our work is structured to make headway into this problem. Using a regression discontinuity design, we analyze incumbency disadvantage together with candidate criminality to estimate the effect of criminal opponents on incumbent reelection probabilities. Our goal is to evaluate whether the proliferation of criminals on the ballot is causally linked to the growing incumbency disadvantage in India.

Finally, many studies of political competitiveness and accountability report that accountability is enhanced by electoral competition. The most compelling statement of this view is the observation advanced by Sen (1981) that famines occur only under non-democratic regimes where leaders do not face reelection prospects. In the Indian context, this view is extended by Besley and Burgess (2002), who show that public food distribution across Indian states in response to falls in grain production and flood damage is improved where newspaper circulation and political competition are greater. Our results fail to corroborate the view that political competition enhances accountability and responsiveness Our empirical results show that where partisan competition is more intense and when a party has less assurance about whether it will win, that party is more likely to list a criminal on the ballot. The aggregate outcome of intense partisan competition is thus to diminish political accountability.

We are not the first to note that legislators may respond to competitive electoral pressures by engaging in criminal or corrupt behavior. In a study of Japan, Nyblade and Reed (2008) differentiate rent-seeking by politicians wishing to line their pockets from electoral corruption. The authors characterize electoral corruption as a way to raise additional funds for political campaigns. In their account, incumbents are more likely to seek additional campaign funds illegally when they experience more electoral uncertainty. Hence,

politicians are more likely to engage in electoral corruption when they are unsure of their prospects for victory.

Our explanation for why Indian political parties list criminal candidates is similar. We contend that they risk listing known criminals on the ballot only in marginal electoral circumstances. When a party anticipates that it will win or lose a seat, it has no reason to alienate its partisan supporters by putting a criminal on the ballot. Instead, criminal candidates are advantageous when parties face electoral uncertainty, especially where there are large numbers of illiterate voters, because these votes, who are poor, can be easily bought by the resources provided by criminal candidates. The reputational costs of criminal candidates are lower among the illiterate. In our story, like that of Nyblade and Reed (2008), criminals are politically productive for parties because of the ability of criminals to increase the returns on political campaigns conditional on the costs of criminality among voters. We do not have information that allows us to assess directly how criminals secure votes, although ethnographic work (Berenschot (2008)) shows that they often use violence and intimidation.

3 A Model of Candidate Selection and Criminality

In this section we present a formal model that is designed to illustrate the logical connections between electoral competition, the costs of fielding a criminal candidate, and the electoral advantage of criminality, on the one hand, and the incentive of political parties to field criminal candidates, on the other. It thereby provides a useful framework for interpreting the empirical results that follow. The model highlights what we believe is the central trade off facing party officials tasked with selecting a candidate in a particular electoral district. The trade off is between the extra votes that criminal candidates can muster for the party and the reputational cost to the party of selecting a candidate with a known criminal record. We stress that in focusing on the demand side, i.e., on why parties may want to select candidates with a criminal record, we take the supply of criminal candidates as given, bracketing the question of why criminals want to run for public office.

3.1 Assumptions

We consider two political parties $j \in \{A, B\}$ that compete in a district for a seat in parliament. In contrast to the standard Downsian or probabilistic voting model (see, e.g., Persson and Tabellini (2000)), we assume that the platforms of the parties are fixed or pre-determined.[2] This allows us to focus on candidate selection: the choice variable of the two parties is the type of candidate to field in the district. Candidates have different types, index by $t \in R^+$, depending on their criminal record. We denote by t_j the type of candidate fielded by party j for $j \in \{A, B\}$ and assume that candidates with a higher index t have worse criminal records. We can think of candidates with a very small t, one close to zero, as being essentially non-criminals and candidates with a large t as having been convicted of serious crimes.[3]

The reason for selecting criminal rather than non-criminal candidates is that criminal candidates are able to muster extra votes for the party. We do not model this explicitly and are agnostic about the sources of this advantage. It could be related to the capacity of criminal networks to intimidate voters to vote for a particular party or to induce (or to discourage) them to turn out to vote in elections, or it could be related to the financial resources that criminals bring with them. Here, we simply postulate that the vote production

---

[2]This assumption is justified by our focus on a particular election district. The platforms of political parties are typically decided at the national level. Accordingly, the party officials tasked with selecting candidates for a given district can be assumed to take the platform of the party as given. By platform, we understand programmatic or identity characteristics of the national party that are distinct from the individuals selected to represent the party in public office.

[3]We conceptualize criminality as a continuous rather than dichotomous variable for reasons of presentational simplicity. In Appendix B, we respecify the model for criminal and non-criminal candidates separately rather than for degrees of criminality.

function of party j is a function of the type of the candidate that it fields $f(t_j)$ and that this function is strictly increasing in t_j. Votes translate into a probability of wining the seat. For party A, this win probability is given by

$$p_A(t_A, t_B; a) = \frac{af(t_A)}{af(t_A) + (1-a) f(t_B)}, \tag{1}$$

where $a \in (0, 1)$ and the win probability of party B is $1 - p_A$. Snyder (1989) uses a similar functional form in his seminal analysis of campaign spending as advertising. It captures the competitiveness of a race in a simple and transparent way and thereby makes it easy to draw the distinction between safe and marginal seats. In particular, the degree of competition is controlled by the parameter a. To see how, begin by supposing that both parties field a candidate of the same type. In that case, the win probability of party A is a and that of party B is $1 - a$, no matter how bad (or good) the criminal records of the two candidates are. Therefore, if $a > \frac{1}{2}$, party A enjoys a "natural advantage" in the district and vice versa if $a < \frac{1}{2}$. This advantage may relate to incumbency or to other factors but is unrelated to the type of candidates fielded. This implies that the race in the district is close and electoral competition at its maximum when $a = \frac{1}{2}$ and any increase or decrease in a from $\frac{1}{2}$ represents a reduction in closeness or competitiveness. In the extreme, if $a = 1$ or $a = 0$, then one of the parties will win with certainty and the seat is not contested in any real sense.

Fielding a criminal candidate is, however, not a free lunch. We specify the cost to party j of fielding a criminal candidate of type t_j as

$$C(t_j) = \mu t_j \tag{2}$$

with $\mu > 0$ representing the (constant) marginal cost of fielding a candidate with a worse criminal record. This fielding cost could, of course, differ by party, but, for simplicity, we assume that it is the same. We think of the fielding cost as being distinct from any specific electoral cost associated with running criminals. Such costs are build into the vote production function, $f(t_j)$, directly, which should be understood as a *net* vote production function. Thus, the fielding cost, C, represents those disadvantages that the party must suffer irrespective of its electoral success in the particular district as a consequence of allowing a criminal of type t_j on its party list. This, among other things, includes reputational costs for the party nationally or locally and the inconvenience for the local party organization that comes from having to associate with criminals.

3.2   Analysis

The objective of each of the two parties is to select and field a candidate of a type that maximizes its win probability net of the fielding cost, i.e., $p_A(t_A, t_B; a) - \mu t_A$ for party A and $1 - p_A(t_A, t_B; a) - \mu t_B$ for party B. The two parties choose candidates simultaneously, taking the choice of the other party as given. The Nash equilibria of the candidate selection game are summarized in the next proposition.

Proposition 1 *Assume that* $f(t_j) = t_j$. *The candidate selection game has one unique symmetric pure strategy Nash equilibrium. In this equilibrium, the two parties select the same type of candidate. The candidate type is given by*

$$t^* = t_A^* = t_B^* = \frac{a(1-a)}{\mu}. \tag{3}$$

The win probability of party A (B) is $a(1-a)$.

Proof. The first-order conditions are:

$$t_A \quad : \quad \frac{D(t_A, t_B) + \left(a - a^2\right) t_B}{\left(t_B + a t_A - a t_B\right)^2} = 0 \tag{4}$$

$$t_B \quad : \quad \frac{D(t_A, t_B) + \left(a - a^2\right) t_A}{\left(t_B + a t_A - a t_B\right)^2} = 0, \tag{5}$$

where $D(t_A, t_B) = -a^2 \mu t_A^2 - 2a\mu \left(1 - a\right) t_A t_B - \mu \left(1 - a\right)^2 t_B^2$. Inspection shows that $t_A = t_B = t > 0$ and so the common type choice must satisfy $D(t, t) + \left(a - a^2\right) t = 0$ or

$$-t \left(-a + t\mu + a^2\right) = 0 \tag{6}$$

which for $t > 0$ implies equation (3). The second-order conditions are satisfied and uniqueness is implied by the general characterization results discussed in (Konrad 2009, ch. 2). ∎

The proposition shows that electoral competition forces the two parties to select criminal candidates despite the fact that it is costly for them to do so and, at equilibrium, it does not actually help them win the seat (the outcome is determined by the natural advantage or disadvantage of the parties, captured by the parameter a). The parties face a prisoners' dilemma-type situation and would be better off if they could agree to field candidates without criminal records.

The type of the chosen candidates depends on two key considerations (see equation (3)). Firstly, the fielding cost obviously plays a role. Intuitively, if it is more costly to field a criminal candidate at the margin, then the two parties have an incentive to field less criminal ones ($\frac{\partial t^*}{\partial \mu} < 0$). In the limit, if it becomes infinitely costly to field a criminal, only non-criminals will be fielded (i.e., $\lim_{\mu \to \infty} t^* = 0$).

Secondly, electoral competition is also important for candidate selection. The parties have the strongest incentive to field criminal candidates in districts in which electoral competition is close. In fact, the "criminal type" is maximized at equilibrium when $a = \frac{1}{2}$ and the race is decided by the toss of a coin. As one of the parties gains a natural advantage and the race becomes less competitive, i.e., a moves away from $\frac{1}{2}$ in either direction, the value of fielding a criminal candidate diminishes. For the party that enjoys the advantage in the race, the reason for this is straightforward: the party does not need to field a criminal to boost its prospects of winning. For the "underdog," or the party lacking a natural advantage, two effects are at work. On the one hand, this party has an incentive to field criminals to catch up with the favorite party. On the other hand, the need to do so is diminished by the fact that the favorite fields a less criminal candidate. At equilibrium, the later effect dominates and both parties select the most criminal candidates for the most open races and vice versa. In the limit, if one of the parties is virtually sure to win the seat (i.e., if a is either 0 or 1), neither of the parties fields a criminal candidate. This is illustrated in Figure 1, which shows the (common) equilibrium type choice (t^*) on the y-axis as a function of the competitiveness parameter a on the x-axis. The relationship is hump-shaped, with a maximum at $a = \frac{1}{2}$ and t^* is zero at $a = 0$ and $a = 1$.

[Figure 1 about here]

Thirdly, the model suggests that the two aspects — the fielding cost and the competitiveness of the race — interact in interesting ways. Mathematically, we notice that

$$\frac{\partial t^*}{\partial a \partial y} = \frac{2a - 1}{\mu^2} \begin{array}{c} < \\ = \\ > \end{array} 0 \text{ for } \begin{array}{c} a < \frac{1}{2} \\ a = \frac{1}{2} \\ a > \frac{1}{2} \end{array} . \tag{7}$$

The interpretation of this cross derivative is that the (positive) effect on the criminal type chosen by the two parties of an increase in the competitiveness of the race (a moves closer to $\frac{1}{2}$) is bigger when the cost of fielding is low. To see this more clearly, imagine two districts, one with a close race and one with a candidate

6

who is a clear favorite. Then, ceteris paribus, the candidates fielded in the former should have worse criminal records than those fielded in the later. Now, imagine that the cost of fielding criminals falls in both districts. This will induce the parties to field candidates with worse criminal records in both districts, but the effect will be larger in the district with the close race because the marginal value of fielding a criminal is larger in that district.

Another way to grasp the interaction effect is to consider Figure 2. It is similar to Figure 1 but shows the relationship between the choice of candidate type and electoral competitiveness for two different values of the fielding cost. The dashed curve represents district L with low fielding costs and the unbroken curve represents district H with high fielding costs. Imagine that party B is the favorite in both districts with $a = a' < \frac{1}{2}$. As a consequence of the lower fielding costs, candidates with worse criminal records are fielded in district L than in district H, i.e., $t_L^* > t_H^*$. Now, suppose that the races become more competitive in both districts, i.e., that a increases to $a' + \Delta a$, where $\Delta a > 0$ represents the increase in a. Criminals with worse records are then fielded in both districts, but we can see graphically that the increase in t^* is larger in district L (the increase in t_L^* is indicated by the length of the dotted arrow) than in district H (the increase in t_H^* is indicated by the length of the unbroken arrow).

[Figure 2 about here]

3.3 Empirical Implications

Empirically, we observe if a party fields a candidate with a criminal record or not in each district rather than the degree of criminality of the candidates as such. It is, therefore, useful to restate the three main predictions of the model in terms of either the probability of fielding a criminal candidate in a given district or in terms of an index of candidate criminality in a district. We base our empirical analyses on these hypotheses.

Let Y_{ijt} be an indicator variable equal to 1 if party j fields a criminal in district i in election t and let $COMP_{it}$ and $COST_{it}$ be empirical measures of the competitiveness of the race in district i in election t and the fielding cost in district i in election t, both defined to be increasing in competitiveness and cost, respectively. Given this, we are interested in estimating a probit model of the following type:

$$\Pr\left[Y_{ijt} = 1\right] = F\left[\beta_0 + \beta_1 COMP_{it} + \beta_2 COST_{it} + \beta_3\left(COMP_{it} * COST_{it}\right) + ..\right] \tag{8}$$

where F represents the cumulative normal distribution and we might condition on observable factors other than $COMP$ and $COST$. Given that, we can restate the model's predictions as:

1. The probability that a party fields a criminal candidate is increasing in the closeness of the race ($\beta_1 > 0$);

2. The probability that a party fields a criminal candidate is decreasing in the fielding cost ($\beta_2 < 0$);

3. The effect on the probability of fielding a criminal candidate associated with an increase in the closeness of the race is larger in districts with low fielding costs.

As an alternative to the probit model, we also consider a linear model where we define the outcome as the proportion of criminal candidates fielded in district i in election t, i.e., $y_{it} = \sum_j Y_{ijt}$. Formally, we can write

$$y_{it} = \alpha_0 + \alpha_1 COMP_{it} + \alpha_2 COST_{it} + \alpha_3\left(COMP_{it} * COST_{it}\right) + .. + \varepsilon_{it} \tag{9}$$

where ε_{it} is an error term with zero mean. In this formulation, the model predicts that $\alpha_1 > 0$, $\alpha_2 < 0$ and $\alpha_3 < 0$.

4 Data and Empirical Analysis

We test the three hypotheses outlined above using each of the two possible operationalizations of the outcome variable discussed: as the probability that a candidate will appear on the ballot and as the rate of criminality of all candidates in the district. We operationalize competitiveness as the percent difference between the share of the vote secured by the winning candidate and the share secured by the runner up. Although we might operationalize this using information from the prior election — when deciding whom to put on the ballot, parties look back to the most recent election and assess the degree of competitiveness in the district on the basis of those results — extensive redistricting just prior to the 2009 elections makes this infeasible. We therefore use a measure of competitiveness drawn from the same election instead of the prior election. One justification for this, in addition to the sheer necessity of doing so, is that we believe that candidates are selected on the basis of relatively current and accurate information collected by the parties operating in the district as to how competitive the race is likely to be. Especially given the costs of fielding a criminal candidate, parties are likely to have used various instruments to collect information about the sentiments of the electorate. Therefore, if the decision to list a criminal on the ballot is in part a function of electoral competitiveness, assessing the degree of competitiveness in the same election simply means that we believe that parties have a relatively good estimate of how close the election is likely to be even in the period leading up to the election when they must make final decisions about which candidate to list.

We proxy the fielding costs to running a criminal candidate with literacy rates in the district. Our rationale for this is that the reputational costs to fielding a criminal candidate are higher where more voters are literate. Literacy carries with it an entire array of characteristics buffering voters from political manipulation, allowing them to exercise greater independence in their vote choices, and as a result rendering it more costly for a party to field criminal candidates. Literate voters have access to written information, including newspaper reports, about the quality of the candidates; they are relatively well educated and hence equipped with some measure of human capital; they are thus more employable and potentially able to operate independently in the labor market; they are as a result less dependent on government resources for everyday survival. Using literacy as a proxy for fielding costs means that we believe that the reputational costs associated with criminal candidates are lower where voters are poor, ignorant, and vulnerable.

Our empirical specifications of the two sets of three hypotheses outlined in the previous section thus invoke two independent variables of theoretical interest as well as their interaction: the competitiveness of the race and the extent of voter literacy. We study these using probit models and, for the proportion of criminal candidates, OLS. As pointed out by Ai and Norton (2003), the interaction effect in the probit model is not captured by the parameter β_3 alone but must be evaluated separately for each combination of the control variables. We return to this point below, but note that the model predicts a negative interaction effect.

4.1 Institutional and Political Context

2004 was the first year in which candidates for the federal legislature across India's 543 single-member electoral districts were legally required to file affidavits in which they reported criminal histories or pending criminal charges. In Figure 3, we map India's electoral districts, differentiating them according to whether at least one criminal appeared on the ballot or not in 2004 and 2009. [Note: Map not yet available for 2009.] In 2004, fully half of the districts witnessed criminal candidates; in 2009, criminal candidates appeared on the ballot in three-quarters of the Lok Sabha's districts, attesting to a diffusion of criminality in national political life.

[Figure 3 about here]

In 2004 and again in 2009, a quarter of the legislators elected had criminal records or faced pending charges. As the data presented in Table 1 has indicated, it appears to have been politically advantageous to

have been malfeasant.

We omit from analysis candidates who are unaffiliated with any political party. The reason is that we are interested in studying when parties list criminal candidates and independent candidates self-nominate. In addition, independent candidates are not politically relevant. In 2004, 43 percent of India's legislative candidates were unaffiliated with any political party and in 2009 independent candidates rose to comprise 47 percent of all candidates. Despite their numbers, unaffiliated candidates had almost no chance of winning seats: a mere five of the $2,385$ independent candidates were elected in 2004 and in 2009 nine of $3,831$ won their seats. As a result, only about 1 percent of the Lok Sabha's 543 members are unaffiliated with any party. Unaffiliated candidates are therefore numerous but politically almost irrelevant. The median vote share collected by the winner and the first runner-up together is 87 percent, making most races effectively two candidate contests.[4] The vote share of the unaffiliated candidates was less than 1 percent in 2004 and 2009 and no independent candidate won more than 7 percent of the vote in either election.

Examining only those candidates who are put on the ballot by a political party and excluding independents, we find that the ability of criminal candidates to gain seats in the legislature is even more pronounced than for all candidates regardless of partisan affiliation. We document this with data presented in Table 1. For partisan-affiliated candidates in 2004, being charged more than doubles the rate of winning a seat, increasing it from 15 percent to 36 percent. In 2009, the difference is less marked but even so success rates among the criminals are 20 percent compared with 11 percent for non criminals. For candidates listed as affiliated with one of India's numerous political parties, we find that being charged thus proves especially electorally advantageous.

4.2 Data and Measures

4.2.1 Criminality

Candidates for the two Lok Sabha elections that we analyze were required to file sworn affidavits in which they reported criminal histories or pending criminal charges. The variable *CRIMINAL* is coded 1 if the affidavit reports charges against the candidate at any time regardless of the court's outcome and 0 otherwise. We use this as our dependent variable in the first set of specifications. In the second, we use the proportion of candidates in the district who are coded 1 on CRIMINAL out of the total number of candidates in the district; we label this variable PCRIM.

One potential objection to using charges drawn from affidavits is that criminal charges against candidates may be politically motivated rather than genuine. Skilled politicians may be charged with crimes by their rivals in efforts to discredit them politically. If this is the case, criminals could naturally be expected to enjoy an electoral advantage. This would not be due to the intimidation of voters or other attributes specifically associated with the criminal status of candidates but rather because of selection effects. Skilled politicians are more adept at winning elections. If skilled politicians are more likely to be charged with crimes, criminal candidates are more likely to win elections.

One way to refute this line of argument is to examine the specific charges against candidates and to estimate the proportion that are plausibly politically motivated. For instance, we would expect politically motivated charges to involve crimes such as libel and slander (of other candidates) as well as activities for which there is little or no direct physical or eyewitness evidence. Such a minute investigation of the pattern of charges is beyond the scope of this paper. However, a 2004 press release by the Association for Democratic Reforms (ADR) sheds light on the nature of the criminal charges against Indian politicians elected to the Fourteenth Lok Sabha. Table 2, taken from the ADR report, details some of the more serious

---

[4](Chakrabarti, Gangopadhyay and Krishnan N.d., p. 4) note that the share of the vote received by candidates not among the top two averages 17 percent, enough to unsettle the final outcome, but that is much less true if we exclude the unaffiliated, who have virtually no chance of winning the seat to begin with.

charges against legislators. The ADR finds that there were 229 charges for serious violent crimes whereas 87 charges were related to "dishonesty, cheating, fraud, forgery and dealing in stolen property." Although it is possible that some charges were politically motivated, the large proportion of serious violent crimes suggests that *CRIMINAL* is mainly picking up a true criminal element in Indian politics. It is easy to fabricate a charge of libel against a political opponent and much more difficult to fabricate a charge of murder.

[Table 2 about here]

The information displayed in Table 2 shows that even prominent political parties, including India's largest governmental party, the Indian National Congress, list criminals among their candidates. There is also large variation in the number of criminal candidates listed by the six main national political parties. The Communist Party of India has only a single legislator among its ten with a criminal record whereas the 16 legislators elected as part of the Bahujan Samaj Party's delegation face a total of 66 serious charges among them. This suggests that some parties may be more prone to selecting criminals to run for national office and that we will want to consider party affiliation in our analysis.

Finally, how the information about the criminality of legislative candidates is used within India is also relevant. The Indian press and several non-governmental organizations use the information from the affidavits to publicize the criminal element in Indian politics. An implication of this is that being a criminal is considered a potential electoral liability. Even if a candidate is unfairly charged with criminal activity for political reasons, the general public is unlikely to know this. The candidate would have to use scarce campaign resources convincing voters that the charges against him were false. For a party to deliberately select a criminal as its candidate thus carries with it fielding costs that are higher than those associated with other candidates. Given the added cost of running as a criminal, the presence on the ballot of such a large number poses a puzzle.

4.2.2 Electoral Competitiveness

We hypothesize that a party's propensity to field a criminal candidate is conditional on how uncertain it is about the electoral outcome. A party that is sure of winning or losing is not likely to field a criminal because of the fielding costs involved whereas a party that is on the cusp of winning or losing is more likely to do so.

Our measure of competitiveness, *COMP*, is the percentage difference between the share of the votes obtained by the winning candidate and the arty of the candidate in question in the district in the same election. The reason that we construct a separate measure of competitiveness for each party, rather than a single district-level measure, as is common in single-member districts, lies with the large number of parties that operate in the typical Indian electoral district. Even a third or fourth runner up will gauge the competitiveness of the race in terms of how close the candidate of his party is to winning, not by how close the runner up was. In the decision of whether to field a criminal on the ballot, we assume that each party's judgement depends on their own competitiveness position in the district.

We include four additional measures that we believe affect electoral competitiveness. These are whether the candidate is the incumbent, whether the seat in the district is reserved for a representative of a scheduled caste or tribe, whether the candidate is affiliated with a political party or is independent, whether the party of affiliation is nationally organized.

INCUMB is coded 1 if a candidate was an incumbent and zero otherwise. The standard argument is that incumbents enjoy an electorally advantage because their names are more likely to be familiar to voters and because they have had the opportunity to use government resources to consolidate their electoral hold. However, research finds that in India, legislators experience an incumbency disadvantage (Linden (2004); Uppal (2007)). This, however, is apparent only using a regression discontinuity design. Raw electoral returns show that incumbents do well in Indian elections. In 2004, 40 percent of those elected to the Lok Sabha were incumbents and in 2009 the equivalent figure rose to 53 percent. Put another way, in 2009, 51 percent of incumbents who ran again won their seat; in 2004, which featured many fewer candidates overall,

53 percent of incumbents who ran were reelected.

RESERVED is coded 1 if a district is reserved for a member of a scheduled caste or tribe. A seat is designated reserved or not by each delimitation order that sets electoral boundaries. The electoral boundaries used in the Lok Sabha elections of 2004 had been set by the Delimitation Order of 1976. In 2009, a new Delimitation Order was issued that redrew electoral boundaries. Preliminary research shows that these boundaries reflect population changes and not political bias Iyer and Shivakumar (2009). A seat is reserved on the share of population that is comprised of members of scheduled castes or tribes. There were 120 seats reserved in the Fourteenth Lok Sabha and 131 in the Fifteenth out of a total of 543 seats. Thus, nearly a quarter of the seats are now reserved.

Because caste features prominently in Indian politics, it seems reasonable to incorporate its potential impact on criminality in elections. We cannot measure it directly since information on caste is not included in the Indian census. But seeing if the selection of criminal candidates differs between reserved and non-reserved legislative seats is an indirect way to assess it.

Reserved districts may differ from other district in three important ways with respect to criminality. First are the demographics of the population resident in the jurisdictions which are reserved. Reserved seats, which by definition have a higher scheduled caste and tribe population, may also have more illiterate voters. Thus, one would expect higher rates of criminals nominated in districts whose seats are reserved. Literacy rates are lower in reserved districts — 52 percent compared to 56 percent generally. Even though we control for literacy in our estimations, this suggests that reserved districts may have other characteristics related to literacy rates that make the parties there more likely to nominate criminals, justifying our decision to code them separately.

A second difference relates to the barriers to entry in the political market. Although all voters are eligible to vote for reserved seats, only members of scheduled castes or tribes are eligible to stand as candidates. It is possible that this feature constitutes a barrier to entry, making reserved seats safer.

The third possible difference between reserved and other seats is that aspects of political competition may be different in districts with large proportions of the population in scheduled castes and tribes. Politics in these districts may be "ethnified" Banerjee and Pande (N.d.). Banerjee and Pande (N.d.) argue that candidate quality is a function of the proportion of a district's population that shares the same caste as the political party representing it. Their model predicts that candidate quality deteriorates as there is a closer match between district population characteristics and the representative's characteristics. Since reserved seats, by definition, have high shares of scheduled caste and tribe voters, their model predicts that in districts with reserved seats, low caste parties will have lower quality candidates. Their model predicts therefore that more criminals will be nominated for reserved seats.

On the basis of these three considerations, we expect that *RESERVED* will have a positive impact on the likelihood of criminals being nominated.

We also control for whether the party with which the candidate is affiliated is considered a "national" party (*NP*). According to the Election Commission of India, six national parties contested the 2004 elections. These parties were the Bharatiya Janata Party (*BJP*), the Bahujan Samaj Party (*BSP*), the Communist Party of India (*CPI*), the Communist Party of India (Marxist) (*CPM*), the Indian National Congress (*INC*), and the Nationalist Congress Party (*NCP*). In 2009, national parties are defined as these same six parties plus the Samajwadi Party, Samata Party, and the Shiv Sena. National parties are defined as those parties that run candidates in a certain number of districts and across a minimum number of states.

We include this variable because we suspect that the decision making calculus of national parties about recruiting criminals onto the ballot is different from that of the decision making calculus of non-national parties. National parties make decisions across electoral districts and are concerned with the extra-district ramifications of their candidate selection. Parties that are organized only locally or in a state or two are more parochial in their decision making calculus.

4.2.3 Fielding Costs

We explained previously why we proxy the specific fielding costs, *COST*, associated with criminal candidates with a measure of literacy. *LIT* refers to the literacy rate in each electoral district.[5] As the literacy rate increases, we expect that the likelihood of fielding criminal candidates decreases.

4.2.4 Demographic Control Variables

In addition to variables that may affect electoral competitiveness, our regressions include two control variables for which data is available at the district level. Other variables that might typically feature in models of electoral behavior, such as income, are available only at the state level for India and we therefore are not able to include them.

POP is the population of each district. Although India's electoral system is single member, electoral districts are not equally sized. We have no theoretical expectation for the sign of this variable, but we suspect that the number of voters in a district could affect information flows, party organization, and the availability of criminal candidates.

URB is the percent of the population located in urban as opposed to rural areas in the electoral district. We include this variable because we believe that urban voters, regardless of their level of literacy, have greater access to political information and are likely to be more sensitive to the criminality of elected officials. We therefore expect that the sign on this variable will be negative; with greater urbaness, expect the likelihood of a criminal to appear on the ballot to fall.

Summary statistics for the independent variables appear in Table 3. The basic electoral outcomes for 2004 and 2009 are presented in Table 4.

[Table 3 and Table 4 about here]

4.3 Analysis of the Probability a Criminal Appears on the Ballot

We employ a logistic analysis to test the effects of political competitiveness and the literacy rate on the likelihood that a criminal candidate appears on the ballot. The unit of analysis is the candidate-district. Because of changes in electoral district boundaries that took place just prior to the 2009 elections, we are not able to match districts precisely across the two elections. Our strategy for pooling the data from both elections is therefore to attach to each candidate characteristics relevant to the electoral boundaries in which he operates; population, for instance, is the estimate of total population in the district for either 2004 or 2009, as appropriate. The basic model that we estimate is:

$$CRIMINAL = \beta + \beta * COMP + \beta * LIT + \beta * COMP * LIT + \beta * CONTROLS + \epsilon$$

where *COMP* is a measure of the margin of victory of the winning candidate; *LIT* is the proportion of the population that it literate; *COMP*LIT* is the interaction; *CONTROLS* comprise *INCUM*, *RES*, *NP*, *POP*, and *URB*; and ϵ is an error term. For ease of interpretation, we report odds ratios; the results of our first set of estimations appear in Table 13.

[Table 13 about here]

Our baseline model appears as Model 1. It tests the likelihood a party will field a criminal candidate as a function of *COMP* and *LIT*, our main theoretically relevant variables. Model 2 adds the additional control variables, *INCUM*, *RES*, *NP*, *POP*, and *URB*. Model 3 adds the interaction variable of *COMP*LIT*. Model

---

[5]Data on literacy are available at the level of India's census (administrative) districts, unlike other data, such as income, which are available only at the state level. We choose not to work with data available only at the state level because it would drastically reduce the number of observations.

4, finally, adds state fixed effects to control for possible unobserved state-level heterogeneity.[6] Interpretation of the odds ratios follows standard guidelines. Results greater than 1 imply that the covariate increases the likelihood that a party fields a criminal candidate and results less than 1 imply the opposite. For each model we run three analyses. The first is for candidates who ran in 2004, the second is for 2009, and the third is a pooled analysis with both sets of observations.

In our baseline model, our principal independent variables (*LITERACY* and *COMP*) are statistically significant at the 0.01 percent level. They retain significance even with the inclusion of our statistical controls, though the inclusion of the interaction term (Model 3) reduces the level of significance for *COMP*. The inclusion of state level dummies diminishes the statistical significance for *LITERACY* to the 0.10 level in 2004, and it loses significance entirely for 2009, and in the pooled analysis. Given the difficulty of comparing logit coefficient across different models, a marginal effects table was created for the baseline model (Model 1) and Model 3 which includes controls and the interaction term. This table was created using Clarify.

[Table 14 about here]

We vary *LITERACY* and *COMP* one standard deviation below and above their means. Thus *LITERACY* varies from 42% to 66%, and *COMP* varies from 14% to 44%. All other variables are set to their mean values and in the case of dummy variables, to zero.

The substantive impact of our IVs is quite substantial. For example, the probability that a party fields a criminal candidate in a high literate district versus a low literate district is reduced by nearly 40% in the baseline model in 2004, 30% in 2009 and 34% in the pooled analysis. The relationship between *LITERACY* and the probability of fielding a charged candidate is even stronger when control variables are added for the 2004 model (52%), though weaker in 2009 (17%). *COMP* also has a substantive impact. In 2004 for the baseline model, a party that is 14% away from winning or losing is 40% more likely to field a criminal candidate than a party that is 44% away from winning or losing. The marginal effect increases to 50% in 2009. A second noteworthy result is that parties in general are more likely to field criminal candidates in 2009 than in 2004. Using Model 3, parties are about 55% more likely to field a criminal candidate in 2009 than in 2004.

Models 2 - 4 uses *RESERVED* and *NATIONAL* as control variables. In order to explore their impact on criminality, we create another marginal effects table. Here, we use Model 3 as our estimating model and set all covariate values to their mean, or to zero for dummy variables.

[Table 15 about here]

Parties competing in reserved districts are far less likely to field a criminal candidate. We estimate that in 2004, a party was 33% less likely to field a criminal candidates and 37% less likely in 2009. National parties were *more* likely to field a criminal candidate in 2009, in contrast to 2004 when being a national party made no difference. In fact, they were nearly 33% more likely to field a charged candidate in 2009. We are currently unsure as to why we see such differences across one election period.

Finally, our interaction term is not statistically significant. We will further study this in the future. Include that interaction not significant

4.4   Analyses of the Proportion of Criminals Listed on the Ballot

[This section yet to be completed.]

---

[6]The coefficients for state effects are not included in the tables reporting our results. Note that when state dummies are included, states without variation on the dependent variable drop out of the analysis. This implies that if all the candidates in all the districts of a state are non-criminals, the state is dropped.

5 Regression Discontinuity Analyses

We now turn our attention to the role criminality may play in contributing to the incumbency disadvantage that has been identified in India (Linden (2004); Uppal (2007)). In particular, we study two issues. First, given that criminal candidates have additional resources unavailable to non criminals, do they experience an equivalent incumbency disadvantage? Second, can the Indian incumbency disadvantage be explained by the presence of incumbents facing criminal opponents?

We examine these questions using regression discontinuity techniques and assessing the extent of incumbency (dis)advantage for different groups of candidates. First, we conduct a baseline analysis on all candidates and calculate the incumbency disadvantage in the 2009 elections. Second, we conduct an analysis for criminal and non criminal candidates separately. Third, we conduct separate analyses for candidates facing a criminal opponent and for those who do not. Finally, for those candidates who do face a criminal opponent in 2009, we conduct an analysis for criminal and non criminal candidates.

We find that the criminality does play a role in the electoral disadvantage experienced by incumbents in 2009. Incumbents who face a criminal opponent are less likely to be elected than their non-incumbent counterparts, though criminal incumbents do not face an incumbency disadvantage against criminal opponents. Tests of statistical significance have not yet been calculated and our results are still preliminary.

5.1 Regression Discontinuity Design

The regression discontinuity design has been used to assess the existence and extent of incumbency (dis)advantage. Lee (2008) estimates the level of incumbency advantage for incumbent parties in the U.S. House of Representatives and finds that incumbent parties are 40 to 45 percent more likely to win an election. Using similar analytic techniques, an incumbency disadvantage has been identified in India. Linden (2004) estimates that since 1991 incumbents in the Lok Sabha are 14 percent less likely to win, and Uppal (2007) estimates that since 1991 incumbents in state assemblies are 25 percentage points less likely to win reelection.

Generally speaking, a regression discontinuity analysis compares the probability of winning an election across two groups: candidates who barely won office in the previous election and candidates who barely lost. The underlying logic is that whether a candidate barely wins or barely loses an election is essentially random and as such, candidates who are close to winning or losing are otherwise identical. This approach controls for any observed or unobserved characteristics of incumbents that may bias the impact of incumbency on election rates when they are estimated using a regression-based empirical strategy.

There are three main steps in using this approach. First, we must estimate the probability that a candidate wins an election. We do this by using a logistic regression that estimates the probability that a candidate wins the election in 2009 as a function of the candidate's 2004 margin of victory. We do this for incumbents and non incumbents separately. The margin of victory of candidates who won in 2004 is calculated by subtracting the vote share for the second place candidate; i.e. we use the same formula as for COMP, studied above. The margin for those who lost in 2004 is calculated by subtracting the vote share of the winning candidate from the share of the near-loser.[7]

Second, we must restrict our sample space to ensure that the two groups of candidates (incumbents and non-incumbents) are equivalent. There is a trade-off between sample size and group comparability. If we set the margin of victory cutoffs too narrowly, then the likelihood of having comparable groups is high but at the cost of a small sample size. Conversely, if we set our margin of victory cutoffs too widely, then the likelihood of having comparable groups is lower (potentially biasing the results) but we would have a larger sample size. Since our analysis only covers one election period, it is particularly sensitive to sample size

---

[7]In contrast to our approach, Linden (2004) uses a non-parametric design and Uppal (2007) utilizes logistic regression with a fourth order polynomial for margin of victory as well as their interactions and fixed effects.

problems. Though this introduces bias, we show later that the direction of the bias is towards overestimating an incumbency advantage.

Finally, we must estimate the discontinuity between near-winners and near-losers. Graphically, we examine the impact of incumbency on the probability of being elected in 2009 by estimating the gap between the probability of being elected for candidates who lost in 2004 (those with a margin of victory less than 0) and the probability of being elected for candidates who won in 2004 (those with a margin of victory above 0). These election probabilities are calculated by running two logit regressions. The first estimates the probability of winning a seat in 2009 for candidates who just won in 2004 and the second estimates the probability of winning a sear for candidates who just lost in 2004.

In order to calculate the size of the gap, we subtract the probability of being elected in 2009 for incumbents from the probability of being elected in 2009 for non-incumbents. If incumbency has no impact, there is no gap between the probabilities of the two groups. If there is a negative effect, then the line representing the probability of being elected in 2009 will be "lower" for incumbents than non-incumbents. We calculate the gap at discontinuity, or in this case, where the margin of victory is zero.

5.2 Data

Of the approximately 8,000 candidates who ran for the Lok Sabha in 2009, 572 had also run in 2004. Of these 572, we restrict our attention to candidates within 25 percent of winning or losing in 2004, thereby reducing our pool of candidates to 398. Tables 5 through 11 compare incumbents and non-incumbents in a variety of ways: with respect to the probability of winning in 2009; with respect to their vote shares in 2004 and 2009; share of the candidates that have criminal charges; share of the candidates that faced a charged candidate in 2009; and the proportion of candidates that were members of the Congress or BJP party.

Bias-free results of a regression discontinuity analysis require that both groups be balanced on observed as well as unobserved traits. Since in order to obtain large enough samples we set a wide margin of victory cutoff (25 percent), our groups are not comparable along certain important dimensions. For example, in all the analyses that we perform, incumbents have a higher vote share in 2004 than non-incumbents. This implies that incumbents as a whole had better political skills than their non-incumbent counterparts. The main implication for our analysis is that results may be skewed in favor of incumbents. Thus results that show an incumbency advantage need to be viewed with caution whereas results that show an incumbency disadvantage may understate the true level of this disadvantage.

5.3 Results

Table 12 summarizes the results of our analyses and Figures 4 through **??** graphically reproduce them. The results of our baseline analysis are consistent with those reported by previous studies. We find that incumbents are at an electoral disadvantage in the Indian legislature. They are at 6 percentage point disadvantage for winning office in 2009, which translates into being 19 percent less likely to win.[8] We find that criminal candidates and their non-criminal counterparts face incumbency disadvantage, and with roughly the same magnitude. Non-criminal candidates are 8 percentage points less likely to win in 2009 and criminal candidates are 6 percentage points less likely to win.

[Tables 5–12 and Figures 4–10about here]

Figure 6 reveals a surprising, and puzzling relationship between the margin of victory for non-incumbents and the probability of winning in 2009. Although we expect there to be a positive relationship between the

---

[8]Our results do not exactly match those reported by Linden (2004) or citeuppal07 but our analysis differs in three important respects. First, unlike Linden (2004), we use a parametric approach and estimate a logit function. Second, unlike Uppal (2007), we do not use control variables in our estimation strategy. Instead, we only use margin of victory in 2004 as an independent variable. Finally, by using a wider margin of victory cutoff than either prior study, or than Clots-Figueras (2005), who also performs a regression discontinuity analysis using Indian data, our results understate the incumbency disadvantage.

two, our results show a negative relationship. This negative relationship may be due to a misspecification of our logit function. Possibly either a non-parametric design or more control variables are needed.

Figures 7 8 represent the main results of the regression discontinuity analysis. In 2009, candidates not facing a criminal opponent have an incumbency advantage (5 percentage points) whereas candidates facing a criminal opponent have an incumbency disadvantage (9 percentage points). The analysis depicted in Figures 9 and refdiscon-fig7 confirms our intuition that criminal candidates would fare better against criminal opponents in comparison to their non-criminal counterparts. Non-criminal candidates have a 17 percentage point incumbency disadvantage when facing criminal opponents whereas criminal candidates have a 6 percentage point incumbency advantage. These results show that the disadvantage of the Indian legislative incumbent is systematically related to the likelihood that the incumbent squares off electorally against a criminal. That fully three-quarters of India's electoral districts saw at least one criminal on the ballot in 2009, representing a massive diffusion of candidate criminality over 2004 when only half of the country's electoral districts witnessed criminal candidates, raises the suspicion that criminals are driving out non-criminals in the Indian polity.

We do not calculate test of statistical significance. The simulation results indicate wide confidence intervals for each estimate, which may imply that our results do not have enough power to pass statistical scrutiny. We intend to explore this in the future.

Finally, we note that our analysis may have misspecification problems. One finding that raises this possibilty is the negative relationship between margin of victory and the probability of winning in 2009 for criminal non-incumbents. This is the reverse of what we would expect: the margin of victory and the probability of winning should be positively related. We need to consider whether the functional form of the relationship between electoral performance and winning differs for non-criminals and criminals. If so, this could be due to unobserved attributes of criminal candidates or to unobserved attributes of districts that criminal candidates run in.

We are also concerned about misspecification because of a paradoxical result concerning criminal candidates. Overall, criminal incumbents exhibit an incumbency disadvantage yet criminal incumbents facing an opponent with a criminal record have an incumbency advantage. If both results are true, then criminal incumbents have an electoral disadvantage when facing non-criminal opponents. This in turn contradicts our initial finding that criminal candidates are more likely to win than non-criminal candidates. Again, it is possible that there is reason for this, but it is also possible that our empirical estimation strategy needs additional calibration. Future analyses may require the use of non-parametric estimation methods or the use of control variables in the logisitic regression.

5.4 Discussion

Our preceding analysis presents suggestive evidence that the well-known incumbency disadvantage in India is due to the presence of criminal candidates. Our results demonstrate not only that criminality is an asset in Indian politics but that criminality trumps incumbency. This is truly remarkable. This means that incumbents have either failed to deliver benefits to constituents sufficient to warrant reelection or that they lack the resources to mount effective electoral campaigns against criminal opponents. In either case, the benefits of being in office do not counteract the electoral pressures that criminality brings to Indian politics. This is especially troubling since theories of democratic accountability start with the assumption that politicians are office seeking individuals. If electoral pressures and competitiveness result in conditions that favor criminal candidates and if criminal candidates have different policy preferences once in office, then the ability of political competition to ensure normatively good outcomes is completely compromised.

Linden (2004) notes that the rise of the incumbency disadvantage in India begins with the demise of the political dominance of the Congress Party. Chhibber (1999) argues Congresss initial dominance was based on clientelistic arrangements between the party and regional elites. The demise of Congress's dominance

was the result of the rising importance of caste based politics, the erosion of the old clientelistic networks, and increasing political competition throughout India. As clientelistic relationships have their effectiveness in the electoral arena, the increase in political competitiveness makes finding new ways to target and mobilize voters even more important. The increase in political competitiveness appears to have resulted in the political incorporation of unsavory elements of the Indian political economy. Further research needs to be done to clarify the link between criminal candidates and their capacity to mobilize votes.

The importance of incumbency is apparent in seniority based systems like the United States. There, the ability of legislative representatives to deliver benefits to voters is not only a function of whether or not their party is in power but also a function of their length of tenure in Congress. The ability to deliver benefits in a Westminster political institution is a function of being in government and thus measuring incumbency at the party level would be more appropriate. Doing such an analysis between the 2004 and 2009 election period is difficult since India underwent a new delimitation between elections.

6 Conclusions

In this preliminary analysis, we have shown that criminal candidates are more likely to appear on the ballot in India in districts with more illiterate voters and where electoral competitiveness is higher. We interpret these results in light of the theory laid out: parties risk placing criminals on the ballot only when they think it might change the outcome of the election; and the cost of fielding a criminal candidate makes this more likely where there are more illiterate voters. Our main results were consistent across various specifications and corroborated our hypotheses. The main hypothesis that thus far our results do not corroborate involves the interaction of competitiveness and the costs of fielding a criminal. We anticipated that the interaction would be significant, but our preliminary results were not. We expect to employ additional techniques of analysis to further explore this.

References

Ai, Chunrong and Edward C. Norton. 2003. "Interaction Terms in Logit and Probit Models." *Economics Letters* 80:123–29.

Banerjee, Abhijit, Donald Green, Jennifer Green and Rohini Pande. 2009. "Can Voters be Primed to Choose Better Legislators? Evidence from Two Field Experiments in Rural India." Unpublished paper.

Banerjee, Abhijit V. and Rohini Pande. N.d. "Parochial Politics: Ethnic Preferences and Political Corruption." Unpublished paper.

Berenschot, Ward. 2008. Moneypower and Musclepower in a Gujarati Locality: On the Usefulness of *Goondas* in Indian Politics. ASSR Working Paper 08/04 Amsterdam School for Social Science Research, University of Amsterdam.

Besley, Timothy and Robin Burgess. 2002. "The Political Economy of Government Responsivenss: Theory and Evidence from India." *Quarterly Journal of Economics* 117(4):1415–51.

Bobonis, Gustavo J. and Luis R. Cámara Fuertes. 2009. "Does Exposing Corrupt Politicians Reduce Corruption?" Unpublished paper, University of Toronto.

Brollo, Fernada. 2009. "Who is Punishing Corrupt Politicians — Voters or the Central Government? Evidence from the Brazilian Anti-Corruption Program." Unpublished paper, Bocconi University.

Chakrabarti, Rajesh, Shubhashis Gangopadhyay and Shagun Krishnan. N.d. "Incumbency Effects in Indian National Elections — A Preliminary Exploration." Unpublished paper.

Chang, Eric C.C., Miriam A. Golden and Seth J. Hill. 2010. "Legislative Malfeasance and Political Accountability." *World Politics* 62(2):177–220.

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

Chhibber, Pradeep. 1999. *Democracy Without Associations: Transformation of the Party System and Social Cleavages in India*. Ann Arbor: University of Michigan.

Clots-Figueras, Irma. 2005. Women in Politics: Evidence from the Indian States. Political Economy and Public Policy Series PEEP/14 Suntory Centre, London School of Economics and Political Science.

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

Iyer, Lakshmi and Maya Shivakumar. 2009. "Electoral Redistricting in the World's Largest Democracy: An Evaluation." Unpublished paper, Harvard Business School.

Konrad, Kai A. 2009. *Strategy and Dynamics in Contests*. Oxford: Oxford University Press.

Lee, David S. 2008. "Randomized Experiments from Non-Random Selection in U.S. House Elections." *Journal of Econometrics* 142(2):675–97.

Linden, Leigh L. 2004. "Are Incumbents Really Advantaged? The Preference for Non-Incumbents in Indian National Elections." Unpublished paper.

Nooruddin, Irfan and Pradeep Chhibber. 2007. "Unstable Politics: Fiscal Space and Electoral Volatility in the Indian States." *Comparative Political Studies* 41(8):1069–91.

Nyblade, Benjamin and Steven R. Reed. 2008. "Who Cheats? Who Loots? Political Competition and Corruption in Japan, 1947–1993." *American Journal of Political Science* 54(4):926–41.

Persson, Torsten and Guido Tabellini. 2000. *Political Economics: Explaining Economic Policy*. MIT Press.

Reinikka, Ritva and Jakob Svensson. 2005. "Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda." *Journal of the European Economic Association* 3(2-3):259–267.

Sen, Amartya. 1981. *Poverty and Famines: An Essay on Entitlement and Deprivation*. Oxford: Oxford University Press.

Snyder, James. 1989. "Election Goals and the Allocation of Campaign Resources." *The Economist* 57(3):637–60.

Uppal, Yogesh. 2007. "The Disadvantaged Incumbents: Estimating Incumbency Effects in Indian State Legislatures." Unpublished paper.
URL: *http://mpra.ub.uni-muenchen.de/8515/*

Viashnav, Milan. 2010. "The Market for Criminality: Money, Muscle and Elections in India." Unpublished paper, Columbia University.

Wilkinson, Steven I. 2007. Explaining Changing Patterns of Party-Voter Linkages in India. In *Patrons, Clients, and Policies: Patterns of Democratic Accountability and Political Competition*, ed. Herbert Kitschelt and Steve I. Wilkinson. Cambridge: Cambridge.

Appendix A: Data and Data Sources

Data units are of four types:

1. National elections occur in India's 543 parliamentary constituencies.
2. Demographic information from the 2001 Indian Census is collected in 593 administrative districts. The administrative districts do not coincide neatly with the parliamentary constituencies.
3. India's 35 states and territories hold elections across 4,140 assembly constituencies.
4. The assembly constituencies aggregate perfectly to the parliamentary constituencies.

To use demographic information at the level of the parliamentary constituencies, we estimate values based on parliamentary voter-weighted values aggregated from administrative districts. For additional details, see below, under *population*.

Caste information is provided by the 2001 Indian census. The information available refers to numbers of persons who are members of scheduled castes and tribes. Data are not available for Manipur and Nagaland. Caste information downloaded from `http:/www.indiastat.com` and `http:/censusindia.gov.in` in April 2008.

Indian Census 2001 . Many demographic variables below are taken from the 2001 Indian census, available as CensusInfo [electronic resource], India 2001 (Office of the Registrar General, New Dehli, India), Version 1.0.

Criminal charges are based on sworn affidavits provided by all parliamentary candidates to the Lok Sabha in the 2004 elections. The affidavits contain the candidate name and party affiliation organized by state and parliamentary constituency. Downloaded from the Liberty Institute, `http:/www.empoweringindia.org/new/home.aspx`.

Electoral returns from 2004 and 2009 are available from the Electoral Commission of India. Information includes the number of votes received by each candidate where candidates are organized by state, parliamentary constituency, and party. We matched the returns for each candidate with the Liberty Institute's coding on whether the candidate's affidavit reported pending criminal charges. Election results downloaded from `http:/eci.nic.in/StatisticalReports/ElectionStatistics.asp`.

Incumbents are coded 1 and non-incumbent candidates are coded 0. Incumbency status as of 2004 coded by matching the names of the winners in the 1999 Lok Sabha elections and winners in any subsequent bye-elections with the names of candidates in 2004. Prior election results from the website of the Electoral Commission of India.

Independent vote shares were calculated for both 1999 and 2004. Refers to the share of the vote in each district that was won by candidates with no partisan affiliation. Our calculations based on data from electoral results.

Literacy rates are calculated using 2001 Indian Census (see above); we divided the values of the number of illiterate persons by the total population.

National parties are parties that have met the requirements to be designated as national by the Election Commission of India. A national party must be a state party in four or more states. In 2004, the six national parties were the BJP, the BSP, the CPI, and CMP, INC, and NCP. In 2009, the Nationalist Congress Party lost its designation as a national party and the Samajwadi Party, Samata Party, and the

Shiv Sena were all designated national parties in addition to the other five parties that had enjoyed national status in 2004.

Reserved seats are coded 1 and regular parliamentary seats coded 0. In the Fourteenth Lok Sabha, there are 79 seats reserved for scheduled castes and 41 for scheduled tribes. In the Fifteenth Lok Sabha, 84 seats are reserved for scheduled castes and 47 for schedeuled tribes. i Seats are designated for reservation by delimitation orders based on the population of scheduled caste and tribes. The 2004 elections were conducted under the Delimitation Order of 1976 and the 2009 elections under the Delimitation Order of 2009; the latter was based on 2001 census data.

Population in each parliamentary constituency was calculated using data from the 2001 Indian Census. Population data is available at the level of administrative districts. We estimate population for parliamentary constituencies in two steps. First we estimate population totals for the state assembly constituencies on the basis of the fraction of votes cast in each state assembly constituency out of the total number of votes cast in the corresponding administrative district. (Parliamentary vote totals are available from the Election Commission of India at the smaller level of administrative districts.) We then aggregate the estimated population totals from assembly constituencies to parliamentary constituencies. Similar manipulations are performed for all variables drawn from the Indian census.

Total votes cast in the 2004 and 2009 Lok Sabha elections for each parliamentary constituency from the Election Commission of India.

Urban is the percent of the population that lives in urban areas. We calculated this by taking the urban population in each administrative district and dividing by the distric's total population. Data available from India Census 2001.

Vote shares for each candidate in 2004 and 2009 are calculated by dividing the number of votes each candidate receives by the vote totals received by all candidates in each parliamentary constituency and multiplying by 100. We calculate the total votes cast in each parliamentary constituency by adding up the votes for all candidates listed by the Election Commission of India in each parliamentary constituency.

Matching 2004 and 2009 parliamentary constituencies with 2001 census districts occurred as follows. For 2004, the Delimitation Order of 1976 was used to match administrative districts to state assembly districts. Second, each assembly district's population figures from the census is estimated as a fraction of the population in the corresponding administrative district. Each assembly district's fraction was determined by its fraction of votes cast in the 2004 election of the entire administrative district. Third, since state assembly districts are perfect subsets of parliamentary constituencies, population estimates are aggregated up to the parliamentary constituency level.

Some administrative districts were partitioned after 1976. The following parliamentary constituencies were reaggregated to their 1976 boundaries: in the state of Karnataka, Bagalkot was reaggregated to Bijapur, Koppal was reaggregated to Raichur, Gadag and Haveri were reaggregated to Dharwad, Davangere was reaggregated to Chitradurga, Bangalore Rural was reaggregated to Bandalore, Udupi was reaggregated to Dakshina Kannada, and Chamarajnagar was reaggregated to Mysore. In the state of Rajasthan, Bharapur was reaggregated to Dhaulpur, Hanamungarh was reaggregared to Ganganagar, Dausa was reaggregated to Jaipur, Baran was reaggregated to Kota, Kkarauli was reaggregated to Sawai Madhopur, and Rajsamand was reaggregated to Udaipur. For the state of Tamil Nadu, Ariyalur was reaggregared to Perambular. In the union territories, Andaman and Nicobar were combined as were Daman and Diu.

For the 2009 elections, we replicated the above procedures using the new electoral districts.

Appendix B: Alternative Specification of the Formal Model

This appendix briefly describes an alternative specification of the model where there are only two types of candidates, criminal (C) or non-criminal (H). This formulation allows one to make predictions about the probability of fielding a criminal candidate and about the number of criminal candidates fielded in a given district. The down side is that the analysis is more complex, involving mixed strategy equilibria and the interaction effects are harder to draw out and interpret.

Suppose that there are only two types of candidates, criminal (C) and non-criminal (H) candidates i.e., $t \in \{t_C, t_H\}$ with $t_C > t_H$. In this case, the candidate selection game can be represented as a matrix game with four possible pure strategy equilibrium outcomes. The payoffs associated with the four configurations are shown in the table below. The vote production function is $f(t) = t$ and the win probability is given by equation (1). We have normalized the vote productivity of the criminal type to $t_C = 1$ and denote $t_H = t < 1$ the productivity of the non-criminal type. The first entry in each cell represents the payoff of party A while the second is that of party B.

A/B	Criminal	Non-criminal
Criminal	$a - \mu$ $1 - a - \mu$	$\frac{a}{a+(1-a)t} - \mu$ $\frac{(1-a)t}{a+(1-a)t}$
Non-criminal	$\frac{at}{at+(1-a)}$ $\frac{1-a}{at+(1-a)} - \mu$	a $1 - a$

This game has two pure strategy Nash equilibria and one equilibrium in mixed strategies. For expositional purposes, let us assume that $a \geq \frac{1}{2}$ so that party A enjoys a natural advantage in the race. Define the following two thresholds:

$$T_1(a) \equiv a - \frac{at}{at + (1 - a)}$$

$$T_2(a) \equiv 1 - a - \frac{(1 - a)\,t}{a + (1 - a)\,t}$$

with $T_1(a) \geq T_2(a) > 0$ for $a \geq \frac{1}{2}$. Threshold $T_2(a)$ decreasing in a while $T_1(a)$ is increasing in a for a close to $\frac{1}{2}$. We can then summarize the outcome of the game as follows.

1. For $\mu > T_1(a)$, both parties field an honest candidate for sure;

2. For $\mu \in [T_2(a), T_1(a)]$, the two parties randomize and field a criminal with positive probability. The outcome may be that both, neither, or one of the two parties actually fields a criminal. The probability that party A selects a criminal is

$$p_A(C) = \frac{(\mu - T_1)\,(at + 1 - a)\,(a + t(1 - a))}{(t - 1)^2\,(1 - a)\,(1 - 2a)\,a}$$

   while the probability that party B selects a criminal is

$$p_B(C) = \frac{(\mu - T_2)\,(at + 1 - a)\,(t + a\,(1 - t))}{(t - 1)^2\,(1 - a)\,(2a - 1)\,a}$$

3. For $\mu < T_2(a)$, the two parties field a criminal candidate for sure.

Intuitively, we see that the expected number of criminals fielded in each race decreases with the cost of fielding criminals. Likewise, a reduction in the electoral advantage of criminals (t increases towards 1) reduces the expected number of criminals fielded (T_1 and T_2 are both decreasing in t).

The impact of competitiveness is more complex. To see this, let us compare a close race to a race in which party A has a clear advantage. A decrease in a towards $\frac{1}{2}$, which makes the race more competitive, shifts threshold T_2 up. As a consequence, for a given cost of fielding, it is more likely that both parties field criminals. The effect on T_1, however, is ambiguous. For a close to $\frac{1}{2}$, T_1 increases in a (the opposite is true for a close to 1). Thus, starting from an a close to $\frac{1}{2}$ an increase in a, that makes the race less competitive, means that, on the one hand, it is less likely that both parties field a criminal for sure, and, on the other hand, it also becomes less likely that both field non-criminal candidates. In other words, the region with the mixed strategy equilibrium "eats" into the regions of the two pure strategy equilibria and the prediction regarding competitiveness and fielding criminals becomes more blurred.

Table 1: Criminality and Electoral Outcomes, 14th and 15th Lok Sabha

	Number of Candidates	Number Elected	Percent Elected
All candidates 2004			
Not criminals	4,960	415	8.37
Criminals	475	128	26.95
Total	5,435	543	9.99
All candidates 2009			
Not criminals	7,177	414	5.77
Criminals	893	129	14.45
Total	8,070	543	6.73
Party-affiliated candidates 2004			
Not criminals	2,691	410	15.24
Criminals	359	128	35.65
Total	3,050	538	17.64
Party-affiliated candidates 2009			
Not criminals	3,596	406	11.29
Criminals	643	128	19.91
Total	4,239	534	12.60
Independent candidates 2004			
Not criminals	2,269	5	.22
Criminals	116	0	0
Total	2,385	5	.21
Independent candidates 2009			
Not criminals	3,581	8	.22
Criminals	250	1	.4
Total	3,831	9	.23

Table 2: Serious Criminal Charges Against MPs by Party, Fourteenth Lok Sabha

Party	BJP	INC	CPM	CPI	BSP	NCP	Other	Total
Murder, attempted murder, etc.	7	4	2	1	17	0	56	84
Robbery	0	4	0	0	8	0	5	17
Kidnapping	0	1	0	0	2	0	9	11
Theft and extortion	1	0	0	0	3	0	24	28
Rape	0	0	0	0	0	0	0	1
Other violent crimes	9	7	3	0	13	2	54	88
Total violent crimes	17	16	5	1	43	2	149	229
Dishonesty, cheating, fraud, forgery, dealing in stolen property	5	17	6	0	23	0	36	87
False oaths	5	4	0	0	0	0	7	16
Defiling place of worship	1	0	0	0	0	0	0	1
Total other serious crimes	11	21	6	0	23	0	43	104
Total all crimes	28	37	11	1	66	2	192	333

Notes: Adapted from "Lok Sabha Elections: Press Release July 21, 2008," issued by the Association for Democratic Reforms (ADR) and partner NGOs from All India Election Watch Network; downloaded from `www.adrindia.org/downloads/LokSabha_High_Level_Analysis.doc`. Data refer to the number of crimes committed not number of MPs charged. BJP: Bharatiya Janata Party. INC: Indian National Congress. CPM: Communist Party of India (Marxist). CPI: Communist Party of India. BSP: Bahajan Samaj Party. NCP: Nationalist Congress Party.

Table 3: Summary Statistics for Independent Variables

	Mean	SD	Min.	Max.
Electoral year 2004				
COMP	37.87	16.94	0.06	72.95
INCUM	.0765409	.2658861	0	1
RES	.1762649	.3810805	0	1
NP	.2485741	.4322261	0	1
LIT	55.00925	12.29996	25.8629	85.42501
POP	1949653	538711.1	60595	5410783
URB	29.84113	22.34279	3.484043	100
Electoral year 2009				
COMP	37.03	14.52	0.04	78.24
INCUM	.0354399	.1849004	0	1
RES	.3095415	.4623333	0	1
NP	.2008674	.4006739	0	1
LIT	55.0358	12.14263	25.42116	85.293
POP	1928146	371871.3	186189	4013609
URB	29.85517	21.69144	3.484043	100

Notes: COMP is percentage margin of victory; INCUM is whether the candidate is the incumbent; RES is reserved seat; NP is affiliation with national party; LIT is percent of population that is literate; POP is total population; URB is percent of total population in urban areas. COMP, RES, LIT, POP, and URB measured at the level of the electoral district. INCUM and NP measured at the level of the candidate.

Table 4: Fourteenth and Fifteenth Lok Sabha Electoral Results

Party	No. of Candidates	No. of Seats Won	Percent Winning	Percent of Lok Sabha
2004 Electoral Results				
BJP	364	138	37.9	25.4
BSP	435	19	4.4	3.5
CPI	34	10	29.4	1.8
CPM	69	43	62.3	7.9
INC	417	145	34.8	26.7
NCP	32	9	28.1	1.7
Other	1,699	174	10.2	32.0
Total	3,050	538	17.6	99.1
2009 Electoral Results				

Table 5: Comparison of Candidate Characteristics (2009 Lok Sabha)

	Won in 2004	Lost in 2004	Difference
Proportion winning in 2009	0.45	0.33	0.12**
	(0.03)	(0.04)	(2.28)
Vote Share 2009	0.37	0.29	0.07***
	(0.01)	(0.01)	(5.41)
Vote Share 2004	0.46	0.33	0.13***
	(0.00)	(0.01)	(14.69)
Criminal Charge	0.23	0.16	0.07
	(0.03)	(0.03)	(1.60)
Faced charged opponent 2009	0.68	0.73	-0.05
	(0.03)	(0.04)	(-1.13)
Proportion of INC Candidates	0.34	0.34	0.00
	(0.03)	(0.04)	(0.04)
Proportion of BJP Candidates	0.24	0.28	-0.04
	(0.03)	(0.04)	(0.91)
Number of observations	253	145	

Notes: Standard errors in parenthesis except for the Differences column, which has t-stats in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6: Comparison of Characteristics of Candidates With No Criminal Charge (2009 Lok Sabha)

	Won in 2004	Lost in 2004	Difference
Proportion winning in 2009	0.47	0.32	0.15***
	(0.04)	(0.04)	(2.59)
Vote Share 2009	0.37	0.29	0.08***
	(0.01)	(0.01)	(5.63)
Vote Share 2004	0.46	0.34	0.13***
	(0.01)	(0.01)	(12.83)
Criminal Charge	0.00	0.00	0.00
	(0.00)	(0.00)	(.)
Faced charged opponent 2009	0.66	0.70	-0.04
	(0.03)	(0.04)	(-0.69)
Proportion of INC Candidates	0.38	0.37	0.01
	(0.03)	(0.04)	(0.17)
Proportion of BJP Candidates	0.24	0.27	-0.04
	(0.03)	(0.04)	(-0.71)
Number of observations	121	194	

Notes: Standard errors in parenthesis except for the Differences column, which has t-stats in parenthesis.
*** p<0.01, ** p<0.05, * p<0.1.

Table 7: Comparison of Characteristics of Candidates with a Criminal Charge (2009 Lok Sabha)

	Won in 2004	Lost in 2004	Difference
Proportion winning in 2009	0.41	0.42	-0.01
	(0.06)	(0.10)	(-0.08)
Vote Share 2009	0.33	0.29	0.04
	(0.02)	(0.03)	(1.26)
Vote Share 2004	0.45	0.29	0.15***
	(0.01)	(0.02)	(7.69)
Criminal Charge	1.00	1.00	0.00
	(0.00)	(0.00)	(.)
Faced charged opponent 2009	0.75	0.92	-0.17
	(0.06)	(0.06)	(-1.76)
Proportion of INC Candidates	0.19	0.17	(0.02)
	(0.05)	(0.08)	(0.21)
Proportion of BJP Candidates	0.25	0.33	-0.08
	(0.06)	(0.10)	(-0.72)
Number of observations	59	24	

Notes: Standard errors in parenthesis except for the Differences column, which has t-stats in parenthesis.
*** p<0.01, ** p<0.05, * p<0.1.

Table 8: Comparison of Characteristics of Candidates not Facing a Criminal Opponent (2009 Lok Sabha)

	Won in 2004	Lost in 2004	Difference
Proportion winning in 2009	0.49	0.24	0.25***
	(0.06)	(0.07)	(2.64)
Vote Share 2009	0.40	0.29	0.11***
	(0.01)	(0.03)	(4.49)
Vote Share 2004	0.47	0.35	0.13***
	(0.01)	(0.01)	(8.82)
Criminal Charge	0.19	0.05	0.13*
	(0.04)	(0.04)	(1.96)
Faced charged opponent 2009	0.00	0.00	0
	(0.00)	(0.00)	(.)
Proportion of INC Candidates	0.38	0.26	0.11
	(0.05)	(0.07)	(1.19)
Proportion of BJP Candidates	0.30	0.39	-0.09
	(0.05)	(0.08)	(-1.02)
Number of observations	80	38	

Notes: Standard errors in parenthesis except for the Differences column, which has t-stats in parenthesis.
*** p<0.01, ** p<0.05, * p<0.1.

Table 9: Comparison of Characteristics of Candidates Facing a Criminal Opponent (2009 Lok Sabha)

	Won in 2004	Lost in 2004	Difference
Proportion winning in 2009	0.44	0.37	0.07
	(0.04)	(0.05)	(1.08)
Vote Share 2009	0.35	0.29	0.06***
	(0.01)	(0.01)	(3.51)
Vote Share 2004	0.45	0.32	0.13***
	(0.01)	(0.01)	(11.83)
Criminal Charge	0.25	0.21	0.05
	(0.03)	(0.04)	(0.93)
Faced charged opponent 2009	1.00	1.00	0.00
	(0.00)	(0.00)	(.)
Proportion of INC Candidates	0.32	0.36	-0.05
	(0.04)	(0.05)	(-0.80)
Proportion of BJP Candidates	0.21	0.24	-0.03
	(0.03)	(0.04)	(-.057)
Number of observations	173	107	

Notes: Standard errors in parenthesis except for the Differences column, which has t-stats in parenthesis.
*** p<0.01, ** p<0.05, * p<0.1.

Table 10: Comparison of Characteristics of Candidates without a Criminal Charge Facing a Criminal Opponent (2009 Lok Sabha)

	Won in 2004	Lost in 2004	Difference
Proportion winning in 2009	0.46	0.36	0.09
	(0.04)	(0.05)	(1.34)
Vote Share 2009	0.36	0.29	0.07***
	(0.01)	(0.02)	(3.73)
Vote Share 2004	0.46	0.33	0.12***
	(0.01)	(0.01)	(9.87)
Criminal Charge	0.00	0.00	0.00
	(0.00)	(0.00)	(.)
Faced charged opponent 2009	1.00	1.00	(0.00)
	(0.00)	(0.00)	(.)
Proportion of INC Candidates	0.36	0.41	-0.06
	(0.04)	(0.05)	(-0.81)
Proportion of BJP Candidates	0.22	0.21	0.01
	(0.04)	(0.04)	(0.22)
Number of observations	129	85	

Notes: Standard errors in parenthesis except for the Differences column, which has t-stats in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 11: Comparison of Characteristics of Candidates with a Criminal Charge Facing a Criminal Opponent (2009 Lok Sabha)

	Won in 2004	Lost in 2004	Difference
Proportion winning in 2009	0.39	0.41	-0.02
	(0.07)	(0.11)	(-0.18)
Vote Share 2009	0.31	0.29	0.02
	(0.02)	(0.03)	(0.70)
Vote Share 2004	0.45	0.29	0.15***
	(0.01)	(0.02)	(6.92)
Criminal Charge	1.00	1.00	0.00
	(0.00)	(0.00)	(.)
Faced charged opponent 2009	1.00	1.00	0.00
	(0.00)	(0.00)	(.)
Proportion of INC Candidates	0.20	0.18	0.02
	(0.06)	(0.08)	(0.22)
Proportion of BJP Candidates	0.18	0.36	-0.18
	(0.06)	(0.10)	(1.63)
Number of observations	44	22	

Notes: Standard errors in parenthesis except for the Differences column, which has t-stats in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table 12: Summary of Regression Discontinuity Results

	2004 Incumbents	2004 Non Incumbents	Difference	% Difference
All observations	0.31	0.37	-0.06	-19
Non charged candidates	0.33	0.41	-0.08	-24
Charged candidates	0.28	0.34	-0.06	-21
Did not face charged opp.	0.37	0.32	0.05	14
Did face a charged opp.	0.29	0.38	-0.09	-31
Non charged candidate with a charged opponent	0.29	0.46	-0.17	-59
Charged candidate with a charged opponent	0.32	0.26	0.06	19

Figure 1: Degree of Competition and Equilibrium Candidate Choice
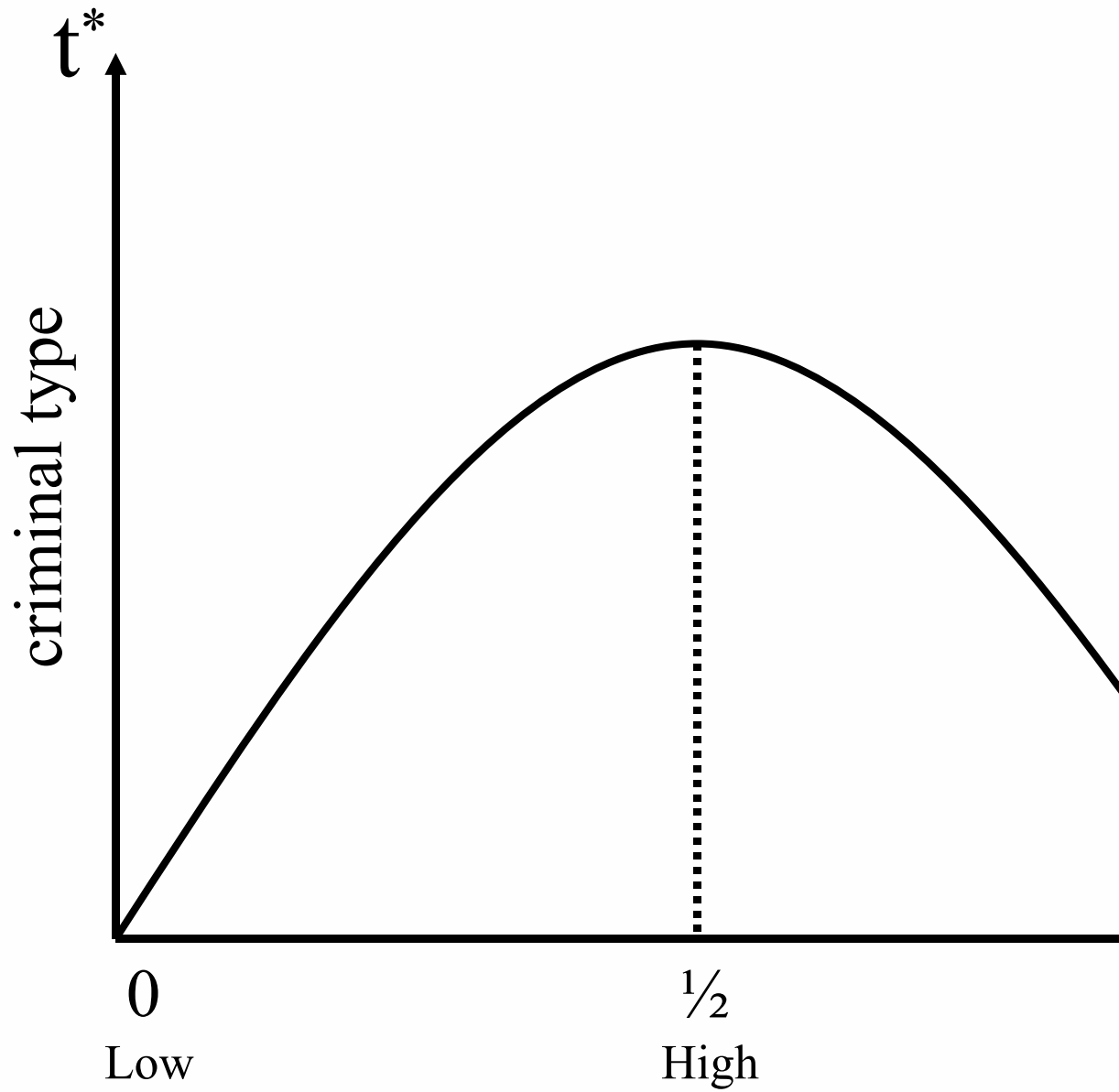
Figure 2: Degree of Competition and Equilibrium Candidate Choice, Including the Interaction between Competition and Fielding Costs
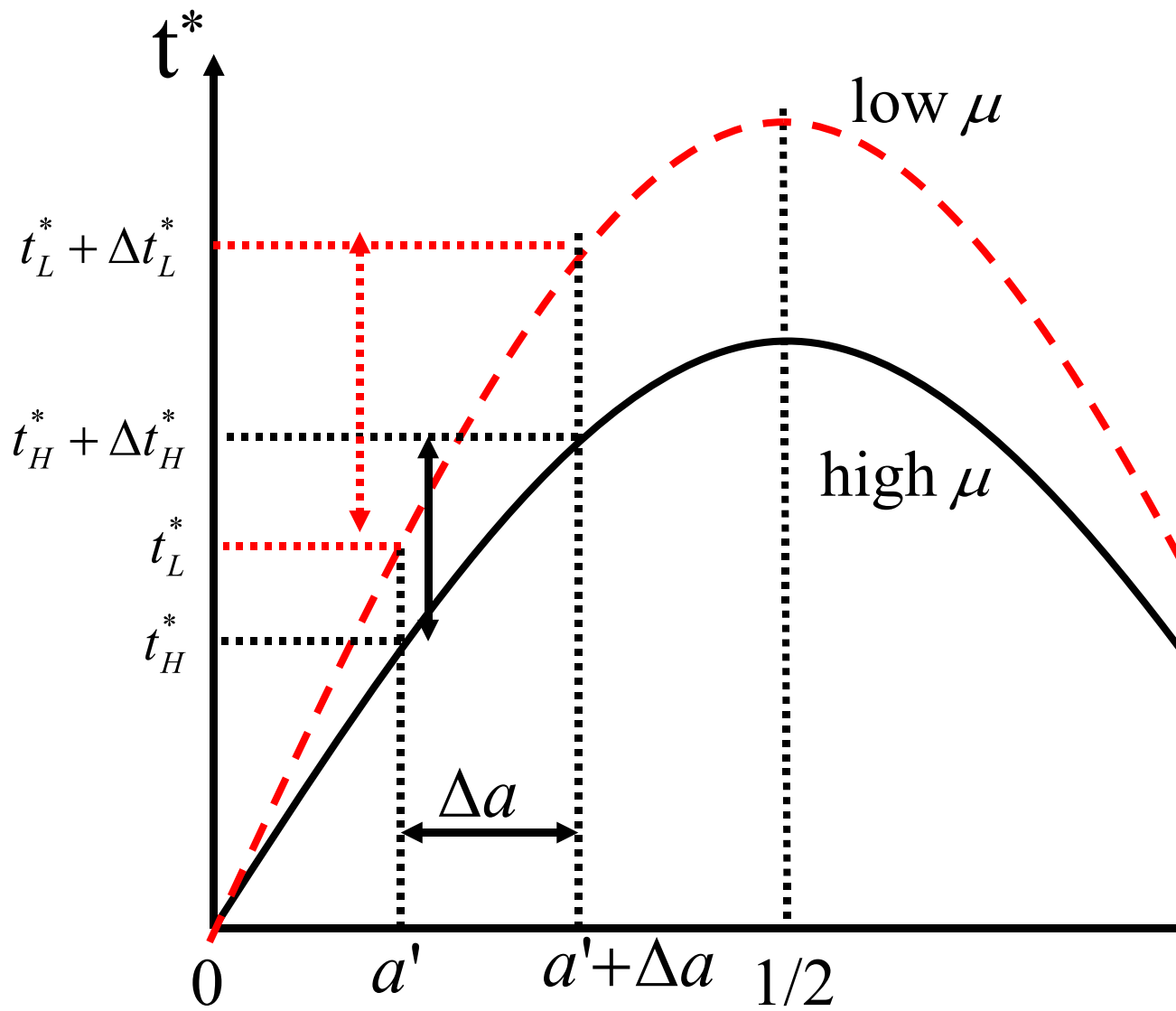
Figure 3: Maps of India's Electoral Districts Showing Where Criminal Candidates Appeared on the Ballot for Elections to the Fourteenth (2004) and Fifteenth (2009) Lok Sabha
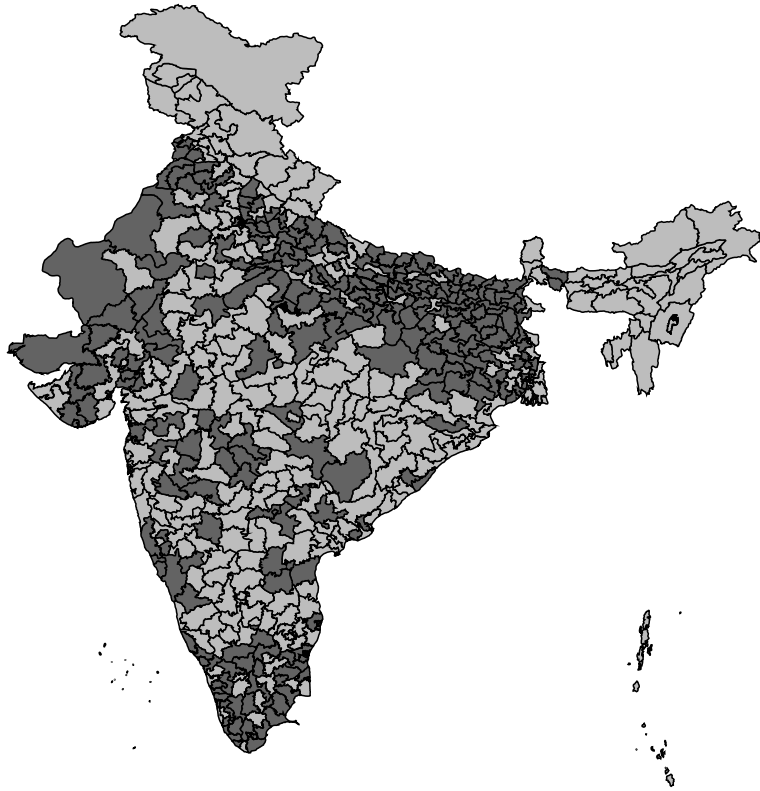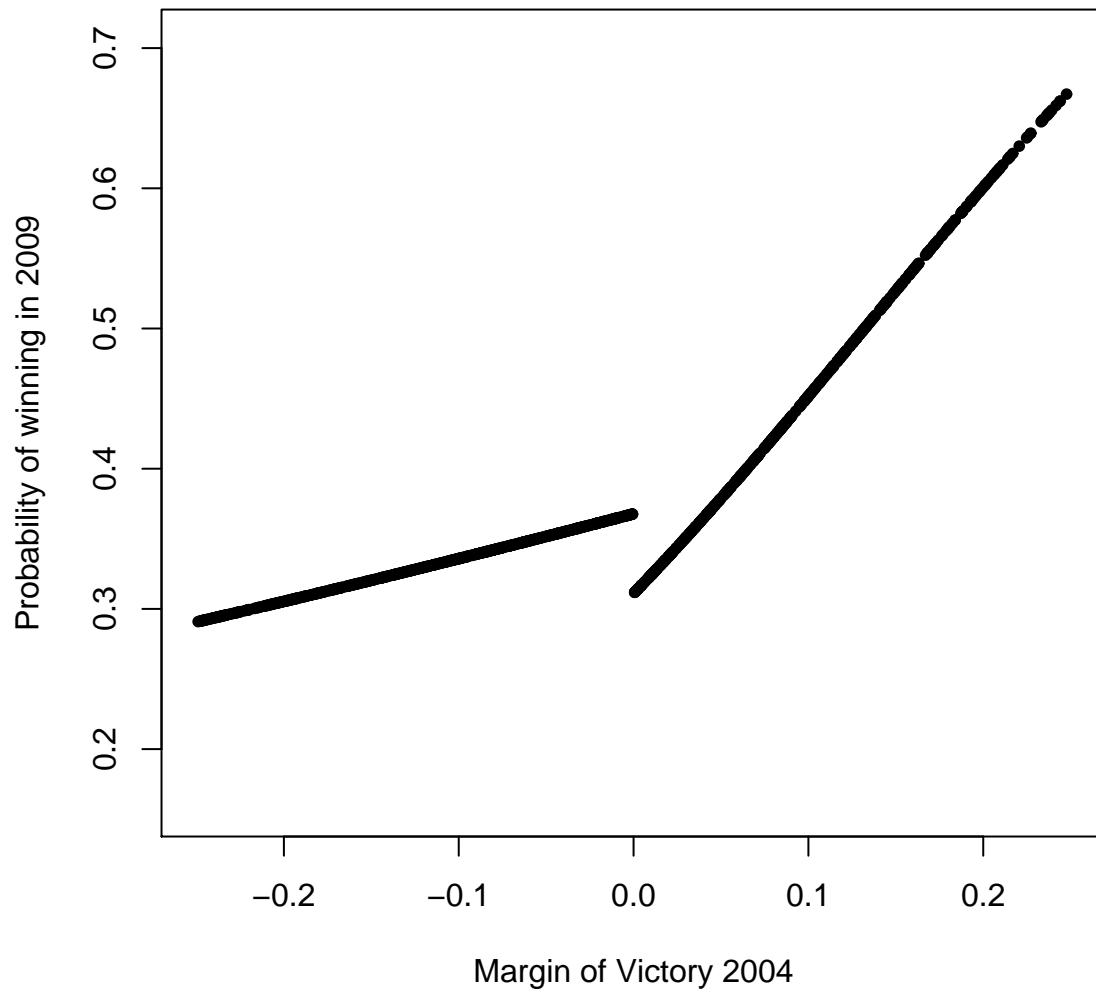
Figure 4: Regression Discontinuity Results of the Probability of Being Elected, All Candidates

Note: Candidates limited to those within 25 percent of winning.

Figure 5: Regression Discontinuity Results of the Probability of Being Elected, Non Criminal Candidates
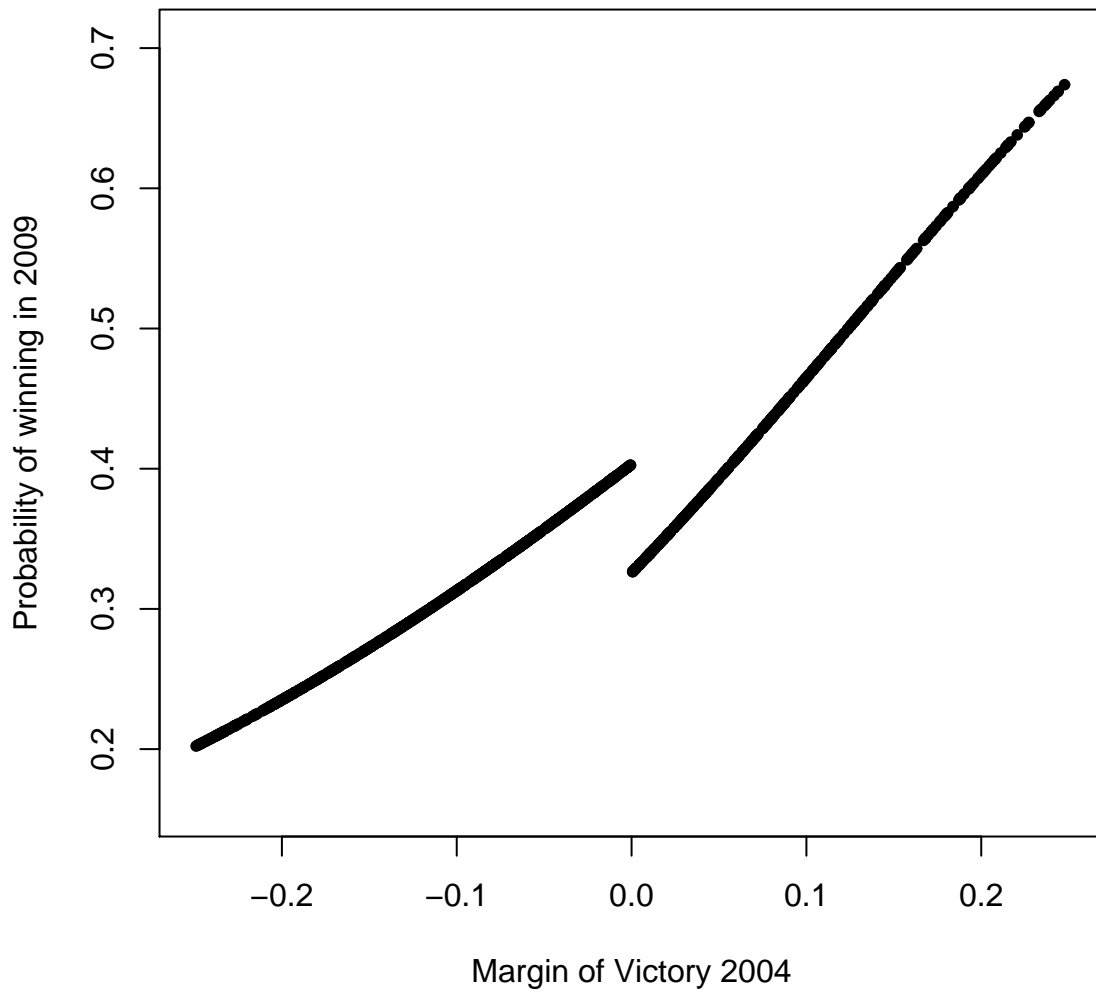
Figure 6: Regression Discontinuity Results of the Probability of Being Elected, Criminal Candidates
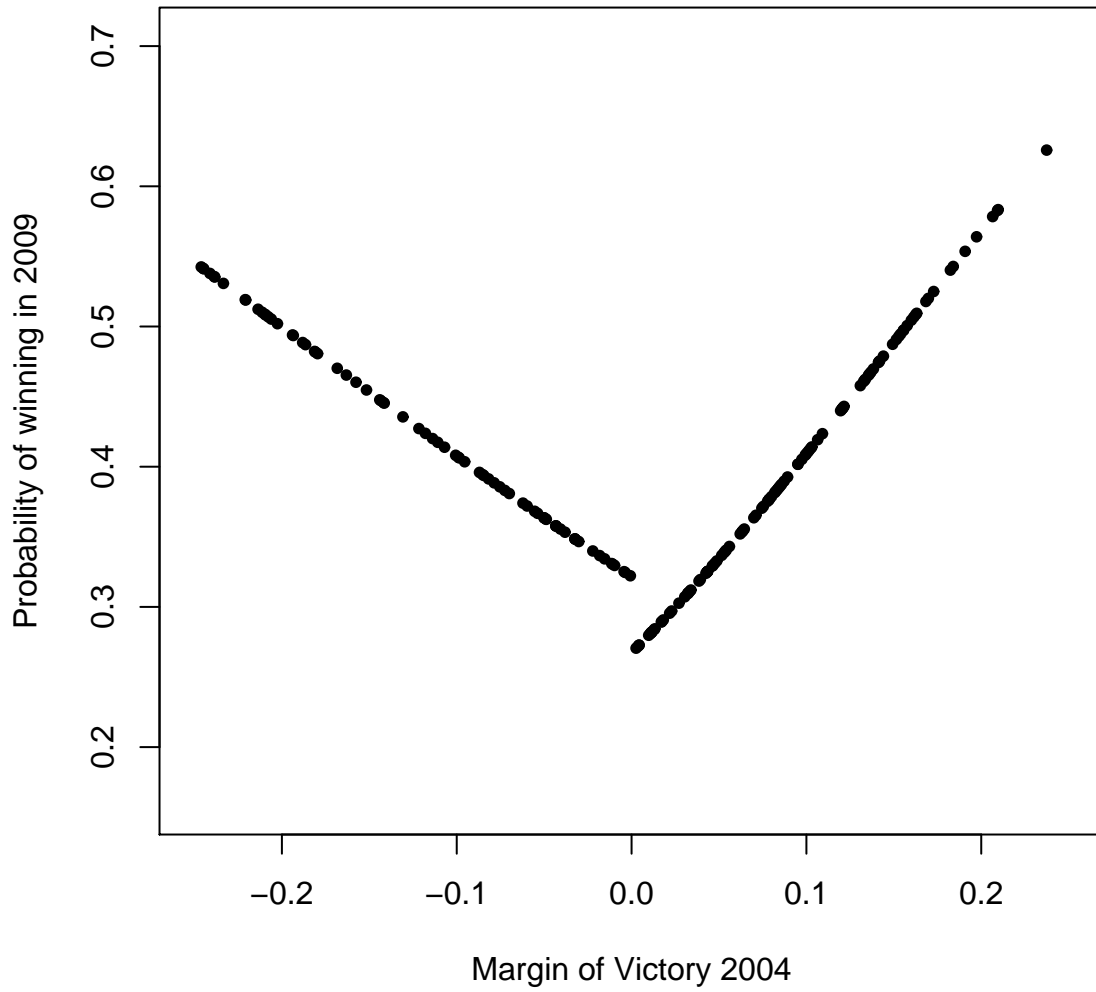
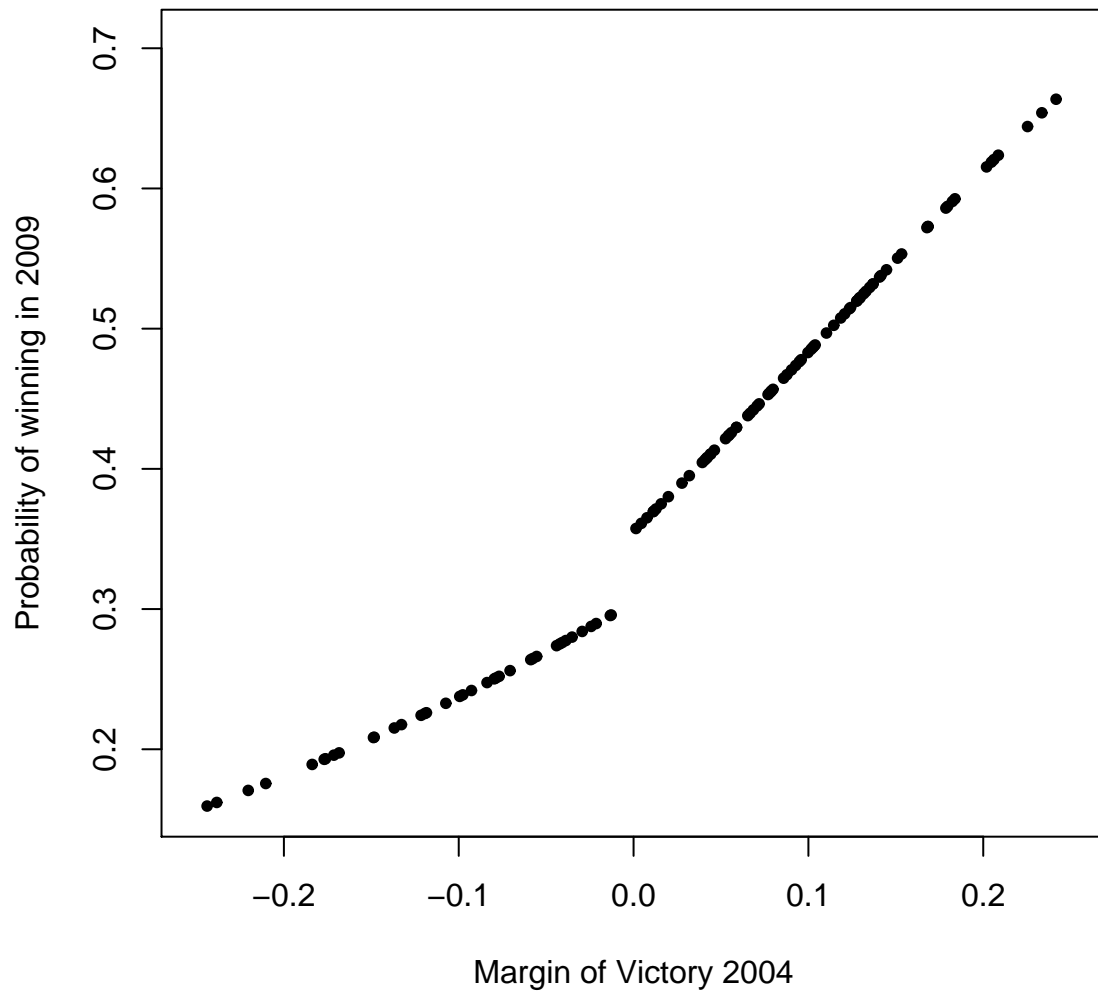Figure 7: Regression Discontinuity Results of the Probability of Being Elected Facing Non Criminal Opponent

Figure 8: Regression Discontinuity Results of the Probability of Being Elected Facing Criminal Opponent
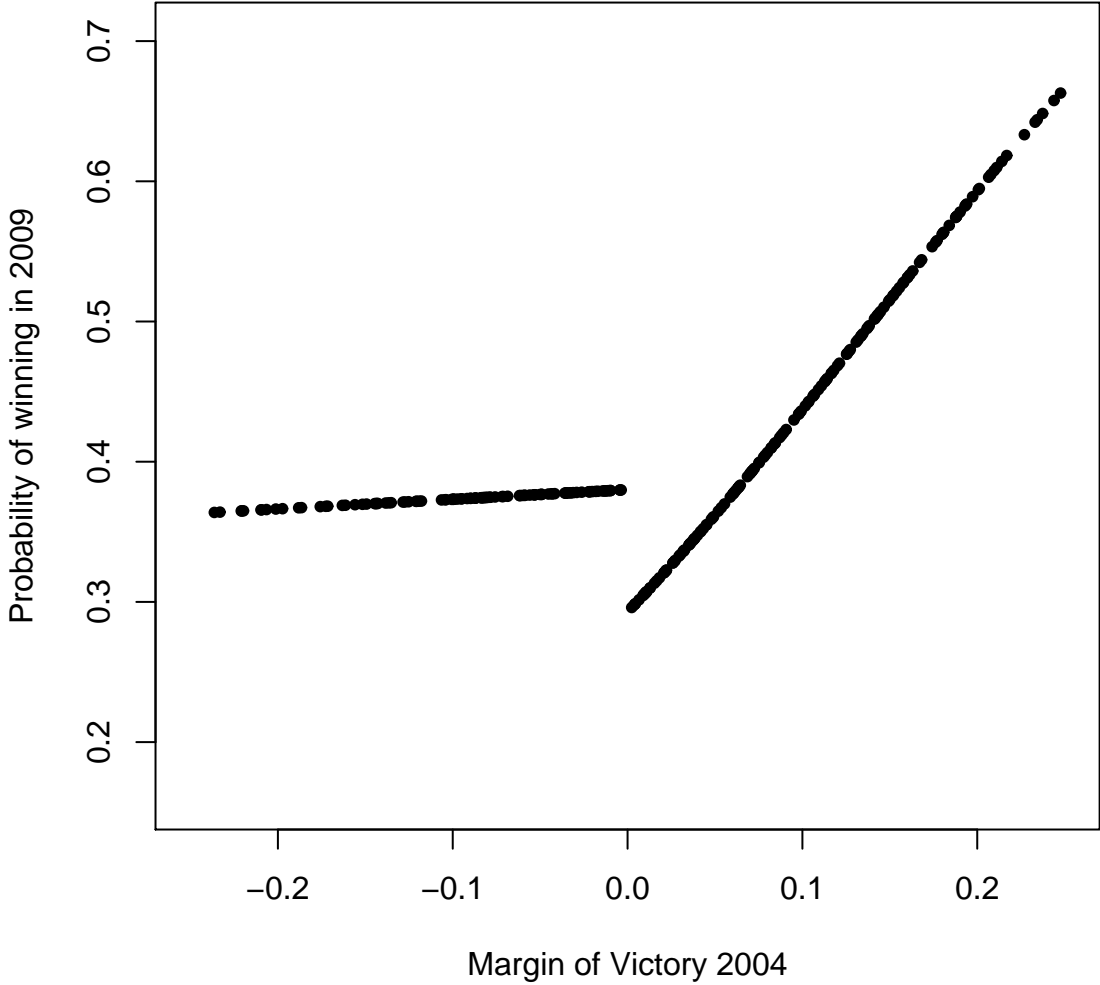
Figure 9: Regression Discontinuity Results of the Probability of Being Elected, Non Criminal Candidate Facing Criminal Opponent
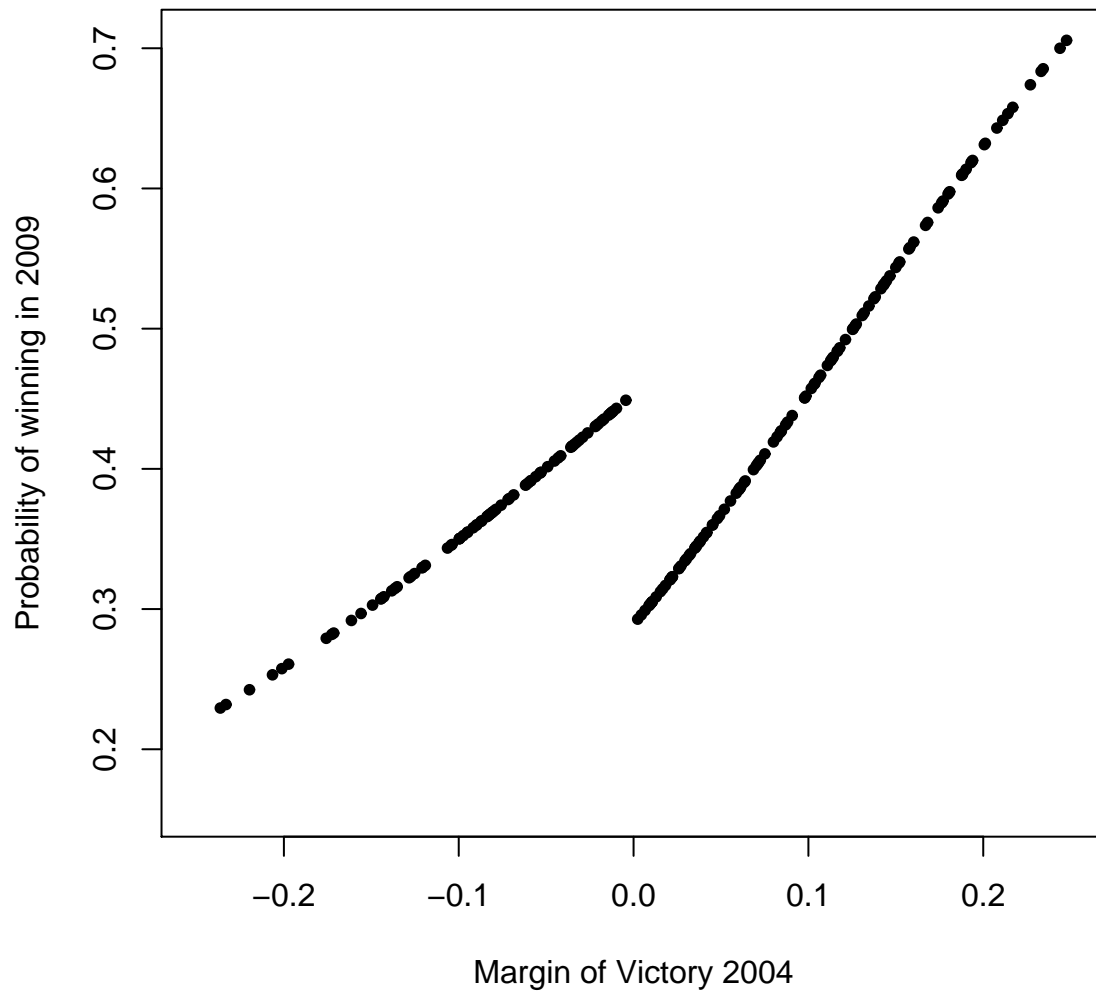
Figure 10: Regression Discontinuity Results of the Probability of Being Elected, Criminal Candidate Facing Criminal Opponent
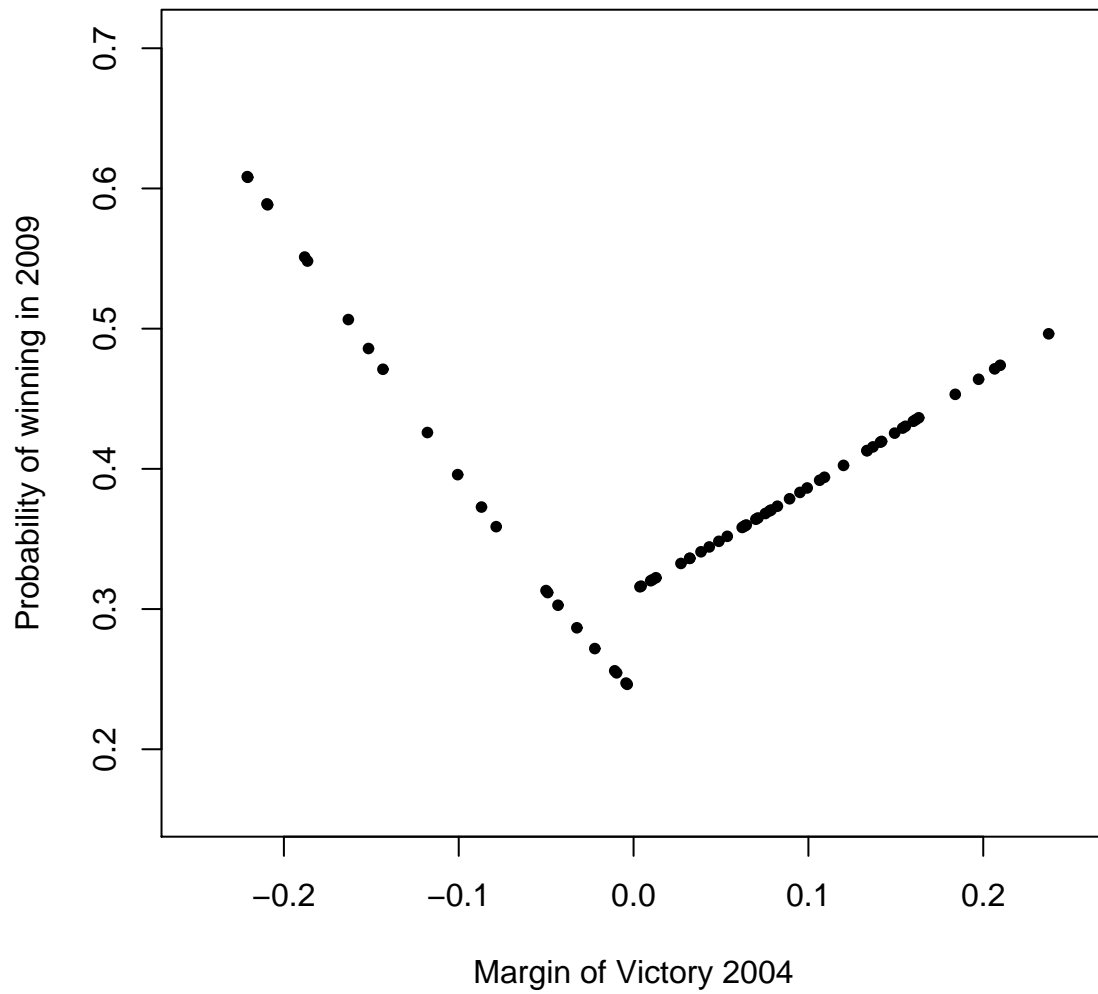
Figure 11: Incumbency (Dis)advantage:Charged Candidate and Charged Opposition
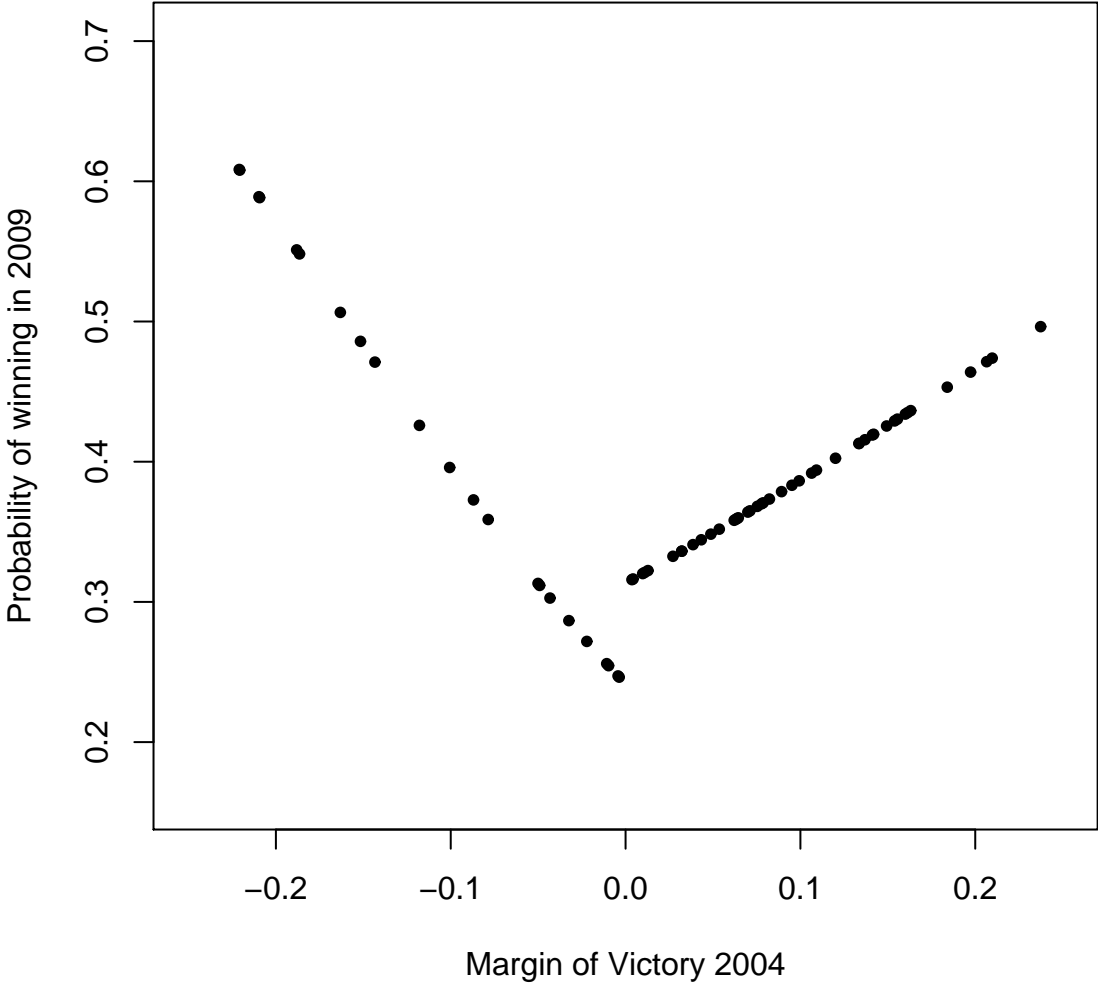
Table 13: Odds Ratios of Estimations of the Probability a Party Fields a Criminal Candidate. DV = Charged Candidate

Election	Model 1 2004	2009	Pooled	Model 2 2004	2009	Pooled	Model 3 2004	2009	Pooled	Model 4 2004	2009	Pooled
COMP	0.981***	0.973***	0.977***	0.982***	0.976***	0.979***	0.975*	0.979	0.975**	0.972**	0.989	0.976**
	(0.003)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.014)	(0.014)	(0.010)	(0.014)	(0.014)	(0.010)
LITERACY	0.977***	0.983***	0.981***	0.967***	0.986**	0.976***	0.964***	0.988	0.974***	0.978*	1.003	0.988
	(0.005)	(0.004)	(0.003)	(0.007)	(0.006)	(0.004)	(0.009)	(0.009)	(0.006)	(0.012)	(0.011)	(0.008)
COMPxLIT							1.000	1.000	1.000	1.000	1.000	1.000
							(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
INCUMBENT				1.278	1.001	1.026	1.278	0.998	1.027	1.255	1.159	1.096
				(0.205)	(0.232)	(0.131)	(0.206)	(0.232)	(0.131)	(0.212)	(0.275)	(0.144)
RESERVED				0.638***	0.575***	0.618***	0.637***	0.575***	0.618***	0.625***	0.594***	0.635***
				(0.098)	(0.071)	(0.059)	(0.098)	(0.071)	(0.059)	(0.101)	(0.077)	(0.064)
NATIONAL				0.982	1.417***	1.197**	0.983	1.422***	1.195**	1.047	1.458***	1.251***
				(0.121)	(0.147)	(0.095)	(0.121)	(0.148)	(0.095)	(0.138)	(0.158)	(0.103)
POPULATION				1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
PCT-URBAN				1.007*	0.997	1.002	1.006	0.997	1.002	1.000	0.999	1.000
				(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.005)	(0.004)	(0.003)
Constant	0.812	0.983	0.890	1.618	0.669	1.110	1.916	0.603	1.258	0.377	0.147***	0.325**
	(0.213)	(0.204)	(0.145)	(0.726)	(0.290)	(0.338)	(1.080)	(0.357)	(0.505)	(0.288)	(0.107)	(0.160)
State FX	NO	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES
Observations	3050	4052	7102	3050	3667	6717	3050	3667	6717	2828	3641	6637

z-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 14: Marginal Effect of Literacy Rate and Competitiveness on the Probability of Fielding a Charged Candidate

Model	Baseline levels	Literacy = 42%	Literacy = 66%	Marginal Effect	Margin = 14%	Margin = 44%	Marginal Effect
Baseline Model 2004	10.9	14.1	8.5	-39.72	14.6	8.7	-40.41
Baseline Model 2009	14.2	16.9	11.8	-30.18	21.0	10.5	-50.00
Baseline Model Pooled	12.8	15.7	10.4	-33.76	18.1	9.8	-45.86
Controls+Interaction 2004	10.1	14.6	7.0	-52.05	13.9	8.1	-41.73
Controls+Interaction 2009	15.6	17.2	14.2	-17.44	20.1	13.7	-31.84
Controls+Interaction Pooled	11.4	14.6	8.5	-41.78	16.6	9.0	-45.78

Note: Each cell represents the probability, in percent, that a party fields a criminal candidate.

Note: Covariate levels are set to their mean or zero for dummy variables unless noted otherwise.

Table 15: Marginal Effect of Reserved and National Party on the Probability of Fielding a Charged Candidate

Model	Baseline levels	Reserved = 1	Marginal Effect	National = 1	Marginal Effect
Controls+ Interaction 2004	10.1	6.8	-32.7	9.9	-2.0
Controls + Interaction 2009	15.6	9.8	-37.2	20.7	32.7
Controls + Interaction Pooled	11.4	7.4	-35.1	13.3	16.7

Note: Each cell represents the probability, in percent, that a party fields a criminal candidate.

Note: Covariate levels are set to their mean or zero for dummy variables unless noted otherwise.