

Incumbents and Criminals in the Indian National Legislature 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

June 9, 2011
Version 4.1. Comments welcome.
Reports work in progress.

¹Portions of the work reported here were presented at the 2009 and the 2010 Annual Meetings of the American Political Science Association, the University of California at Riverside, Georgetown University, the World Bank (New Delhi), the Johns Hopkins School of Advanced International Studies, the University of Virginia, Loyola Marymount University, and the University of California at Los Angeles. For comments, we are grateful to Scott Desposato, Philip Keefer, Studi Khemeni, and John Echeverri-Gent. Funding for the research reported here was provided to Miriam Golden by the Academic Senate, the International Institute, and the Center for International Business Education and Research at the University of California at Los Angeles.

Abstract

Utilizing data on criminal charges lodged against candidates to the Fourteenth and Fifteenth Lok Sabha, India's lower house of representatives, we study the conditions that resulted in approximately a quarter of members of parliament elected in 2004 and in 2009 facing or having previously faced criminal charges. Indian political parties are more likely to select alleged criminal candidates when facing greater electoral uncertainty and in parliamentary constituencies whose populations exhibit lower levels of literacy. We interpret allegations of criminality of parliamentary candidates as related to their capacity to intimidate voters, and we show that criminal candidates depress electoral turnout. Finally, we document that the well-known incumbency disadvantage characterizing Indian legislative elections stems from the superior electoral performance of allegedly criminal candidates, who drive incumbents from office. Our results raise questions for democratic theory, which claims that electoral competition improves accountability, and for the future of the Indian polity, which is experiencing a growing criminalization of the national political arena.

1 Introduction

Why would a political party in a competitive democratic system recruit a known criminal to run for national public office? There are varied puzzling aspects to this phenomenon. Not only is it puzzling that a party would stand a candidate facing criminal charges since, especially if the charges involve acts of non-familial violence, this ought to prove electorally disadvantageous, it is puzzling that voters, instead of repudiating such a 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 parliamentary constituencies not once but twice in a row in a highly competitive multiparty system in a well-established, long-standing democratic polity known for its vigilant civil society and aggressive free press.

We analyze data on the 2004 and 2009 Indian national elections, the first and second legislative elections in which candidates to the national parliament were required by a 2003 Supreme Court ruling to file sworn affidavits that included, among other pieces of information, declarations of criminal records and then-current indictments.¹ Our dataset contains information drawn from all the affidavits filed by the more than 12,000 parliamentary candidates in these two elections. Empirically, our goal is to isolate the main electoral, demographic, and other factors associated with the appearance of allegedly criminal candidates on the ballot. Theoretically, we seek to shed light on how electoral democracies generate situations in which accountability appears so seriously compromised.

Our story is simple although troubling. We contend that allegedly criminal politicians are selected by political parties because these candidates specialize in the use of electoral tactics not employed by their rivals. We believe these tactics are related to expertise in technologies of violence, and we document accordingly that allegedly criminal candidates appear on the ballot when the parliamentary seat is highly contestable, where there are more illiterate and therefore more vulnerable voters, and that these candidates

¹We discuss below the validity of the information reported. We do not distinguish between persons who report having been convicted of a criminal offense from those who report facing indictment, and we refer interchangeably to persons who report facing indictment on criminal charges as “self-reported,” “acknowledged,” or “publicly known” criminals.

depress electoral turnout. We interpret this as evidence in support of our claim that these candidates intimidate opposition voters from going to the polls. In addition, we show that these candidates successfully drive incumbents out of office. We interpret this as evidence that holding a seat in the national legislature in India is highly valued — indeed, so highly valued that individuals regularly use violence to obtain one.

Substantively, who wins an election is of greater interest than who runs for office. If alleged criminals are listed on the ballot but win only a handful of seats, the phenomenon would be of only marginal interest and of limited political importance. But in fact, the reverse is the case. As the data reported in Table 1 shows, in both 2004 and 2009 Indian candidates to the lower house whose affidavits reveal criminal charges have a much larger likelihood of winning than other candidates. In 2004, more than a quarter of those facing criminal charges won their seats compared with a success rate of only 8 percent for other candidates. The 2009 elections were much more competitive — the total number of candidates rose 50 percent over the 2004 total — but even so, allegedly criminal candidates 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, those reporting criminal charges were two to three times more likely to win than other candidates. This obvious electoral advantage makes candidates reporting criminal charges unusually attractive to political parties. Although in most circumstances we expect criminality to constitute an electoral hindrance, in contemporary India the reverse appears to be the case. We seek to understand why.

[Table 1 about here]

Our paper is organized as follows. First, we briefly review various strands of literature related to our study. Second, we formalize the decision by a political party to select a criminal candidate. We then turn to the main statistical estimations. Here, we first conceptualize our dependent variable as the likelihood that an allegedly criminal candidate appears on the ballot, and we analyze the determinants of candidate selection. We then conceptualize our dependent variable as the number of candidates in each constituency who are self-reported criminals to see if our results continue to hold. In the fourth section, we show that

self-reported criminals on the ballot depresses voter turnout and in the fifth, we analyze patterns in the numbers of self-reported criminal candidates in each constituency. A sixth section investigates selection effects. Finally, our concluding section raises unanswered questions and suggests future research paths.

Our study has two theoretically unexpected implications. First, as we discuss in the next section, various studies document that once information about political malfeasance is released and disseminated to voters, they will use the information to elect honest challengers or, what comes to the same thing, public officials, anticipating voter backlash, will substantially reduce the frequency of corrupt or malfeasant activities. We observe voters successfully using new information to reduce the frequency and scope of political corruption in settings as diverse as Uganda (Reinikka and Svensson 2005), Brazil (Ferraz and Finan 2008), and Italy (Chang, Golden and Hill 2010). On the basis of such studies, we expect that the release of public information about criminal charges lodged against members of the Indian political elite should have reduced the number of alleged criminals elected — if not immediately, then certainly by the second election in which candidates filed affidavits reporting their legal profiles. Instead, we observe a nearly uniform persistence in the rate of self-reported criminality among the elected. This suggests that information alone is inadequate to reduce political corruption and criminality, and that prior studies may have omitted important variables from consideration. Other research that we refer to below indicates that ethnic or programmatic attachments to a party may dissuade voters from turning against that party even if its leaders engage in corrupt, criminal, or otherwise malfeasant behavior. Our research refocuses attention on the importance of violence as well. It thereby builds on other studies of the strategic use of violence in electoral contests, especially Wilkinson (2004): where voters are impoverished, illiterate, and easily coerced, organized groups may strategically deploy violence to reelect corrupt or criminal political representatives.

A second implication of our study speaks to the literature on legislative incumbency. A number of papers document an incumbency disadvantage for legislators at the federal as well as the 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 alleged criminality of candidates. Incumbency disadvantage occurs because self-reported criminals boast an extraordinary political advantage against even experienced politicians. In India, we appear to be witnessing the nearly wholesale take-over of large chunks of an established democratic political system by persons with criminal records. Moreover, this is occurring even as Indian party politics becomes increasingly competitive. This suggests that democratic accountability may encounter unexpected obstacles that have not been previously identified and that are not well understood or analytically appreciated.

2 Related Literature

Our paper builds on four distinct strands of literature: first, prior studies that also use Indian candidate affidavit information; second, studies of electoral responses to political malfeasance; third, studies of incumbency disadvantage in India; and, fourth, studies of political competition and accountability. We briefly review these various classes of studies.

We have located three other scholarly studies that analyze Indian candidate affidavits, although for somewhat different purposes or at levels of government below the national. Chemin (2008) studies local outcomes after the 2004 parliamentary elections and finds that bureaucratic corruption falls but poverty rates increase in constituencies that elect self-reported criminals.² These outcomes are consistent with a theoretical framework in which criminal elements enter and capture the polity; such a view expects political control by organized crime to reduce the unorganized criminality of the petty bureaucrat while also encouraging economic inequality. Banerjee and Pande (N.d.) examine the 2004 affidavits of candidates to the assembly in the state of Uttar Pradesh, a state well known for a high level of political corruption and criminality, and find that parliamentarians who are elected are more likely to be alleged criminals in constituencies with more low-caste voters. The broad lines of their argument is that corruption and criminality among politicians in Uttar Pradesh have been due to the rise of low-caste and ethnic voting and that ethnic party politics creates

²In India, electoral districts for the national legislature are known as parliamentary constituencies. We use this terminology since the term “district” is used in the Indian context to refer to administrative units.

incentives for political corruption. These findings are not inconsistent with ours. We find that areas with more illiterate voters (which broadly coincide with those with more low-caste voters) also see more alleged criminals on the ballot.

The study closest to ours, finally, is Vaishnav (2010), who uses a political selection framework to analyze the affidavit information of party-affiliated candidates to state (not federal) elections in India. His main argument is that allegedly criminal candidates are attractive to political parties because such candidates are self-financing, perhaps because of assets amassed thanks to criminal activities. However, in contrast to the results we report, Vaishnav finds no evidence that electoral competitiveness increases the likelihood that an alleged criminal appears on the ballot.

In addition to studies that analyze the affidavits of Indian parliamentary candidates, various studies, 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), have analyzed how voters respond to revelations of wrongdoing on the part of public officials. They find that revelations of malfeasance by elected officials result in electoral retribution under certain conditions — for instance, when information about malfeasance is disseminated widely to voters. When it is not, elected officials with records of wrongdoing may be repeatedly reelected.

We highlight another reason why, even if they are informed of candidate malfeasance, voters may end up electing such individuals into public office. Information is not fully effective when voters are subject to physical intimidation. This may compromise political accountability by scaring voters into staying home on election day. If potential opponents of reputedly malfeasant candidates do not turn out to vote for another candidate, the former may win the seat. Such practices are likely to be all too common in environments characterized by weak rule of law, endemic political corruption, and widespread organized crime.

A third strand of relevant literature studies the electoral (dis)advantage of incumbent politicians in India. 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 2009; 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 offers an entirely new understanding of these phenomena. We analyze incumbency disadvantage together with allegations of candidate criminality to estimate the effect of allegedly criminal opponents on incumbent reelection probabilities. We evaluate whether the proliferation of alleged criminals on the ballot is causally linked to the growing incumbency disadvantage in India and show that it is.

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 argument has been extended by Besley and Burgess (2002), who show that public food distribution across Indian states in response to declines in grain production and flood damage is improved where newspaper circulation and political competition are greater. Our results fail to corroborate that political competition enhances accountability and responsiveness. Instead, we show that where partisan competition is more intense and when a party has less assurance about whether it will win the seat, that party is more likely to list an alleged criminal on the ballot. The aggregate outcome of intense partisan competition is thus to diminish political accountability.

We are not the first to observe that competitive electoral pressures may encourage wrongdoing by elected officials (Nyblade and Reed 2008). However, the mechanism that we identify to explain why Indian political parties list allegedly criminal candidates is new. We contend that the self-reported criminal candidates that we study are distinguished not by their involvement in illegal fund-raising or any financial advantage they may bring with them but rather by their expertise in the use of violence. Although we cannot observe this

directly, we test some implications of this theory using as much data as we have been able to assemble and in more than one way.

3 A Model of Candidate Selection and Criminality

We present a formal model designed to illustrate possible links between electoral competition, voter turnout, and the incentives of political parties to field candidates who publicly acknowledge facing criminal indictment. The model highlights what we believe is the central mechanism that makes known criminal candidates attractive to political parties: their capacity to intimidate opposition voters into staying home on election day. Formalizing these issues allows us to develop a framework of testable hypotheses that also facilitates interpreting and understanding the empirical results that we report in subsequent sections. In the model, we assume that self-reported criminals are in fact actually engaged in criminal activities; in reality, of course, these two phenomena may not be identical.

3.1 Assumptions

We consider two political parties $k \in \{A, B\}$ that compete in a constituency for a seat in parliament. The parties have fixed ideological positions, and we take their electoral platforms as given.³ Each party has a natural support base of voters who for ideological or identity reasons are committed to support it over other party contenders. We denote the number of party k supporters by N_k . Since voting is costly, only a fraction of the supporters will turn out to vote on election day. The turnout cost is, for simplicity, the same for all voters and denoted by $c > 0$. It captures the opportunity cost of voting, net of electoral intimidation. We assume that voting is an expressive act and that each voter gets some benefit out of the act

³This is in contrast to the standard Downsian or probabilistic voting model (see, e.g., Persson and Tabellini (2000)), where political competition centers on the platform choice. While this is an important consideration in many contexts, here our focus lies with selection. For our purpose, the platform choice is of secondary importance and can be thought of being determined centrally by the party organization for all constituencies and so as given for any particular constituency.

of voting irrespective of whether the votes influence the outcome or not.⁴ The benefit of voting is distributed according to a uniform distribution on the interval $[0, \theta_k]$ and may differ between the supporters of the two parties. The parameter θ_k captures how much value the average voter in each group attaches to voting and, as we shall see below, determines how resilient supporters of the two parties are in the face of violence and attempted political intimidation.

Potential candidates are of two different types, which we refer to as criminal and non-criminal, indexed by C and N .⁵ The two parties select the type of candidate they want to field simultaneously before the election. On the election day, voters decide whether to go to the polls to vote for their favored party or to stay home. We refer to this as the candidate fielding game.

We conceptualize the electoral benefit of fielding a known criminal rather than a non-criminal candidate as the access enjoyed by the former to an organized network of persons who can target opposition voters with threats of or the actual use of violence, thereby intimidating some of them not to show up at the polls.⁶ In particular, we assume that a criminal candidate through this mechanism increases the cost of voting for opposition voters. This reduces turnout rates of opposition voters, potentially swinging the election in favor of the party represented by the criminal. Criminals, however, have limited resources to produce electoral violence and therefore to intimidate. We capture this by assuming that a criminal candidate can at most increase the cost of voting for each opposition voter by $t > 0$. The parameter t can be interpreted as a measure of the productivity of violence. Non-criminal candidates do not have access to this technology.⁷ To ensure some turnout even with criminals in the race, we assume that $\theta_k > c + t$ for all k .

⁴This formulation circumvents the paradox of voting. For a discussion of expressive political behavior see, for instance, Jennings and Hamlin (N.d.).

⁵We focus on the demand side — that is, on why parties select candidates with criminal records — taking the supply of criminal candidates as given. However, there is good reason to believe that public office is highly valued in India. Recent research shows that winning office increases the average candidate’s family assets by 28 percent over five years Bhavnani (2011).

⁶Empirically, other work supports our hypothesis that violence reduces turnout. Research in Africa documents the effects of voter intimidation on turnout; in that context, as perhaps also in India, violence is especially effective in reducing turnout among the rural poor, where even highly selective intimidation penetrates larger communities of voters. See Bratton (2008).

⁷We could assume that voters, in general, do not like known criminal candidates and that a party that selects such a candidate erodes its own support base. This feature is not essential for our results as long as criminal candidates compensate for this loss of votes through their capacity to reduce turnout of opposition voters. To simplify the analysis, we ignore this effect but it would be straightforward to incorporate it into the model.

We assume that there is a small (but strictly positive) cost associated with fielding a criminal candidate. We denote the cost by μ and assume, for simplicity, that it is the same for both parties.⁸

In the absence of criminal candidates in the race, the number of voters turning out in support of party k is $N_k(1 - \frac{c}{\theta_k})$ and party A is elected if

$$\Delta \equiv (1 - \frac{c}{\theta_A}) - \gamma(1 - \frac{c}{\theta_B}) > 0 \quad (1)$$

where $\gamma = \frac{N_B}{N_A}$ is the relative size of group B .⁹ We assume that party A would win the seat with certainty if both parties field non-criminal candidates (i.e., $\Delta > 0$). We refer to party A as the top dog and party B as the underdog in the race. The closeness or contestability of the election is determined by the relative size of the two groups of supporters. We can express the assumption that party A is the top dog as a condition of the relative group size:

$$\gamma < \frac{\theta_B(\theta_A - c)}{\theta_A(\theta_B - c)} \equiv \hat{\gamma}. \quad (2)$$

If γ is large (and close to $\hat{\gamma}$), then Δ is close to zero and the race is close; if γ is small (and close to zero), then Δ is large and the seat is safe for party A . We can therefore use γ as an index of electoral competition: a low value corresponds to a situation with little competition whereas a high value corresponds to a situation in which the race, in the absence of any criminal candidates, is close and the seat is therefore competitive.

Finally, we assume that the value to a party of winning the seat is $M > 0$. It is reasonable to assume that the benefit of winning the seat for the party is always greater than the cost of fielding a criminal ($M - \mu > 0$).

If not, there would be no reason to do so.

⁸The fielding cost represents those disadvantages that a party suffers irrespective of its electoral success in the particular constituency as a consequence of allowing a criminal on the ballot under its symbol. This, among other things, includes reputation costs for the party nationally as well as the inconvenience for the local party organization of having to associate with criminals. The cost can be arbitrary small and it plays a role only as a tie breaker.

⁹If $\Delta = 0$, we assume that the election is determined by a toss of a coin.

3.2 Analysis and Results

Depending on how close the race is, three different electoral regimes can emerge. We name these regimes: safe seat, competitive defensible seat, and competitive indefensible seat. Recall that a criminal candidate can increase the voting cost of opposition voters by t . The effect of this on the number of voters intimidated depends on how vulnerable or resilient they are (i.e., on how big θ_k is). The maximum fraction of the supporters of party A that a criminal candidate fielded by party B can intimidate not to vote is $\Delta_B = \frac{t}{\theta_A}$ and similarly $\Delta_A = \frac{t}{\theta_B}$ is the maximum fraction of party B voters that a criminal candidate fielded by party A can intimidate into not voting. We say that the race is *competitive* if party B — the underdog — can swing the election by fielding a criminal, i.e., if $\Delta - \Delta_B < 0$, and that the seat is *defensible* if party A — the top dog — can swing the election back if, in response, it fields a criminal, i.e., if $\Delta - \Delta_B + \gamma\Delta_A > 0$. With these preliminary remarks, we are ready to characterize the outcome of the candidate fielding game in the three regimes in which the two parties simultaneously field a candidate. We do so intuitively here (the technical details appear in Appendix B).

1. Safe seat: $\gamma < \hat{\gamma} - \frac{\theta_B t}{\theta_A(\theta_B - c)} \equiv \underline{\gamma}$. Even if party B fields a criminal candidate and party A does not, party A wins the seat. The candidate fielding game has a unique pure strategy Nash equilibrium in which both parties field non-criminals. The intuition is that fielding a criminal candidate is a dominated strategy for party B . It would not swing the election if it were to do so but it would incur the (small) fielding cost. Hence, party B does not field a criminal in this setting.
2. Competitive defensible seat: $\underline{\gamma} \leq \gamma < \frac{\theta_B(\theta_A - c - t)}{\theta_A(\theta_B - c - t)} \equiv \bar{\gamma}$. Party B can swing the election if it fields a criminal candidate and if party A does not. Party A can, however, defend the seat if it, in response, fields a criminal. The candidate fielding game has a unique Nash equilibrium in mixed strategies where the two parties each field criminals with positive probability.¹⁰ The probability that party A

¹⁰The reason the game does not have a pure strategy Nash equilibrium is that party A will field a criminal to defend the seat if party B fields one. Given that, party B does not want to field a criminal. But then party A prefers not to field a criminal either which, in turn, gives party B an incentive to do so, etc.

fields a criminal is $\frac{M-\mu}{M}$ and the probability that party B fields a criminal is $\frac{\mu}{M}$. The equilibrium outcome, then, may be that one, both or neither of the parties field a criminal candidate. The probability that we observe at least one criminal candidate is $\frac{\mu(M-\mu)}{M^2} > 0$.

3. Competitive indefensible seat: $\theta_B > \theta_A$ and $\hat{\gamma} > \gamma \geq \frac{\theta_B(\theta_A - c - t)}{\theta_A(\theta_B - c - t)} \equiv \bar{\gamma}$. Party B can swing the election if it fields a criminal candidate and party A does not, but party A is unable to defend the seat. The candidate fielding game has a unique pure strategy Nash equilibrium in which party B fields a criminal candidate while party A does not. This is an equilibrium because party B — the underdog — by fielding a criminal wins the election with a net gain of $M - \mu > 0$ irrespective of what party A — the top dog — does. Given this, the best response for party A is to avoid the (small) fielding cost and to select a non-criminal candidate. The condition that $\theta_B > \theta_A$ is required to ensure that $\bar{\gamma} < \hat{\gamma}$. Intuitively, for the seat to be indefensible, the supporters of party A must be relatively easy to intimidate.

3.3 Empirical Implications

Empirically, the model generates five sets of predictions that we investigate in the remainder of this paper.

The first is that criminal candidates are more likely to be fielded in races where the seat is defensible or indefensible — what we have called competitive seats — rather than safe. When the seat is safe, criminals are not given tickets, as the decision by a party to select a candidate is called in India. Figure 1 illustrates this result. On the horizontal axis, we record the value of γ which controls the degree of electoral competition, with larger values corresponding to a closer race and thus more electoral competitiveness. On the vertical axis we record Δ , $\Delta - \Delta_B$ and $\Delta - \Delta_B - \gamma\Delta_A$ for the case where $\theta_B > \theta_A$.

[Figure 1 about here]

Consider an electoral district or constituency located at point 1 with a value of γ close to zero. From the point of view of party A , this seat is completely safe because its support base is much larger than that of party

B: no matter what party *B* does, the latter cannot swing enough votes to win a majority. In constituencies like this, with low electoral competition, criminal candidates are not fielded by any party. In contrast, a constituency located at point 2 is home to a more competitive race. Here, although party *A* is the favorite, party *B* can, in principle, swing the election with the help of a criminal candidate ($\Delta - \Delta_B < 0$). However, the position of party *A* is sufficiently safe to make the seat defensible ($\Delta - \Delta_B - \gamma\Delta_A > 0$). In constituencies like this, parties occasionally field criminals and we may observe situations with two criminals, one criminal, or no criminal on the ballot.

When competition becomes even more intense, the seat may become indefensible ($\Delta - \Delta_B - \gamma\Delta_A < 0$). This is the case for constituency 3 in the figure. In this locality, party *A* cannot counter the challenge from party *B* by fielding a criminal and thus refrains from doing so. In a constituency such as this, we expect to observe a race with one criminal on the ballot.

The testable implication is that criminal candidates will appear on the ballot more often where the race is competitive and, conversely, fewer criminal candidates will appear in races for seats that are safe. We model this empirically by assessing systematically whether, all else equal, parties are more likely to list self-reported criminals where their party is closer to winning the seat.

A second and related prediction of the model concerns the numbers of acknowledged criminals that are likely to be listed in any one constituency. Unlike other models that study when criminal candidates are listed (for instance, Vaishnav (2010), ours does not generate predictions about the specific numbers of criminal candidates that should be observed at the constituency level. For instance, we do not necessarily expect to see criminal candidates appear in pairs, as occurs when a party matches another party's decision to list a self-reported criminal by also selecting an alleged criminal. Instead, overall we expect to observe the full range of possibilities: constituencies with none, with one, with two, or with more than two alleged criminals on the ballot. We can assess this empirically by examining the distribution of the numbers of self-reported criminals on the ballot across India's parliamentary constituencies.

A third prediction of the model is that criminals are more likely to be fielded in constituencies where criminality is more effective in reducing voter turnout. This effect can operate through two channels: the productivity of criminal candidates, t , and the vulnerability or resilience of voters (θ_k). Firstly, an increase in t lowers the threshold $\underline{\gamma}$ that determines if the seat is safe or not, and so makes it more likely that criminals are fielded. Second, suppose for the sake of argument that all voters are equally resilient, i.e., $\theta_A = \theta_B = \theta$. Then a reduction in resilience unambiguously decreases the threshold $\underline{\gamma}$ and so, again, makes it more likely that criminal candidates are fielded.¹¹

In our empirical models, we use illiteracy as a proxy for how effective criminal technologies are likely to be. This is because where illiteracy is higher, we expect criminal candidates to depress turnout more effectively. Our reasoning is that literate voters are endowed with a whole panoply of resources, including access to information and access to police protection, that together offers them greater resilience in the face of electoral violence. In India, as is the case in many less developed nations, there is evidence that higher socio-economic status, of which literacy is both a component and a marker, is associated with more frequent recourse to police services [cite]. Citizens with low socio-economic status are less likely to access police services even when confronting circumstances that would warrant using them. Higher status voters, by contrast, are more knowledgeable about their legal rights, less economically vulnerable, and more capable in their interactions with government services. As a result, we argue that literate voters are less likely to be intimidated by threats of violence and to stay home on election day. We therefore expect to see more criminal candidates on the ballot where illiteracy is higher.

A fourth implication, directly following the preceding, is that we expect criminal candidates to depress turnout more than other candidates. We study this by investigating whether self-reported criminal candidates significantly affect turnout rates across parliamentary constituencies in the 2004 and 2009 elections.

¹¹The comparative statics is

$$\frac{\partial \gamma}{\partial \theta} \Big|_{\theta_A = \theta_B} = \frac{t}{(\theta - c)^2} > 0.$$

Fifth, our model carries second-order implications for longevity in office. In some contexts, it is natural to think about party *A*'s electoral advantage as an incumbency advantage. If we do so, our model suggests that in constituencies with what we have called competitive seats, acknowledged criminal candidates can erode whatever incumbency advantage exists. In other words, even if the incumbent party has an ex ante advantage, perhaps because of the size of its natural support base in the constituency, the fact that opposition parties can (and do) field criminals to intimidate the incumbent's voters may turn the ex ante incumbency advantage into an ex post incumbency disadvantage. This effect arises in cases with what we have labeled indefensible seats. To examine this empirically, we utilize a regression discontinuity design to study whether the proliferation of self-reported criminals on the ballot systematically disadvantages incumbent legislators.

Our empirical estimations examine when self-reported criminals appear on the ballot but not when they win the seat. The reason is that winning is endogenous to receiving a ticket. That is, parties select the candidates they believe are most likely to be able to win. If criminals are in fact more likely to win, then they will be more often selected. Hence, to avoid endogeneity issues, we study the conditions in which self-reported criminals are selected and not whether or not they win.

To summarize, we study the following hypotheses:

1. Criminal candidates are more likely to appear on the ballot where the race is close;
2. Criminal candidates are more likely to appear on the ballot where there is a higher proportion of illiterate voters;
3. Criminal candidates depress turnout more than other candidates;
4. Criminal candidates appear on the ballot as singletons, as pairs, or as multiples, and no specific configuration dominates;
5. Criminal candidates erode incumbency advantage.

We investigate the first two hypotheses using two different conceptualizations of the dependent variable: one that studies the probability of observing a publicly known criminal on the ballot and a second that studies the number of self-reported criminal candidates in each constituency. For the first, the unit of analysis is the political party in each parliamentary constituency; for the second, it is the parliamentary constituency. Each is a slightly different way to formulate empirically the determinants of acknowledged criminals on the ballot.

For the first, we estimate logit models of the following type:

$$\Pr [Y_{it} = 1] = F [\beta_0 + \beta_1 COMP_{it} + \beta_2 LIT_{it} + ..]$$

where $COMP$ measures the competitiveness of the race, LIT measures the literacy rate of the electorate, F represents the cumulative normal distribution, i is the parliamentary constituency, and t is the electoral period (2004, 2009, or both).

As an alternative to the logistic model that studies when parties select self-reported criminals as candidates, we also consider a poisson model where we define the outcome as the number of known criminal candidates fielded in constituency i in election t , i.e., $y_{it} = \sum_j Y_{ijt}$. Formally, we can write

$$y_{it} = \alpha_0 + \alpha_1 COMP_{it} + \alpha_2 LIT_{it} + .. + \varepsilon_{it}$$

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

We operationalize electoral competitiveness by measuring how close each party expects to be to winning the seat in the constituency. In the multiparty setting in which we work, this generates a separate measure for each of the many political parties (and thus, for each partisan-affiliated candidate) in each constituency in each of the two legislative elections for which criminal records are disclosed by candidates. Redistricting between elections, which prevents us matching 2009 districts back to those that existed in 2004, requires that

we use the actual constituency-level electoral results from 2009 as proxies for the expected competitiveness of the race. One justification for this, in addition to the sheer necessity of doing so, is that we believe that candidates are selected by parties on the basis of relatively current and accurate information about how competitive the race is likely to be. Especially given the costs of fielding a known criminal, parties are likely to deploy various instruments to collect information about the sentiments of the electorate in the period leading up to the election. Therefore, if the decision to list a self-reported 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 in the weeks prior to the election when they must make final decisions about which candidate to list. For 2004, by contrast, we are also able to estimate the expected competitiveness of the race using the results from the prior (1999) elections. As we show, it makes little difference which of the two possible measures of competitiveness is used. We compare results using the prior election as a proxy for competitiveness with those that use the current election in Appendix C.

4 Data and Empirical Analysis

We test the hypotheses just discussed by studying information drawn from the affidavits filed by candidates to India's Fourteenth and Fifteenth national parliaments. We have coded all candidates by whether they report having been convicted of or face indictment for criminal activity.¹² Although the affidavits that Indian legislative candidates are legally required to file are self-reported, candidates who perjure themselves place themselves in legal jeopardy and may be disbarred from the race. Their electoral opponents, moreover, have obvious incentives to scrutinize their affidavits, as do the many anti-corruption nongovernmental organizations active in the field. So while the measure of criminality that we use may underestimate the real extent of criminal indictments or convictions of parliamentary candidates, we believe the underestimation is modest.

¹²Due to the intricacies of the often multiple charges against individual candidates, we do not attempt to differentiate candidates according to the nature of the charges. However, we shortly present evidence that the charges often involve violence.

Moreover, to the extent that our data underestimate indictment of politicians, the bias operates against our hypotheses.

We have merged the affidavit data on candidates to India's national legislature with data on electoral outcomes in 2004 and 2009, electoral turnout, information on the party affiliation of candidates, information on which parties are national or regional in scope, the incumbency status of each candidate, and theoretically relevant characteristics of the voting population in each constituency, namely, literacy rates. We have also merged in data for relevant constituency-level controls. We detail the operational indicators later.

4.1 The Institutional and Political Context

Figure 2 depicts a map of the 543 parliamentary constituencies as of 2004, visually differentiated according to whether at least one known criminal appeared on the ballot or not.¹³ In 2004, fully half of the constituencies witnessed known criminal candidates; in 2009, self-reported criminal candidates appeared on the ballot in three-quarters of Lok Sabha constituencies, attesting to a diffusion of publicly acknowledged criminality in national political life. The data depicted on the map also shows that, perhaps contrary to common belief, criminality was not geographically clustered in India in 2004.¹⁴ Candidates facing criminal charges are widely dispersed across the subcontinent.

[Figure 2 about here]

Although the data depicted in Figure 2 refers to all Lok Sabha candidates, our statistical analysis omits candidates who are unaffiliated with any of India's many political parties. We are interested in knowing when parties list criminal candidates, and independent candidates by definition self-nominate. Including independent candidates introduces considerable noise into the data analysis, making results difficult to interpret. Finally, independent candidates are politically irrelevant. In 2004, 43 percent of India's legislative

¹³We have not been able to locate the information to create a similar map for 2009.

¹⁴A formal test shows that a state identifier for each of India's 35 states and union territories is not a significant predictor of the percent of acknowledged criminal candidates in parliamentary constituencies in 2004 ($p = .734$) although it becomes statistically significant in 2009 ($p = .056$).

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 have 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 numerous but close to politically irrelevant. The median vote share collected by the winner and the first runner-up together is 87 percent, making most parliamentary races effectively two candidate contests.¹⁵ The vote share of unaffiliated candidates was less than 1 percent in 2004 as well as in 2009, and no independent candidate won more than 7 percent of the vote in either election.

Examining only candidates who are put on the ballot by a political party and excluding independents, we find that the ability of self-reported criminal candidates to gain seats in the legislature is even more pronounced than for all candidates. We refer again to the data presented in Table 1. For partisan-affiliated candidates in 2004, reporting a criminal charge on their sworn affidavit more than doubles the rate of winning a seat, increasing it from 15 to 36 percent. In 2009, the difference is less marked but even so success rates among publicly-identified criminals are 20 percent compared with 11 percent for other candidates. For candidates listed as affiliated with one of India's numerous political parties, we find that reporting a criminal charge 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.¹⁶ The variable *CRIM* is coded 1 if the affidavit

¹⁵Chakrabarti, 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 this is much less true if we exclude the unaffiliated, who have virtually no chance of winning the seat to begin with.

¹⁶We use affidavit information downloaded from the Liberty Institute, an NGO. We initially downloaded the affidavits ourselves from the website of the Election Commission of India. We then paid an Indian data input company to input the data recorded on the affidavits electronically for us. However, preliminary data analysis revealed that dataset produced by this company was unreliable; as a result, we rebuilt our dataset using the information obtained on-line from the Liberty Institute. However, other scholars working

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 number of candidates in the constituency who are coded 1 on *CRIM* out of the total number of candidates in the constituency.

One potential objection to using charges drawn from affidavits is that the charges 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 any activities involving voter intimidation or attributes specifically associated with their criminal status 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, charged candidates are more likely to win elections.

If this line of argument is true, then it should be reflected in the nature of the charges against candidates, a large proportion of which should either be plausibly politically motivated or be charges to which politicians are naturally especially vulnerable given the nature of electoral competition. These charges could involve crimes such as libel and slander (of other candidates, for instance), or perhaps activities for which there is little or no direct physical or eyewitness evidence. Charges such as murder, arson, looting, or assault are less likely to be fabricated and are not charges to which politicians are especially or uniquely vulnerable.

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 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 *CRIM* is mainly picking up a true criminal element in Indian politics.

with the same data (e.g. Bhavnani (2011)) use datasets based on inputting by the same company, which has perhaps become more reliable.

It is easy to fabricate a charge of libel against a political opponent and difficult to fabricate a charge of murder.

[Table 2 about here]

Ethnographic work substantiates the regular involvement of some Indian politicians in criminal activities and violence [cite]. To illustrate the kinds of criminal activities of which prominent politicians are accused, we provide a few examples. Ganesh Singh was elected from the constituency of Satna in the state of Madhya Pradesh on the ticket of the Bharatiya Janata Party (BJP) in both the Fourteenth and Fifteenth Lok Sabha. At the time of each election, Singh stood indicted on multiple counts of cheating and forgery. He remained under indictment while serving as a Member of Parliament.

An example on which more information is publicly available — due to its notoriety — is that of Afajal Ansari, son of a one-time President of the Indian National Congress, who was elected to the 2004 Lok Sabha from the constituency of Ghazipur in the state of Uttar Pradesh (UP) on the ticket of the Samajwadi Party (SP) with 48 percent of the votes polled. In the affidavit filed with his candidacy papers, Ansari reported three criminal charges, including rioting (India Penal Code Section 147), rioting and armed with a deadly weapon (India Penal Code Section 148), and criminal force to deter a public servant from the discharge of his duty (India Penal Code Section 353). We display a page of his affidavit reporting these charges in Figure 3.¹⁷ (We direct the reader's attention to the statute numbers with an arrow.) In 2009, Ansari switched his party affiliation to the Bahujan Samaj Party (BSP) and again stood in the (redistricted) constituency of Ghazipur, where he again won the seat, this time with slightly more than 40 percent of the votes polled. Perhaps the decline in his vote share was related to the fact that Ansari filed his 2009 candidacy papers from jail, where he was being held in connection with the 2005 murder of Krishnanand Rai, a legislator from the BJP.¹⁸

¹⁷The sample is in Hindi, one of the eight possible languages in which affidavits are filed. It illustrates some of the difficulties in working with the original data used in this analysis. Many affidavits are handwritten, and therefore even more difficult to decipher than the sample page featured in Figure 3.

¹⁸In 2010, Ansari was expelled from the BSP, along with his brother, Mukhtar, who had also run for the Lok Sabha in 2009 under the BSP symbol and who at the time of his expulsion faced 30 criminal charges, including murder and kidnapping.

These examples suggest that self-reported criminality is an imprecise but not invalid proxy for genuine criminality. There is little here to make us believe that India's acknowledged criminal politicians are simply more skilled at political contests and therefore more vulnerable to fabricated or politically generated charges.

[Figure 3 about here]

Finally, how the information about the alleged criminality of legislative candidates is used within India is also relevant to our evaluation of the validity of the self-reported charges. The Indian press and multiple non-governmental organizations use the information from the affidavits to publicize the criminal element in Indian politics. An implication of this is that being charged with criminal activity is considered a potential electoral liability. Even if a candidate is unfairly charged, 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 self-reported criminal as its candidate thus carries with it fielding costs that are higher than those associated with other candidates. Given the added costs of running these candidates, the presence on the ballot of such a large number poses a puzzle.

4.2.2 Electoral Competitiveness

We hypothesize first that a party's propensity to field a criminal candidate is conditional on how uncertain it is about the electoral outcome. Intuitively, a party that is sure of winning or losing is not likely to field a criminal because of the 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 (absolute value of the) percentage difference between the share of the votes obtained by the winning candidate and the share of votes obtained by the party of the candidate in question in the constituency in the same election. The reason that we construct separate measures of competitiveness for each candidate-party rather than a single constituency-level measure, as is common in single-member districts, lies with the large number of parties that operate in the typical Indian

parliamentary constituency. The party of each candidate will gauge the competitiveness of the race in terms of how close the party is to winning, not by how close the second runner up was. In the decision to field a self-reported criminal, we assume that each party's judgement depends on its own competitive position in the constituency. According to the theory postulated earlier, the sign on this variable should be negative; recall that smaller values of *COMP* indicate greater electoral competitiveness, and this should be associated with a greater likelihood of observing self-reported criminal candidates.

Even though India uses a simple plurality system (first past the post), it nonetheless has more than two political parties. That often makes it difficult to compare electoral distances across constituencies. For example, a party that wins by 3 percent in a two-party constituency may feel less safe than a party competing against five others that wins the seat with a 3 percent margin. To incorporate this consideration into our analysis, we control for the number of parties (*NUMBER*) in each constituency. If our theory is correct and more competitive electoral environments are more likely to see acknowledged criminals as candidates, the sign on this variable should be positive.

Our regression estimates also include three other measures that we believe affect electoral competitiveness. These are whether the candidate is the incumbent (*INCUMB*), whether the seat is reserved for the representative of a scheduled caste or tribe (*RESERVED*), and whether the candidate, is affiliated with a political party that is nationally organized or instead is local or regional in scope (*NATIONAL*).

INCUMB is coded 1 if a candidate is an incumbent and 0 otherwise. The standard argument is that incumbents enjoy electorally advantages 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 (2009)). 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. All else equal, we expect the sign on this variable to be negative, since incumbents should enjoy electoral advantages that make it unnecessary for them to resort to violence and intimidation in campaigning.

RESERVED is coded 1 if a seat is reserved for a member of a scheduled caste or tribe. A seat is designated reserved or not by each delimitation order setting 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 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 basis of 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 parliamentary 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 caste directly since information on caste is not included in the Indian census. But seeing if the selection of known criminals as candidates differs between reserved and non-reserved parliamentary seats is an indirect way to assess this.

Reserved seats may differ from other constituencies in three important ways affecting their propensity to see known criminals given parliamentary tickets. 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, also have more illiterate voters.¹⁹ Even though we control for literacy in our estimations, reserved constituencies may have other characteristics that make political parties there more likely to nominate criminals, justifying our decision to code them separately.

A second difference relates to barriers to entry in the political market. Although all voters are eligible

¹⁹Literacy rates are 52 percent in reserved seats compared to 56 percent generally.

to vote for reserved seats, only identified members of scheduled castes or tribes are eligible to stand as candidates. This feature may constitute a barrier to entry and make reserved seats less competitive.

The third possible difference between reserved and other seats is that aspects of political competition may be different in constituencies with large proportions of the population in scheduled castes and tribes. Politics in these constituencies may be “ethnified.” Banerjee and Pande (N.d.) argue that candidate quality is a function of the proportion of a constituency’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 constituency 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 constituencies with reserved seats, low caste parties will have lower quality candidates. Their model predicts therefore that more known criminals, which we assume is a low-quality candidate characteristic, 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.²⁰

We also control for whether the party with which the candidate is affiliated qualifies as a “national” party (*NATIONAL*). The Election Commission of India classifies six parties as national in 2004: the BJP, the 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 constituencies and across at least four states.

We include this variable because the decision making calculus of national parties about recruiting self-reported criminals onto the ballot is likely to be different from the decision making calculus of parties that are geographically more restricted. National parties make decisions across parliamentary constituencies and

²⁰This expectation seems also to be commonly shared in India.

are concerned with the extra-constituency ramifications of candidate selection. Parties that are organized only locally or in a state or two are more parochial in their decision making calculus. This may affect the likelihood of selecting known criminals to run. We expect the sign on this variable to be negative, since national parties should be more sensitive to the costs of fielding criminal candidates.

4.2.3 Vulnerability and Resilience to Intimidation and Violence

We proxy the vulnerability of the population to possible political intimidation and violence with a measure of literacy. *LIT* is an estimate of the literacy rate in each parliamentary constituency. As the literacy rate increases, we expect the likelihood that parties will list candidates facing criminal indictments to decrease. Hence, we expect the sign on this variable to be negative.

4.2.4 Demographic Control Variables

In addition to variables that may affect electoral competitiveness, our regressions include two demographic control variables for which we can estimate data at the level of the parliamentary constituency. *POP* is the population of each constituency. Although India's electoral system is single member, parliamentary constituencies are not equally sized. We hypothesize that, because the technology of intimidation is inherently limited in scope, the ability of criminals to intimidate voters will decrease with size. Therefore, we expect the sign on this variable to be negative.

URB is the percent of the population located in urban as opposed to rural areas in the constituency. We include this variable because we believe that urban voters, regardless of their level of literacy, have greater access to government protection from potential intimidation. We therefore expect that the sign on this variable will be negative; in more urban areas, we expect the likelihood of a self-reported criminal on the ballot to fall.

Summary statistics for the independent variables appear in Table 3, where we have also indicated the

sign that we expect for each. Electoral results for 2004 and 2009 are presented in Table 4 to give readers an overall idea of the extent of Lok Sabha competition. Finally, in Appendix D, we present the data on alleged criminal candidates by state. States are the natural units of analysis for Indian politics and knowledgeable observers will be especially interested in seeing the data at this level.

[Table 3 and Table 4 about here]

4.3 Analysis of the Probability a Party Selects a Self-Reported Criminal Candidate

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

$$CRIM = \beta + \beta * COMP_{it} + \beta * LIT_{it} + \beta * CONTROLS_{it} + \epsilon$$

where *COMP* is a measure of how close the candidate is to winning the seat; *LIT* is the proportion of the population that is literate; *CONTROLS* comprise *INCUM*, *RESERVED*, *NUMBER*, *NATIONAL*, *POP*, and *URB*; *i* indexes each constituency; *t* indexes each election; 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 5.

[Table 5 about here]

Our baseline model appears as Model 1. It tests the likelihood a party will field a criminal candidate as a function of only *COMP* and *LIT*, our main theoretically relevant variables. Model 2 adds the control variables *INCUM*, *RESERVED*, *NATIONAL*, *NUMBER*, *POP*, and *URB*. Model 3, finally, adds state fixed

effects to control for possible unobserved state-level heterogeneity.²¹ For each model, we run three separate regressions. The first is for candidates who ran in 2004, the second is for those who ran in 2009, and the third is a pooled analysis that includes both sets of observations. Interpretation of the odds ratios that we report 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.

In our baseline model, our principal independent variables, *COMP* and *LIT*, are each statistically significant at the 0.01 level. They retain significance even with the inclusion of statistical controls, though the inclusion of the state level fixed effects (Model 3) diminishes the statistical significance for *LIT* to the 0.10 level in 2004 and it loses statistical significance entirely for 2009 and in the pooled analysis. The signs on each variable is in the expected direction: greater electoral competitiveness increases the likelihood of observing a self-reported criminal candidate (recall that smaller values of *COMP* imply greater competitiveness, as the difference in the share of the vote received compared to the share received by the winning candidate shrinks).

Given the difficulties of comparing logit coefficients across different models, we also report marginal effects for the baseline model (Model 1) and for the model that includes control variables (Model 2). We vary *COMP* and *LIT* one standard deviation below and above their mean values. Thus *LIT* varies from 42 percent to 66 percent, and *COMP* varies from 14 percent to 44 percent. All other variables are set to their mean values and in the case of dummy variables, to zero. The results reported in Table 6 show that the substantive impact of our theoretically relevant independent variables is substantial. For example, the probability that a party fields a known criminal candidate is reduced by nearly 40 percent in the baseline 2004 model, 30 percent in 2009, and 34 percent in the pooled analysis with an increase in the literacy rate from a low of 42

²¹The coefficients for state effects are not included in the tables reporting our results. About half the state-level coefficients are statistically significant. 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 constituencies in a state are non-criminals, the state is dropped. In 2004, 11 states and six union territories drop out of the analysis reported in Model 3; in 2009, when more self-reported criminals are scattered across India, five states and the same number of union territories drop out. The dropped states and union territories are all quite small.

percent to a high of 66 percent. The relationship between *LIT* and the probability of fielding a candidate under indictment is even stronger when control variables are added for the 2004 model (52 percent), though weaker in 2009 (17 percent). *COMP* also has a substantively important impact. For the baseline 2004 model, a party that is 14 percent away from winning or losing the seat is 40 percent more likely to field a self-reported criminal candidate than a party that is 44 percent away from winning or losing. The marginal effect increases to 50 percent in 2009. A second noteworthy result evident in the results reported in Table 6 is that parties in general are more likely to field acknowledged criminal candidates in 2009 than in 2004. Under Model 2, parties are about 55 percent more likely to field a criminal candidate in 2009 than in 2004.

[Table 6 about here]

4.4 Analyses of the Number of Self-Reported Criminals on the Ballot

We have documented that the probability that a political party offers a ticket to a self-reported criminal is significantly greater under more competitive electoral conditions and as the literacy rate declines. To study whether a similar pattern obtains for the number of self-reported criminal candidates who are given tickets, we estimate poisson regressions that parallel those just reported. Because the unit of analysis is now the parliamentary constituency, all variables are measured at this level. *LIT* and most control variables (*NUMBER*, *POP*, and *UBR*) are already measured at the level of the constituency or convert to that level in a natural way (*INCUM* is now defined as whether an incumbent candidate runs in the constituency). Because there is no obvious way to convert the measure of national party status to the constituency level, we omit this control. *COMP*, which in the regressions reported in the prior section is measured for each party as the difference between the share of votes won by its candidate and the share of votes received by the winner in the constituency, is conceptually more difficult to convert to the constituency level. We operationalize the competitiveness measure as the difference in the share of the vote received by the winner and the first runner-up; this is, however, a rough gauge that does not capture the strategic incentives of

individual political parties. As the results reported in Table 7 show, results for *COMP* measured in this way fail to reach conventionally defined levels of statistical significance although the variable exhibits the correct sign in 2009 and for the pooled results. The second of our theoretically relevant independent variables, *LIT*, behaves as expected, as the results reported in the same table demonstrate. Literacy is consistently statistically significant and negative, meaning that as the literacy rate declines, the number of self-reported criminal candidates in each district increases. We interpret this as corroboration of our hypothesis that as the population become more vulnerable to intimidation, more known criminals are selected by parties as political candidates in the constituency.

[Table 7 about here]

4.5 Analysis of Whether Self-Reported Criminal Candidates Depress Electoral Turnout

We now study whether self-reported criminal candidates depress electoral turnout. If our theory of political criminality in India is correct, acknowledged criminal are selected by political parties as candidates because of their skills in utilizing technologies of violence to intimidate voters. If this is the case, then we should observe lower electoral turnout where such candidates appear on the ballot. In 2004, turnout varied across Indian parliamentary constituencies from a low of 15 percent of the electorate to a high of 91 percent; in 2009, the variation was slightly less. To assess if some of this variation is driven by the presence of self-reported criminals on the ballot, we estimate a series of constituency-level regressions of the following sort:

$$TURNOUT = \beta + \beta * CRIM_{it} + \beta * COMP_{it} + \beta * LIT_{it} + \beta * CONTROLS_{it} + \epsilon$$

where *CRIM* is coded 1 if at least one self-reported criminal is given a ticket in the constituency; *COMP* is a measure of how close the runner up's vote share is from the vote share received by the winner of the seat; *LIT* is the proportion of the population that it literate; *CONTROLS* comprise *INCUM*, *RESERVED*, *NUMBER*,

POP, and *URB*; *i* indexes the constituency; *t* indexes the election; and ϵ is an error term. As is the case for the constituency-level regressions reported in the prior section, the variable *NATIONAL* is not included in these models; other variables are measured in the same way as those just reported. Studies of turnout in India corroborates the importance of such standard explanatory variables as electoral competitiveness [cite]. We utilize an OLS estimator and report results in Table 8.

[Table 8 about here]

As the results reported in Table8 document, turnout is significantly reduced by the presence of at least one self-reported criminal candidate on the ballot in most of the models studied. In a simple t-test of the 2004 data, turnout is about 3 percent higher in constituencies with no acknowledged criminal on the ballot than in those with a single such individual ($p = .0048$) and in 2009, the difference grew to 5 percent ($p = .0002$). Our multiple regression results show that other variables largely affect turnout rates as expected by the literature [cite]. Turnout is significantly higher when literacy increases and in rural as opposed to urban areas. The two variables that do not behave as expected by our theory are those proxying for electoral competitiveness. Results reported show that as the margin of victory of the winner over the runner-up falls, so does turnout and as the number of parties listing candidates falls, so does turnout. Perhaps these results are due to the noise with which electoral competitiveness is measured at the constituency level in a multiparty setting; with so many parties contesting elections in India, the strategic use of electoral violence will vary by party even within the same constituency and this measure is not sensitive to that. Perhaps as a result, the signs on our measures of competitiveness for turnout are the reverse of expected. More importantly for present purposes, *CRIM* reduces turnout, and results are generally significant More importantly for present purposes, *CRIM* reduces turnout, and results are generally statistically significant.

4.6 Analyses of the Number of Self-Reported Criminals in Each Constituency

Our theory of how electoral competitiveness affects the decisions by parties to select criminal candidates leads us to expect to observe varying numbers of such candidates across parliamentary constituencies. Under the mixed strategy equilibrium conditions discussed in our formal model, we may observe constituencies with no self-reported criminal candidates, with one such candidate, with two, or with more than two. Alternate theories, such as Vaishnav (2010), expect instead to observe criminal candidates in pairs; where one party selects a self-reported criminal, a close competitor should do so as well. In Figure 9, we report the frequency with which we observe different numbers of candidates facing criminal indictment in India's parliamentary constituencies in 2004 and in 2009.

[Figure 9 about here]

The data summarized in Figure 9 show that many constituencies see only a single self-reported criminal on the ballot; many more, in fact, than see two such contestants. Many constituencies are home to more than two known criminal legislative candidates. The distribution of these candidates conforms to the expectations of the formal model presented in this paper and fails to corroborate alternate theories that expect self-reported criminals to appear in pairs.

4.7 Regression Discontinuity Analysis of the Impact of Self-Reported Criminals on Incumbents

We now turn our attention to the impact criminals play in the electoral fates of incumbents. We examine whether incumbents who face self-reported criminal opponents are less likely to win reelection than incumbents who do not. That is, are indicted criminals eroding incumbency advantage in India?

We employ a regression discontinuity design to estimate the incumbency (dis)advantage in the 2009 Lok Sabha elections.²² Incumbency effects are estimated for all incumbents (baseline), for incumbents who face

²²Because the design requires electoral results for all candidates from the previous election, we are able to estimate incumbency effects only for 2009 and not for 2004 or any earlier election.

a self-reported criminal opponent, and for incumbents who do not face a self-reported criminal opponent.

The regression discontinuity design is commonly used to assess the existence and extent of incumbency advantages in elections. 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 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 than non-incumbents, and Uppal (2009) estimates that since 1991 incumbent members of state assemblies are 25 percentage points less likely to win reelection than non-incumbents.

Generally speaking, a regression discontinuity analysis compares the probability of winning an election across two groups: candidates who barely won 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 very 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 involved in regression discontinuity. The first is to estimate the probability that a candidate wins an election conditional on his electoral performance in the prior 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; this is done 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 each near-loser.²³

Second, we restrict the sample space to ensure that the two groups of candidates (incumbents and non-

²³In contrast, Linden (2004) uses a non-parametric design and Uppal (2009) utilizes logistic regression with a fourth order polynomial for margin of victory as well as interactions and fixed effects.

incumbents) are as equivalent as possible with respect to observed characteristics. There is a trade-off between sample size and group comparability. If the margin of victory cutoffs are set too narrowly, then the likelihood of having comparable groups is high but at the cost of a small sample size. Conversely, if the margin of victory cutoffs are set too widely, the likelihood of having comparable groups is lower, potentially biasing the results, but the sample size larger. Since, due to the absence of prior data, our analysis only covers one election period, it is particularly sensitive to sample size problems. Though this introduces bias, we show later that the direction of the bias is towards overestimating an incumbency advantage.

The third step in the regression discontinuity analysis is to estimate the difference in election probabilities for incumbents and non-incumbents at the discontinuity (that is, where the electoral distance is zero). This is done by subtracting the probability of being elected in 2009 for non-incumbents from the probability of election for incumbents. This is especially effective when done graphically by plotting the election probability in 2009 on the y-axis as a function of electoral distance in 2004, which is plotted on the x-axis, with the gap in election probabilities at the discontinuity representing the incumbency effect. If there is no incumbency effect, there is no gap between the two groups. If there is a negative effect, then the line representing the probability of being elected in 2009 is “lower” for incumbents than non-incumbents.

Our baseline estimate for all incumbents is calculated by estimating the following equation:

$$Pr(GOTSEAT_{2009i} = 1) = L(\beta_0 + \beta_1 * COMP_{2004i} + \epsilon_i)$$

where L is a logistic function; *GOTSEAT* is a dummy variable that takes a value of 1 if the candidate won a seat in 2009 and 0 otherwise; *COMP* is the electoral distance a candidate was from winning or losing in 2004, and *i* are the constituencies.

In order to estimate the impact on incumbency of facing a self-reported criminal opponent, the basic equation just outlined is expanded to include a dummy variable indicating whether a candidate faced a known criminal opponent. We also include a variable measuring the interaction of a criminal opponent and

electoral marginality (*COMP*). Our theory of criminal intimidation implies that the electoral environment of constituencies with self-reported criminal candidates is fundamentally different from constituencies without them. As a result, we hypothesize that constituencies with known criminals on the ballot may have different “slopes” with respect to the relationship between electoral performance in 2004 and the probability of winning in 2009. We capture this expected difference with the interaction term. Thus we estimate:

$$Pr(GOTSEAT_{2009i} = 1) = L(\beta_0 + \beta_1 * COMP_{2004i} + \beta_2 * CRIMOPP_{2009i} + \beta_3 * COMP_{2004i} * CRIMOPP_{2009i} + \epsilon_i)$$

where *CRIMOPP*₂₀₀₉ is a dummy variable which takes the value of 1 if a candidate faced an opponent with a criminal record and 0 otherwise; and *COMP* * *CRIMOPP* is the interaction term between *COMP*₂₀₀₄ and *CRIMOPP*₂₀₀₉.

4.8 Data

Of the approximately 8,000 candidates who ran for the Lok Sabha in 2009, 572 had also been listed on the ballot in 2004. Of these 572, we restrict attention to candidates within 25 percent of winning or losing in 2004, thereby reducing our pool of candidates to 398.²⁴ Bias-free results require that both incumbents and non-incumbents be balanced on observed as well as unobserved traits. Since in order to obtain a large enough sample size, we set a wide margin of victory as the discontinuity cutoff (25 percent), our groups are not comparable along certain important dimensions. Table 10 compares incumbents and non-incumbents across a variety of dimensions: the probability of winning in 2009; vote shares received in 2004 and 2009; proportions facing criminal indictment; proportions with an opponent facing criminal indictment in 2009; and proportions affiliated with the Congress or the BJP.

²⁴We also ran an analysis with a cutoff of 10 percent. In that analysis, the incumbency disadvantage for all incumbents disappears but our main result remains; incumbents who face a known criminal opponent are at a disadvantage whereas incumbents who face other opponents are not.

[Table 10 about here]

As the data presented in the table document, incumbents on average received higher vote shares in 2004 than non-incumbents. This implies that incumbents as a whole may have had better political skills than their non-incumbent rivals. The main implication for the analysis that follows 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. The bias therefore works against the argument that Indian incumbents are disadvantaged electorally.

4.9 Results

Table 11 summarizes the results of the regression discontinuity analyses. Results are graphically presented in Figures 4 and 5. Results of the baseline analysis are consistent with those reported in previous studies. We find that legislative incumbents face an electoral disadvantage in India. They confront a 6 percentage point disadvantage in winning office in 2009, which translates into being 15 percent less likely to win.²⁵

[Table 11 and Figures 4 and 5 about here]

Figures 4 and 5 represent the main results of the regression discontinuity analysis. Figure 4 shows that in 2009, incumbents and near-losers from 2004 enjoyed similar probabilities of election. There is a noticeable but small incumbency disadvantage, represented by the break in the line at the discontinuity and the slight drop in the line to the right of 0, representing near winners in 2004. The source of this incumbency disadvantage is clarified in Figure 5. The lefthand panel shows that candidates facing a self-reported criminal opponent have an incumbency disadvantage (8 percentage points) whereas those facing an opponent with no known criminal record enjoy an incumbency advantage (3 percentage points). The two figures clearly show that the entire incumbency (dis)advantage is systematically linked to whether an incumbent faces off

²⁵Our results do not exactly match those reported by Linden (2004) or Uppal (2009), 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 (2009), we do not include control variables in our estimations. 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.

against a known criminal or not.

These findings document that the well known disadvantage of the Indian legislative incumbent is a product of whether incumbents face known criminals as electoral opponents. That fully three-quarters of India's electoral districts saw at least one acknowledged criminal on the ballot in 2009, representing a massive diffusion of candidate criminality over 2004, suggests that self-reported criminals are driving out non-criminals at the national level of the Indian polity.

5 Conclusions

Self-reported criminal candidates are more likely to appear on the ballot in India in constituencies with more illiterate voters and when a party faces greater electoral competitiveness. Their selection reduces voter turnout, which we interpret as corroboration of our theory that acknowledged criminals utilize violent tactics to intimidate opposition supporters and keep them at home on election day. Finally, self-reported criminals appear to be successfully driving Indian incumbent legislators out of office, which may explain the proliferation of known criminals as national candidates.

There is much that we still do not know about the use of violence as an electoral strategy.

References

- 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.
- Besley, Timothy and Robin Burgess. 2002. "The Political Economy of Government Responsiveness: Theory and Evidence from India." *Quarterly Journal of Economics* 117(4):1415–1451.
- Bhavnani, Rikhil R. 2011. "Corruption Among India's Politicians: Evidence from Unusual Data." Unpublished paper.
- Bobonis, Gustavo J. and Luis R. Cámara Fuertes. 2009. "Does Exposing Corrupt Politicians Reduce Corruption?" Unpublished paper.
- Bratton, Michael. 2008. Vote Buying and Violence in Nigerian Election Campaigns. Working Paper 99 Afrobarometer, Cape Town, South Africa: .
- Brollo, Fernanda. 2009. "Who is Punishing Corrupt Politicians — Voters or the Central Government? Evidence from the Brazilian Anti-Corruption Program." Unpublished paper.
- 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. Technical report Centre Interuniversitaire sur le Risque, les Politiques Economiques et l'Emploi.
- 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 London: .
- 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.
- Jennings, Colin and Alan Hamlin. N.d. "Expressive Political Behaviour: Foundations, Scope and Implications." *British Journal of Political Science*. Forthcoming.
- 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.
- Uppal, Yogesh. 2009. "The Disadvantaged Incumbents: Estimating Incumbency Effects in Indian State Legislatures." *Public Choice* 138(1):9–27.
- Vaishnav, Milan. 2010. "The Market for Criminality: Money, Muscle and Elections in India." Unpublished paper.
- Wilkinson, Steven I. 2004. *Votes and Violence: Electoral Competition and Ethnic Riots in India*. New York: Cambridge University Press.
- 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 and 2009 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 taken 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 Election Commission of India.

Independent vote shares refers to the share of the vote in each constituency won by candidates with no partisan affiliation. Data from Election Commission of India.

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

National parties are designated 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 scheduled tribes. 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 Indian census data.

Population in each parliamentary constituency 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 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 2001 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. Calculated by taking the urban population in each administrative district and dividing by the district's total population. Data available from the 2001 Indian census.

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.

Procedure used for matching 2004 and 2009 parliamentary constituencies with 2001 census districts:

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 reaggregated to Gangana-gar, 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 reaggregated to Perambular. In the union territories, Andaman and Nicobar were combined as were Daman and Diu.

For the 2009 elections, we replicate the above procedures using the new Delimitation Order. However, we were unable to match census data to electoral data for 25 constituencies. The delimitation order did not have information for 12 constituencies in the state of Assam. In addition, several new

administrative districts were created after 2004. In Bihar, the administrative district of Arwal was split from Jehenabad after 2001. In Karnataka, Anuppur, Ashok Nagar and Burhanpur are new districts created after 2001. In Manipur, all nine administrative districts either underwent name changes or had new boundaries drawn. In Tamil Nadu, Krishnagiri was split from Dharmapuri and Ariyalur was split from Perambalur. Finally, in West Bengal, the administrative districts of Paschim Medinipur and Purbo Medinipur were combined into Medinipur. In addition, the 20xx Delimitation Order did not contain the electoral boundaries for any of the seven union territories. As a result, we were unable to merge census data for 25 electoral constituencies for 2009, all of which were dropped from analysis.

Appendix B: Specification of the Formal Model

This appendix provides a full characterization of the equilibrium of the candidate fielding game. The seat is safe if

$$\Delta - \Delta_B > 0 \Leftrightarrow \gamma < \hat{\gamma} - \frac{\theta_B t}{\theta_A(\theta_B - c)} \equiv \underline{\gamma} \quad (3)$$

and competitive otherwise. When $\gamma < \underline{\gamma}$, party B cannot swing the election and it is dominant strategy for that party not field a criminal. When $\gamma \geq \underline{\gamma}$, the seat is competitive, and the regime depends on whether party A can defend the seat or not. The seat can be defended if

$$\Delta - \Delta_B + \gamma \Delta_A > 0 \Leftrightarrow \gamma < \frac{\theta_B(\theta_A - c - t)}{\theta_A(\theta_B - c - t)} \equiv \bar{\gamma} \quad (4)$$

and is indefensible otherwise.

The normal form representation of the candidate fielding game is

A/B	Criminal	Non-criminal
Criminal	$M - \mu$ $-\mu$	0 $M - \mu$
Non-criminal	$M - \mu$ 0	M 0

where the column player is party A and the row player is party B . We can rule out each of the four potential pure strategy Nash equilibrium by showing that at least one player has a profitable deviation. First, suppose that both parties field a criminal. Given that party A fields a criminal, party B is better off not doing so, as $-\mu < 0$. Second, suppose that both parties field a non-criminal. Given that party A fields a non-criminal, party B will deviate and field a criminal because $M - \mu > 0$. Third, suppose that party A fields a criminal but party B does not. Given that party B does not field a criminal, party A will deviate and field a non-criminal because $M > M - \mu$. Fourth, suppose that party B fields a criminal but party A does not. Given that party B fields a criminal, party A will deviate and also field a criminal because $M - \mu > 0$. To find the mixed strategy equilibrium, let λ_k be the probability that party k fields a criminal. If a party is using a mixed strategy at equilibrium, then it should have the same expected payoff from the pure strategies over which it is mixing. For party A , this means that the expected payoff of fielding a criminal, $\lambda_B(M - \mu) + (1 - \lambda_B)(M - \mu)$, must be equal to the expected payoff of fielding a non-criminal, $\lambda_B 0 + (1 - \lambda_B)M$. This implies that

$$\lambda_B = \frac{\mu}{M}. \quad (5)$$

For party B , the expected payoff of fielding a criminal, $\lambda_A(-\mu) + (1 - \lambda_A)(M - \mu)$ must equal the ex-

pected payoff of fielding a non-criminal, $\lambda_A 0 + (1 - \lambda_A) 0$. This implies that

$$\lambda_A = \frac{M - \mu}{M}. \quad (6)$$

The seat cannot be defended if

$$\Delta - \Delta_B - \gamma \Delta_A < 0 \Leftrightarrow \gamma > \bar{\gamma}. \quad (7)$$

To make sure that $\bar{\gamma} < \hat{\gamma}$, we require that $\theta_B > \theta_A$. If this fails, then the seat is always defensible. The normal form of the candidate fielding game is

A/B	Criminal	Non-criminal
Criminal	$\begin{matrix} -\mu \\ M - \mu \end{matrix}$	$\begin{matrix} 0 \\ M - \mu \end{matrix}$
Non-criminal	$\begin{matrix} M - \mu \\ 0 \end{matrix}$	$\begin{matrix} M \\ 0 \end{matrix}$

We observe that fielding a criminal is a dominant strategy for party B . Given this, the best response for party A is not to field a criminal and the unique pure strategy Nash equilibrium is that party B fields a criminal and party A does not. Since dominated strategies are never used in mixed Nash equilibria, we can rule out mixed strategy equilibrium in this regime.

Appendix C: Results using Alternate Measure of Competitiveness

Appendix D: State-Level Information on Candidates

Appendix E: Robustness of Results

Table 1: Criminal Candidates and Electoral Outcomes, Fourteenth and Fifteenth 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	0.22
Criminals	116	0	0
Total	2,385	5	0.21
Independent candidates, 2009			
Not criminals	3,581	8	0.22
Criminals	250	1	0.4
Total	3,831	9	0.23

Notes: Criminal candidates defined as those who report having been convicted of or currently facing criminal charges in their affidavit filed with candidacy papers.

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

<i>Party</i>	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 and Expected Signs of Independent Variables

	Mean	SD	Min.	Max.	Exp. Sign
Electoral year 2004					
COMP	31.23	18.28	.06	72.58	-
LIT	54.10	11.97	25.86	85.42	-
INCUM	.13	.34	0	1	-
RESERVED	.21	.41	0	1	+
NATIONAL	.44	.50	0	1	-
NUMBER	5.52	2.52	0	15	+
POP	1,957,936	538,101.6	60,595	5,410,783	-
URB	27.57	20.86	3.48	100	-
Electoral year 2009					
COMP	31.93	16.72	.04	78.24	-
LIT	53.99	12.00	25.42	85.29	-
INCUM	.07	.25	0	1	-
RESERVED	.33	.47	0	1	+
NATIONAL	.38	.49	0	1	-
NUMBER	7.84	3.04	1	20	+
POP	1,930,203	35,8941.7	186,189	4,013,609	-
URB	27.89	20.72	3.48	100	-

Notes: COMP is the absolute value of the difference between the share of vote won by the candidate and the winner's share; LIT is the percent of population that is literacy; INCUM is whether the candidate is an incumbent; RES is whether seat is reserved for a scheduled caste or tribe representative; NAT is affiliation with national party; NUM is number of other partisan-affiliated candidates; POP is total population; URB is percent of total population in urban areas. LIT, RESERVED, NUMBER, POP, and URB are measured at the level of the parliamentary constituency. COMP, INCUM, and NATIONAL are measured at the level of the individual candidate.

Table 4: Fourteenth and Fifteenth Lok Sabha Electoral Results by Party

Party	No. of Candidates	No. of Seats Won	Percent of Cands. Winning	Percent of Seats Won
2004 Electoral Results				
BJP	364	138	25.4	37.9
BSP	435	19	3.5	4.4
CPI	34	10	1.8	29.4
CPM	69	43	7.9	62.3
INC	417	145	26.7	34.8
NCP	32	9	1.7	28.1
Other	1,699	174	32.0	10.2
Total	3,050	538	99.1	17.6
2009 Electoral Results				
BJP	433	116	21.36	26.79
BSP	500	21	3.87	4.20
CPI	56	4	0.74	7.14
CPM	82	16	2.95	19.51
INC	440	206	37.94	46.82
NCP	68	9	1.66	13.24
RJD	44	4	0.74	9.09
AITC	27	19	3.50	70.37
DMK	22	18	3.31	81.82
JD(U)	27	20	3.68	74.07
SHS	22	11	2.03	50.00
SP	95	23	4.24	24.21
Other	2,423	67	12.34	2.77
Total	4,239	534	98.34	12.6

Notes: BJP: Bharatiya Janata Party. BSP: Bahujan Samaj Party. CPI: Communist Party of India. CPM: Communist Party of India (Marxist). INC: Indian National Congress. NCP: National Congress Party. RJD: Rashtriya Janata Dal. AITC: All India Trinamool Congress. DMK: Dravida Munnetra Kazhagam. JD(U): Janata Dal (United). SHS: Shivsena. SP: Samajwadi Party. Figures excludes independent candidates.

Table 5: Logit Estimates of the Determinants of Self-Reported Criminal Candidates in Parliamentary Constituencies, Fourteenth and Fifteenth Lok Sabha Elections

Election	Model 1		Model 2		Model 3		
	2004	2009	2004	2009	2004	2009	
	Model 1a	Model 1b	Model 2a	Model 2b	Model 3a	Model 3b	
	Pooled	Pooled	Pooled	Pooled	Pooled	Pooled	
	Model 1c	Model 2c	Model 3c	Model 3c	Model 3c	Model 3c	
Competitiveness	0.981*** (0.003)	0.973*** (0.003)	0.982*** (0.004)	0.977*** (0.003)	0.978*** (0.004)	0.979*** (0.003)	0.979*** (0.002)
Literacy rate	0.977*** (0.005)	0.983*** (0.004)	0.967*** (0.007)	0.986*** (0.005)	0.981* (0.011)	0.997 (0.008)	0.990 (0.006)
Incumbent candidate			1.295 (0.209)	1.188 (0.188)	1.256 (0.212)	1.264 (0.204)	1.208* (0.138)
Reserved seat			0.641*** (0.098)	0.661*** (0.066)	0.625*** (0.101)	0.707*** (0.074)	0.735*** (0.063)
National party			0.998 (0.125)	1.358*** (0.134)	1.048 (0.139)	1.367*** (0.138)	1.263*** (0.100)
Number of other parties			1.022 (0.026)	1.014 (0.017)	1.000 (0.038)	0.972 (0.020)	1.018 (0.015)
Population			1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
Percent of pop. in urban areas			1.006 (0.004)	0.998 (0.003)	1.001 (0.005)	1.000 (0.004)	1.001 (0.003)
Constant	0.812 (0.213)	0.983 (0.204)	1.460 (0.680)	0.534 (0.230)	0.323 (0.225)	0.195*** (0.111)	0.255*** (0.105)
State fixed effects	NO	NO	NO	NO	YES	YES	YES
Observations	3,050	4,052	3,050	4,052	2,828	4,018	7,014

Notes: Standard errors in parenthesis. State fixed effects not reported.
 *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Marginal Effects of Electoral Competitiveness and Literacy on the Probability a Party Selects a Self-Reported Criminal Candidate

Model	At baseline	Literacy = 42%	Literacy = 66%	Marginal effect	Margin = 14%	Margin = 44%
Baseline 2004	10.9	14.1	8.5	-39.7	14.7	8.8
Baseline 2009	14.1	16.8	11.8	-29.6	21.0	10.4
Baseline pooled	12.8	15.6	10.4	-33.6	18.1	9.8
With controls 2004	11.5	15.5	8.0	-48.5	15.1	9.3
With controls 2009	14.2	16.2	12.1	-25.3	20.1	10.9
With controls pooled	13.0	15.9	10.1	-36.3	17.8	10.2

Notes: Each cell represents the probability, in percent, that a party fields a criminal candidate. Covariate levels are set to their mean or zero for dummy variables unless noted otherwise.

Table 8: OLS Estimates of the Determinants of Electoral Turnout in Parliamentary Constituencies, Fourteenth and Fifteenth Lok Sabha Elections

	2004			2009				
	Model 1a	Model 1b	Model 1c	Model 1d	Model 2a	Model 2b	Model 2c	Model 2d
Criminal on ballot	-3.563*** (1.016)	-2.778*** (1.005)	-1.271 (0.927)	-0.139 (0.604)	-4.561*** (1.238)	-3.534*** (1.193)	-1.678 (1.085)	0.415 (0.595)
Competitiveness		0.108** (0.048)	0.087** (0.044)	-0.021 (0.030)		-0.020 (0.063)	-0.063 (0.056)	-0.043 (0.031)
Literacy rate		0.189*** (0.042)	0.348*** (0.050)	0.193*** (0.046)		0.335*** (0.048)	0.429*** (0.058)	0.100** (0.045)
Incumbent candidate			-0.583 (1.054)	-0.725 (0.648)			1.649 (1.019)	0.367 (0.536)
Reserved seat			-1.681 (1.095)	-1.302** (0.655)			-0.131 (1.082)	-1.077* (0.579)
Number of parties			-1.235***	-0.458***			-0.908***	-0.337***
Population			(0.220)	(0.173)			(0.209)	(0.124)
Percent of pop. in urban areas			-0.000 (0.000)	-0.000 (0.000)			-0.000*** (0.000)	-0.000*** (0.000)
Constant	60.525*** (0.670)	48.455*** (2.462)	54.710*** (3.685)	67.997*** (2.791)	61.918*** (0.982)	43.101*** (2.954)	64.717*** (4.584)	81.782*** (2.983)
State fixed effects	NO	NO	NO	YES	NO	NO	NO	YES
Observations	543	543	543	543	519	519	519	519
R-squared	0.022	0.071	0.258	0.770	0.026	0.112	0.316	0.831

Notes: CRIM is whether at least one self-reported criminal candidate appears on the ballot; COMP is the absolute value of the difference between the share of vote received by the winner and the first runner-up; LIT is the percent of population that is literate; INCUM is whether any candidate is an incumbent; RES is whether seat is reserved for a scheduled caste or tribe representative; NUM is number of partisan-affiliated candidates; POP is total population; URB is percent of total population in urban areas. All variables measured at the level of the parliamentary constituency. Standard errors in parentheses. State fixed effects not reported.

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

Table 9: Frequency Distribution of Number of Self-Reported Criminal Candidates Observed in Each Parliamentary Constituency, Fourteenth and Fifteenth Lok Sabha Elections

Number	Freq. 2004	Perc. 2004	Freq. 2009	Perc. 209
0	271		137	
1	154		168	
2	64		89	
3	35		59	
4	11		30	
5	6		22	
6	0		6	
7	2		2	
8	0		3	
9	0		2	
10	0		1	

Table 10: Comparison of Characteristics of Candidates within 25 Percent of Winning or Losing a Seat in 2004

	Won in 2004	Lost in 2004	Difference
Proportion winning in 2009	0.45 (0.03)	0.33 (0.04)	0.12** (2.28)
Vote share 2009	0.37 (0.01)	0.29 (0.01)	0.07*** (5.41)
Vote share 2004	0.46 (0.00)	0.33 (0.01)	0.13*** (14.69)
Self-reported criminal indictment	0.23 (0.03)	0.16 (0.03)	0.07 (1.60)
Criminal opponent 2009	0.68 (0.03)	0.73 (0.04)	-0.05 (-1.13)
Proportion of INC candidates	0.34 (0.03)	0.34 (0.04)	0.00 (0.04)
Proportion of BJP candidates	0.24 (0.03)	0.28 (0.04)	-0.04 (0.91)
Number of obs.	253	145	

Notes: Standard errors in parenthesis except for the differences column, which reports t-stats in parenthesis. INC: Indian National Congress. BJP: Bharatiya Janata Party.

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

Table 11: Summary of Regression Discontinuity Results

	2004 Incumbents	2004 Non-incumbents	Difference	Perc. difference
All observations	0.31	0.37	-0.06	-19
Candidates not reporting indictments	0.33	0.41	-0.08	-24
Candidates reporting indictments	0.28	0.34	-0.06	-21
Candidate opponent does not report indictment	0.37	0.32	0.05	14
Candidate opponent reports indictments	0.29	0.38	-0.09	-31
Candidate not reporting indictment whose opponent reports indictments	0.29	0.46	-0.17	-59
Candidate reporting indictments whose opponent reports indictments	0.32	0.26	0.06	19

Figure 1: Degree of Electoral Competition and Equilibrium Candidate Choice

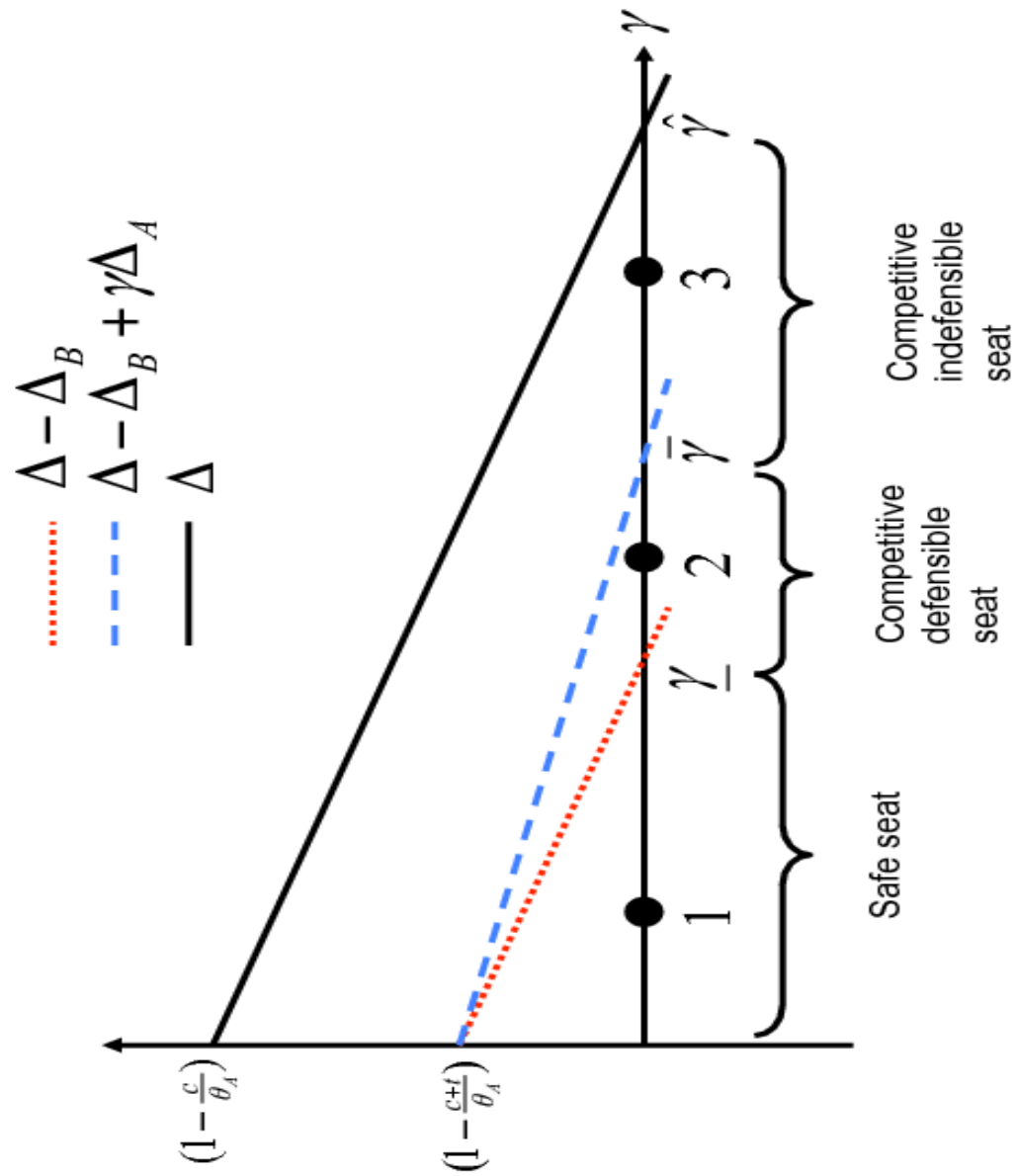
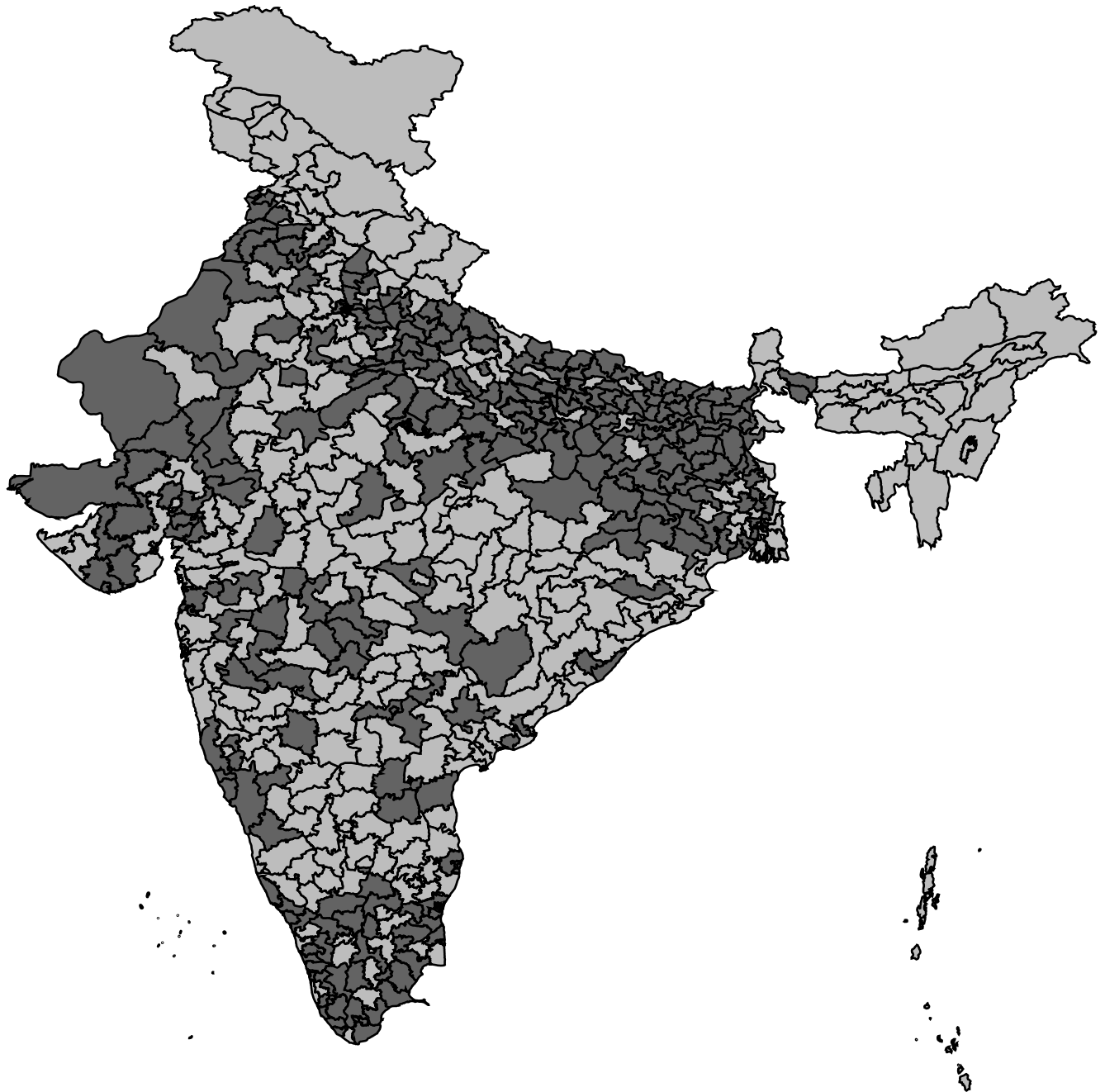



Figure 2: Map of Self-Reported Criminal Candidates in Parliamentary Constituencies for Elections to the Fourteenth Lok Sabha (2004)



Notes: Darker constituencies represent those with at least one self-reported criminal.

Figure 3: 2004 Affidavit of Afajal Ansari, Candidate to the Lok Sabha from Ghazipur (UP), p. 3



Serial No. 055
Date 5-4-2004

शपथ-पत्र
प्रारूप-26 (नियम-4क)

समक्ष : रिटर्निंग आफिसर,
46, गाजीपुर संसदीय निर्वाचन क्षेत्र।

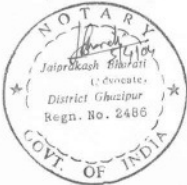
मैं अफजाल अंसारी पुत्र सुभहानुल्लाह अंसारी आयु 50 वर्ष जो निवासी-जमालपुर मुहम्मदाबाद, जिला-गाजीपुर का निवासी हूँ और उपरोक्त निर्वाचन में अभ्यर्थी हूँ, सत्यनिष्ठा से प्रतिज्ञा करता हूँ। शपथ पर निम्नलिखित कथन करता हूँ :-

मैं ऐसे किसी लम्बित मामले में दो वर्ष या अधिक के कारावास से दण्डनीय किसी अपराध (अपराधों) का अभियुक्त नहीं हूँ जिसमें सक्षम अधिकारिता वाले न्यायालय द्वारा आरोप विरचित किया गया है/किए गए हैं। (यदि अभिसाक्षी ऐसे किसी अपराध (अपराधों) का अभियुक्त है तो वह निम्नलिखित जानकारी प्रस्तुत करेगा)।

- 1- मामला / प्रथम सूचना रिपोर्ट संख्या / संख्याएं : 260 / 2001
- 2- पुलिस थाना (थाने) मुहम्मदाबाद, जिला (जिले) गाजीपुर, राज्य - उत्तर प्रदेश
- 3- संबंधित अधिनियम (अधिनियमों) की धारा (धाराएं) और उस अपराध (अपराधों) का संक्षिप्त विवरण जिसके (जिनके) लिए अभ्यर्थी आरोपित किया गया है - धारा 147, 148, 353 आई.पी.सी. व 3 पी.डी. पी.पी. एक्ट व 7 सी.एल.ए. एक्ट
- 4- न्यायालय जिसके (जिनके) द्वारा आरोप (आरोपों) की विरचना की गयी - सिविल जज जूनियर डिवीजन, मुहम्मदाबाद, गाजीपुर।
- 5- तारीख (तारीखें) जिनको आरोप विरचित किया गया था/किए गए थे - 09-08-2001
- 6- क्या सभी या कोई कार्यवाही किसी सक्षम अधिकारिता वाले न्यायालय द्वारा रोकी गयी है - नहीं

मुझे किसी अपराध (अपराधों) लोक प्रतिनिधित्व अधिनियम 1951 (1951 का 43) की धारा 8 की उपधारा (1) के अन्तर्गत में निर्दिष्ट या उपधारा (3) के अन्तर्गत आने वाले किसी अपराध (अपराधों) से भिन्न के लिए सिद्धांतों में कोई दण्डित किया गया है और एक वर्ष या अधिक के लिये कारावास से दण्डादिष्ट नहीं किया गया है।

NOTARY
GHAZIPUR (U.P.)



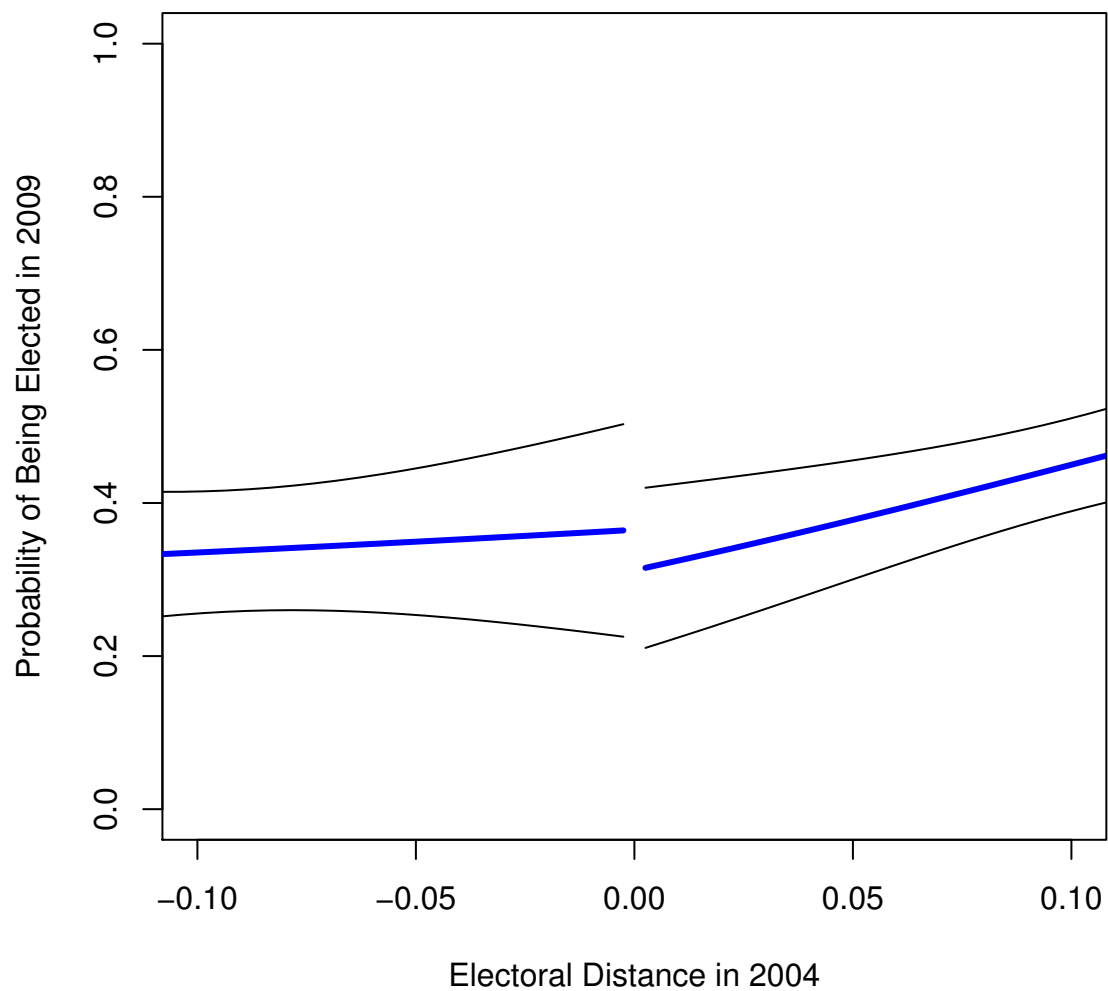
Handwritten Signature

Handwritten Signature

25/4/04

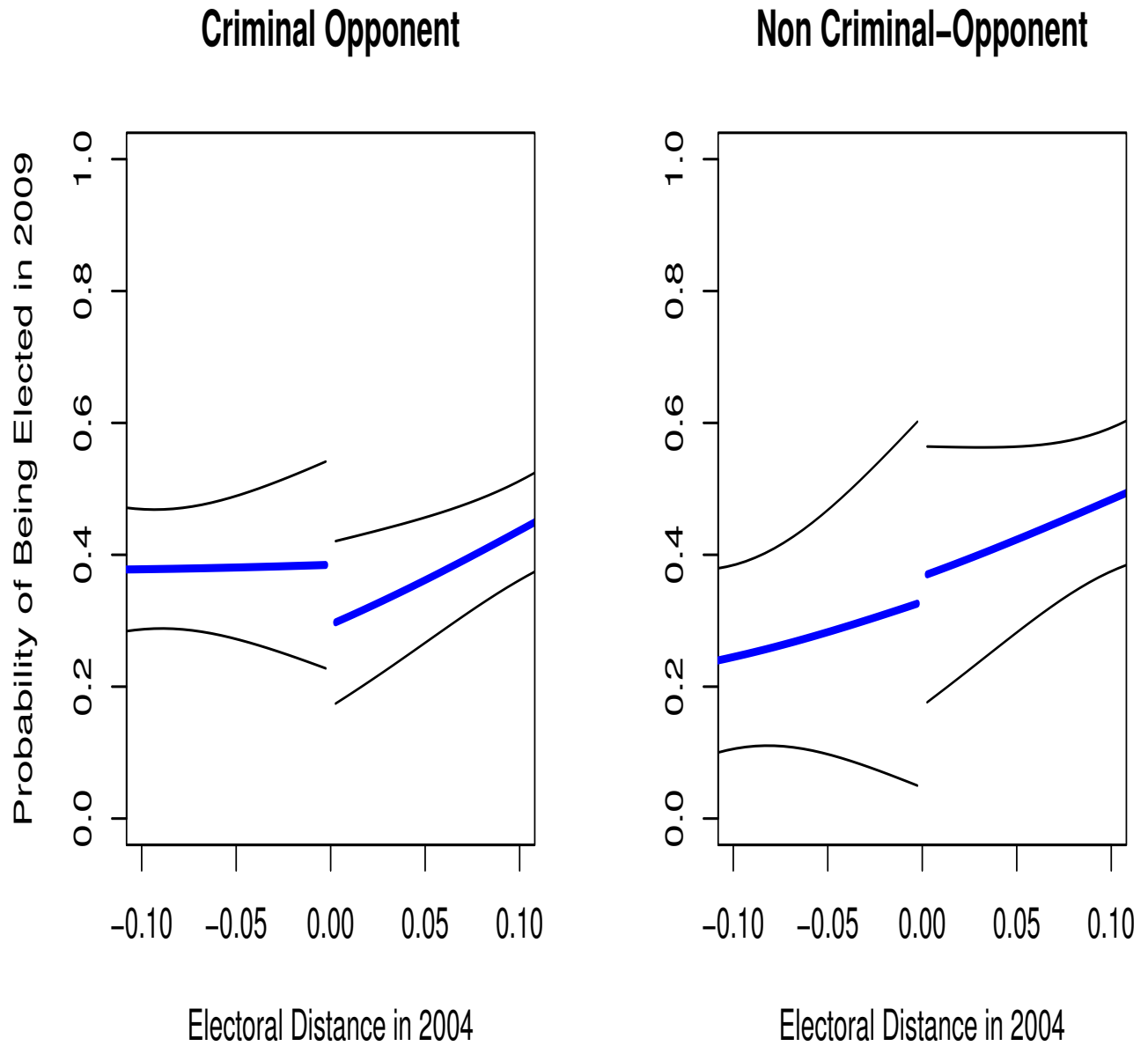
46/22-5/7

Figure 4: Regression Discontinuity Results of the Probability of Being Elected, All Candidates



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

Figure 5: Regression Discontinuity Results of the Probability of Being Elected, Criminal and Non-Criminal Opponent



Notes: Candidates limited to those within 25 percent of winning. Criminal candidates defined as those who report having been convicted of or currently facing criminal charges in their affidavit filed with candidacy papers.