BUILDING CONNECTIONS: POLITICAL CORRUPTION AND ROAD CONSTRUCTION IN INDIA*

Jonathan Lehne† Jacob N. Shapiro‡ Oliver Vanden Eynde§

29 February 2016

Abstract

Corruption is a pervasive challenge for development. We provide empirical evidence that political corruption can impact resource allocation even in programs where politicians have no official role in allocation decisions. Using data from the bidding process for a major rural road construction programme in India – the Pradhan Mantri Gram Sadak Yojana (PMGSY) – we show that contractors benefit when politicians they are connected to – as measured by shared surnames – win office. Our regression discontinuity design exploits close elections to identify the causal effect of a politician coming to power on the composition of contractors winning road construction contracts in their constituency. Relative to the previous term, the share of contractors whose name matches that of the winning politician increases by 63%. Politicians appear to be intervening in the allocation of contracts on behalf of members of their own network, a striking fact given that politicians have no official role in making contracting decisions. Regression discontinuity estimates at the road level show that this political interference raises the cost of road construction, and has no clear offsetting benefits in terms of efficiency or quality.

JEL Codes: D72, D73, L14, O18

---

* We would like to thank the International Growth Centre (IGC) and Cepremap for supporting the data collection of this project. We are also grateful to Nawal Aggarwal, Ashish Modi, Shrenik Sanghvi, Radha Sankar, and Paolo Santini for excellent research assistance. This paper has benefited from discussions with Sam Asher, Dan Keniston, Ariane Lambert-Mogilianski, Karen Macours, Alexander Plekhanov, Akiko Suwa-Eisenmann, Liam Wren-Lewis, and Maiting Zhuang. We also thank seminar participants in PSE's IRG Workshop and the Sussex Development Workshop. All remaining errors are our own.

† Paris School of Economics, jonathan.lehne@gmail.com.

‡ Princeton University, Woodrow Wilson School of Public and International Affairs, jns@princeton.edu.

§ Paris School of Economics, olivervandeneynde@gmail.com.

1. INTRODUCTION

Corruption is a major obstacle to the provision of public goods in many developing countries (World Bank, 2011). When officials are able to siphon off part of the funds allocated to a project, only a fraction of the money reaches the intended beneficiaries. The undue benefits of corruption are often shared within kinship groups (Chandra 2004), which means that corruption and political connections reinforce each other. In spite of growing literatures on both measuring corruption and the value of political connections, there is little micro-evidence on how exactly public officials intervene to favour connected firms and whether these interventions are welfare-reducing. Even less evidence exists about why kinship networks facilitate corruption or whether they can do so in the absence of a formal role for politicians in resource allocation decisions. Our paper addresses these questions using unique data on more than 88,000 contracts awarded in India's major rural roads construction programme.

We find that state-level legislators (MLAs), who do not have any formal role in the allocation of contracts, manage to get more projects assigned to contractors who share their surname (a proxy for caste connections in India). These favoured contractors build roads that are more expensive without observable differences in quality. Hence, we document that democratically elected politicians use their power improperly to benefit connected firms at the cost of the population at large. We also address the question of why it is kinship networks that act as a conduit for corruption in this setting. Election incentives do not appear to strengthen rent-seeking behaviour, suggesting that kinship networks are not targeted for patronage in order to buy votes, but may instead offer the mutual trust required for engaging in illegal corruption.

Any empirical analysis of corruption or political connections must contend with a lack of data. Participants in illegal activities are typically reluctant to provide information on their actions and may go out of their way to conceal them. In this setting the challenge is twofold. Firstly, there is no information on actual connections between politicians and the contractors active in their constituency. Secondly, to the extent that politicians intervene in the allocation of roads on contractors' behalf, such improper interference would not be documented.

We address the first problem by constructing a measure of proximity between state-level legislators (MLAs) and contractors based on their surnames. This approach follows a number of papers that use Indian surnames as identifiers of caste or religion (e.g. Hoff and Pandey 2005, Field et al. 2008, Banerjee et al. 2014b). We measure proximity after a politician is elected and in the term before. This allows us to evaluate how the composition of contractors changes when the politician is voted into office.

Dealing with the second issue – identifying improper intervention – requires isolating the variation in proximity that results from the MLA coming to power. To this end, we employ a regression discontinuity approach that exploits the fact that in close elections, candidates who barely lost are likely to have similar characteristics to those who were barely elected. If MLAs are intervening in the assignment of contracts, one would expect a shift in the allocation towards contractors who share their name, and no equivalent shift for their unsuccessful opponents.

Applying this difference-in-differences approach in conjunction with a non-parametric regression discontinuity (RD) design, we find strong evidence that politicians do indeed interfere in the allocation of PMGSY contracts. The average political candidate shares a surname with 4% of the contractors active in their constituency in the term prior to the election, When an MLA is elected, the estimated increase in the share of contractors of their name is 63% in our preferred specification. Our sample of 4,058 electoral terms from 2001 to 2013 covers 2,632 constituencies for which we have election data and in which PMGSY agreements were signed both before and after the election. This sample is broadly representative of areas receiving PMGSY roads, and covers 24 of 28 states which existed throughout the period. This result can be considered a lower bound on the true level of political interference, as our empirical approach will fail to detect all preferential treatment of contractors who may be connected to politicians but do not share their name. A series of robustness checks corroborate the main result.

The allocation of contracts to those with political connections does not conclusively prove that politicians' motives are corrupt. In an environment of imperfect information, MLAs may be better informed about, and better able to monitor, contractors in their own network, which would imply improvements in road quality within connected contracts. An RD estimation at the road level, provides no evidence that political interventions promote efficiency or quality. On the contrary, roads allocated to connected contractors are both more expensive to construct and more likely to fail subsequent quality inspections, the opposite of what we would expect if politicians were biasing the quality assurance process. This is consistent with politicians putting pressure on officials to reject the lowest bidding or most reliable contractors in favour of someone from their network. Given the lack of any offsetting beneficial effects, these results suggest that political interference is welfare-reducing in this setting.

An open question in the literature is why politicians target patronage along kinship lines. We consider two explanations: (1) MLAs may allocate roads to members of their own group as a form of vote-buying, the standard explanation in the literature; and (2) kinship networks provide trust and social sanctions that facilitate collusion. We find no evidence that the preferential allocation of roads or the cost inflations increase immediately before or after election dates, so if vote-buying is going on it must be a long-run transaction. We also exploit India's 2008 re-drawing of electoral constituency boundaries to study the behaviour of MLAs in regions that have become "politically irrelevant" after the redistricting. We find no evidence of different behaviour in these regions. Our results are more consistent with the trust explanation, particularly in that corruption occurs in ways that are harder to detect ex-post. In particular, we find strong evidence of preferential allocation, ex-ante cost inflation and weak evidence of quality deficiencies, but evidence of increased cost over-runs or delays in construction. Hence, corruption affects precisely those aspects of programme implementation that are least likely to be scrutinised ex-post. Taken together, our findings suggest that the role of kinship networks in our context is not to act as conduits for vote-buying activity, but rather to provide the mutual trust required for risky collusive behaviour.

Corruption in the allocation of PMGSY contracts is associated with significant costs. Our estimates imply that of the contracts in our sample, around 1600 roads, worth roughly 470 million USD, had been allocated to contractors who would not have received them without political connections. But, the interplay between political connections and corruption gives rise to social costs beyond the misappropriation of resources. Where elected office opens up opportunities for corruption, this can have divisive consequences for politics and weaken the quality of governance more generally.[1]

The remainder of the paper proceeds follows. Section 2 reviews the literature on political corruption in public goods provision and discusses our contribution to that literature. Section 3 provides context on PMGSY, the role of MLAs, and Indian surnames as identifiers of caste or religion. Section 4 describes the dataset used in the analysis. Section 5 outlines the empirical strategy. Section 6 presents the main results. Section 7 analyses their robustness. Section 8 discusses extensions to the main analysis. Section 9 concludes.

2. LITERATURE

Our paper relates to a large literature on corruption, political connections, and ethnic favouritism. In the theoretical literature on corruption, a distinction is typically made between the rent-seeking hypothesis, and the "greasing the wheels" hypothesis. Under rent-seeking, public officials use their control over the allocation of contracts or the provision of services to ask for bribes (e.g. Becker and Stigler, 1974; Krueger, 1974; Rose-Ackerman, 1975; Shleifer and Vishny, 1993). This behaviour is most likely to arise in contexts where enforcement is weak and officials are poorly remunerated.[2] In contrast, the "greasing the wheels" hypothesis argues that corruption could be optimal in a second-best world, by allowing agents to circumvent inefficient institutions and regulation (Leff 1964, Huntington 1968, Lui 1985). In principle, both arguments could apply to the preferential assignment of PMGSY roads by Indian MLAs. However, the evidence we present on cost inflation in preferentially allocated contracts support the rent-seeking hypothesis.

Early empirical work on corruption was based on subjective measures of 'perceived corruption',[3] which can be hard to interpret and are subject to cognitive biases (Rose-

[1] Banerjee and Pande (2009) show that when voters prefer candidates of their own caste – in expectation of future preferential treatment – the quality of elected politicians declines.

[2] Becker and Stigler (1974) argue for a form of efficiency wage. In the case of Indian MLAs, this calculation may be complicated by the fact that candidates frequently need to pay their parties significant sums for their place on the ticket, prompting them to engage in corrupt behaviour once elected, so as to get a return on their investment (Jensenius,2013).

[3] Indices compiled by the Economist Intelligence Unit, the World Bank, or Transparency International, have frequently been used in cross-country regressions analysing either the determinants or the effects

Ackerman 1999, Reinikka and Svensson 2002). Recently, a growing number of papers seek to provide objective, quantitative estimates of corruption (Banerjee et al., 2012). Underlining the benefits of such an approach, Olken (2009) finds that villagers' assessments of corruption correlate only weakly with an actual measure of missing expenditures in the context of rural road construction in Indonesia.

In the case of PMGSY, there is no publicly available audit data that would provide a direct measure of corruption.[4] We therefore employ an approach that Banerjee et al. (2012) refer to as "cross-checking": the comparison between (i) an actually observed outcome (which may or may not reflect corrupt activity), and (ii) a counterfactual measure which should be equivalent to the former in the absence of corruption. Our counterfactual is the proximity between contractors and losing candidates in close elections. If politicians are not intervening in the allocation of road projects, they should be no 'closer' to contractors than their successful opponents. In this sense our empirical strategy is close to that of Do et al. (2013), who use a regression discontinuity design to compare the performance of firms connected to winning and losing candidates in close gubernatorial elections in the US. Other exponents of the "cross-checking" approach include Acemoglu et al. (2014), Golden and Picci (2005), Reinnika and Svensson (2004), Olken (2007), Fisman (2001), and Banerjee et al. (2014b). By and large, this literature offers more support for the rent-seeking than for the "greasing the wheels" hypothesis, and our findings point in the same direction.

Given the challenge of collecting data on corruption, most of these existing cross-checking studies have been conducted either in highly localised settings or at the macro-level. The PMGSY data we use are exceptional in that they provide fine-grained local variation and near-to-complete coverage of one of the world's largest countries. Our paper does not just provide evidence of corruption in this highly relevant context,

of corruption (e.g. Mauro 1995; Knack and Keefer, 1995; Treisman 2000). There are also a growing number of sub-national indices, including one for India compiled by Transparency International in 2005. Using this data at the state level, Charron (2010) finds that higher levels of development and greater fiscal decentralisation are negatively associated with corruption perceptions.

[4] Several countries conduct regular audits of local government expenditure and make the results publicly available. Examples of research based on these data include Ferraz and Finnan (2008 and 2011) and Melo et al. (2009) for Brazil, or Larreguy, Marshall and Snyder Jr (2014) for Mexico. Alternatively, studies can be designed to observe corruption independently, as in the Indian driving license experiment by Bertrand et al. (2007). A potential problem with direct measurement is that participants' knowledge that they may be audited is likely to affect their willingness to engage in corrupt behaviour (Olken and Barron, 2007).

the granularity of our data also allows us to document the link between preferential assignment of contracts in kinship networks and the rents implied by the characteristics at the level of contracts. Most existing studies on the value of political connections cannot provide direct evidence on contract characteristics, which constrains the study of underlying mechanisms. Thanks to contract-level data, we can test "efficiency" arguments and show that connected contractors build more expensive roads without achieving any offsetting improvements in quality.

Our work also relates to a pattern described in the literature on "patronage-democracies" (Chandra 2004, Horowitz 1985), where targeting patronage is easier within ethnic groups. Voters' preference for patronage could motivate them to choose politicians of their own caste (Banerjee et al., 2014b), possibly under the influence of local strong men who command the votes of certain groups. An alternative explanation relates to risk of engaging in corruption (Cadot 1987; Lambsdorff, 2002; Tonoyan, 2003). As an illegal activity, corruption requires a degree of mutual trust among collaborators, which is most likely to exist between members of the same family, ethnic group, or network. Exploiting the timing of elections and a major redistricting exercise, we confirm that "electorally relevant" time periods or regions do not experience higher corruption. The latter finding contrasts with one of the few existing papers using contractor-level data. [5] Mironov and Zhuravskaya (2013) link elected politicians to the firms awarded public procurement contracts, and use data on financial transactions in Russia to document an electoral cycle in tunnelling and the allocation of public procurement contracts. Firms who tunnel money in the run-up to elections are significantly more likely to receive procurement contracts after the election. While closely related, the focus of our paper is different, as we document preferential assignment of contracts in kinship networks as opposed to confirmed campaign contributors.

The results of our paper are of particular relevance for India and the functioning of its democratic institutions. A recent literature confirms the large influence of state legislators (MLAs) on economic outcomes. Asher and Novosad (2015a) find higher employment in constituencies whose MLAs are aligned with the state-level

[5] Amore and Bennedsen (2013) exploit an exogenous increase in local Danish politicians' power to show that companies with close family ties to those politicians see an increase in their profits.

government. Higher clearances of mining projects suggest that MLAs use their influence in the administration to push employment generating projects. An important source of influence for these politicians is their ability to reassign bureaucrats, as highlighted by Iyer and Mani (2012). Prakash et al. (2015) also confirm the economic importance of MLAs. These authors find that the election of criminal MLAs leads to lower economic growth in their constituencies. Finally, Fisman et al. (2015) show that the assets of marginally elected MLAs grow more than those of runner-ups, which confirms the idea that there are substantial private returns to holding public office in India. Compared to this recent literature, our paper sheds more light on the long causal chain that connects the characteristics of MLAs with aggregate economic outcomes. By showing how MLAs use their power improperly to favour connected contractors, our paper provides micro-evidence on the channels of influence of these democratically elected politicians. Moreover, our paper suggests that not just the economic performance of the constituency is at stake when MLAs exert influence, through the preferential allocation of contracts they also affect the distribution of public goods between competing patronage networks.

The importance of kinship networks also links our paper to the literature on ethnic favouritism in public good provision. Focusing on road construction as well, Burgess et al. (2014) show that the ethnic homelands of Kenyan presidents receive preferential coverage by road projects, but only under autocracy.[6] Our paper shows how democratically elected politicians use their power to favour caste or family networks through the preferential assignment of unduly lucrative contracts. It suggests that "ethnic" favouritism affects various stages of the public good provision process and is not always mitigated by democratic institutions.

3. BACKGROUND

3.1 PMGSY

In the year 2000, an estimated 330,000 Indian villages or habitations – out of a total of 825,000 – were not connected to a road that provided all-weather access (PMGSY

[6] Kramon and Posner (2016) show similar favouritism in schooling outcomes, although these survive in periods of democracy.

2004). As such, their inhabitants were at least partially cut-off from economic opportunities and public services (such as health care and education). In an effort to address this lack of connectivity, the Indian government launched the Pradhan Mantri Gram Sadak Yojana (PMGSY) in December 2000. Its goal was to ensure all-weather access to all habitations with populations over 1,000 by the year 2003, and to those with more than 500 inhabitants by 2007. In hill states, desert and tribal areas, as well as districts with Naxalite insurgent activity, habitations with a population over 250 were targeted (PMGSY 2004).

The programme has been described as “unprecedented in its scale and scope” (Aggarwal 2015), with roadwork for over 116,000 habitations completed and another 23,000 currently under construction as of January 2016 (OMMS 2015). A second phase of the scheme (PMGSY II), launched in 2013, targets all habitations with populations over 100. According to World Bank estimates, expenditures under PMGSY had reached 14.6 billion USD by the end of 2010, with a further 40 billion USD required for its completion by 2020 (World Bank 2014).

Several studies have focused on the first-order research question that arises in relation to PMGSY: its impact on habitations and the lives of their inhabitants. Asher and Novosad (2015b) analyse the employment effects of the programme in previously unconnected villages. They find that a new paved road raises participation in the wage labour market with a commensurate decrease in the share of workers employed in agriculture. Aggarwal (2015) also finds a positive effect on employment and reduced price dispersion among villages.

While the studies above analyse *what* PMGSY has achieved, this paper looks at *how* it has been implemented. Given the financial expenditures involved, the potential gains for habitations in being allocated roads as quickly as possible, and the potential gains for contractors in being selected to build them, there are significant incentives to bend the programme rules. A number of newspaper reports document alleged corruption in PMGSY.[7] The possible manipulation of road allocations is also one of the principal

[7] Examples include articles in “The Hindu” on April 11 2012, “The Economic Times” on March 8 2013, “The Arunachal Times” on March 6 2013, the online news-platform “oneindia” on July 31 2006, and “Zee News” on 30 August 2014. For example, the “oneindia” article reports that the former Chief Minister of Sikkim accused the current administration of “widescale corruption” in the implementation of PMGSY and “alleged that the works were awarded to relatives of Chief Minister, Ministers and MLAs of the state”.

challenges for impact evaluations of the programme seeking to identify exogenous variation in treatment (Asher and Novosad,2015b).[8] Our paper tests for a specific form of corruption in PMGSY: interventions by state-level parliamentarians (MLAs) in the allocation of roads within their constituencies.

An advantage of focussing on MLAs in this context is that under the programme guidelines, they should be in no way involved in the tendering process or the selection of contractors. In fact, they are granted practically no official role in the implementation of PMGSY whatsoever.[9] Funding for PMGSY comes primarily from the central government. The scheme is managed at the district level by Programme Implementation Units (PIUs) which are under the control of State Rural Roads Development Agencies (SRRDA). These agencies are responsible for inviting tenders and awarding contracts. Given their lack of formal involvement, any systematic relationship between MLAs and the contractors working in their constituencies can therefore, in itself, be construed as evidence for an irregularity in the allocation of contracts.

While there are strict rules for the assignment of PMGSY roads, the process is open to manipulation. Bidding takes the form of a two stage auction. In the first stage, officials are responsible for evaluating contractors' eligibility and "bid capacity" (PMGSY 2001). Only bids deemed to meet the technical requirements make it to the price auction. This structure affords officials considerable discretion and implies that the contract need not be awarded to the lowest bidder. In a case study of 190 contracts in Uttar Pradesh, for example, Lewis-Faupel et al. (2014) found that among the contracts where multiple bids were submitted, all but one bid was disqualified on technical requirements in over 75% of cases. This suggests that it would be possible for an MLA to influence the allocation of contracts in their constituency, provided that they are able to put pressure on the officials in the relevant PIU or SRRDA.

[8] These authors find that the habitation population figures reported to PMGSY had been manipulated, particularly around the 1,000 and 500 population cut-offs used to target the program..

[9] MLAs are mentioned in the PMGSY guidelines, but only in reference to the initial planning stage. Intermediate panchayats and District panchayats were responsible for drawing up a planned "Core Network" which encompasses all future roadwork to be carried out under PMGSY. These plans were to be circulated to MPs and MLAs, whose suggestions were to be incorporated. MLAs could therefore have influenced which habitations were targeted ex-ante through official channels. However, given that the analysis below focuses on the allocation of roads after the commencement of the programme, these interventions should not be relevant.

3.2 The role of MLAs

Is it plausible that MLAs would seek to intervene on behalf of specific contractors? While their official function is to represent their constituents in state legislative assemblies, surveyed MLAs overwhelmingly report this to be a minor part of their work (Chopra 1996). State assemblies meet rarely and according to Jensenius (2013), individual legislators have little impact on political decisions: "much more important to the MLAs are all their unofficial tasks of delivering pork, blessing occasions, and helping people out with their individual problems". Qualitative accounts suggest that MLAs spend much of their time receiving requests from their constituents, including those seeking to overcome or circumvent bureaucratic obstacles. Describing such meetings Chopra (1996) writes "constituents came to ask for favours that clearly contravened rules and laws". MLAs often respond to requests by passing them on to ministers or high-ranking officials, but are also known to put pressure on bureaucrats by threatening them with reassignment (Iyer and Mani 2012, Bussell 2015).

Asher and Novosad (2015a) provide quantitative evidence of state politicians' control over local bureaucrats in India. Using an empirical strategy similar to that of this paper – a regression discontinuity design which exploits close assembly elections – they show that firms perform better when the MLA for their constituency is aligned with the state's governing coalition. They find that the effect is strongest in industries heavily dependent on government inputs controlled by local bureaucrats. While their empirical approach exploits variation in MLAs' influence, this paper focuses on variation in the proximity between MLAs and the potential beneficiaries of their influence.

3.3 Surnames as a measure of interpersonal proximity in India

Absent direct proof of politicians meddling in contract allocation, a systematic relationship between the identity of those in office and the identity of those receiving contracts constitutes the best available indirect evidence. To measure proximity between MLAs and contractors we construct a proxy based on politicians' and contractors' surnames.[10]

Indian surnames can (but need not) be an indicator of caste affiliation, religion, or geographic provenance. The strength of these associations varies regionally and across names within regions. Overall, the correlations are sufficiently strong for Indian surnames to have been used as identifiers of caste or religion in many empirical studies (Banerjee et al. 2014, Hoff and Pandey 2005, Vissa 2011, Fisman et al. 2012, Field et al. 2008). This paper treats a match between the names of a politician and a contractor as a rough overall measure of proximity, without seeking to establish whether the individuals are of the same religion, caste, or (potentially) family. All of these types of connections are likely to increase the probability that a contractor would approach an MLA when bidding for a contract, and that the MLA would be receptive.

Name-based matching is an imperfect measure of proximity. Contractors may have connections to politicians without sharing a name, or equally, share a name but have no connection. Surnames that are not caste-identifiers, former honorific titles for example, are likely to dilute the accuracy of the measure. Hence, the estimates in this paper can be viewed as a lower bound for MLAs' true effect on contract allocation.

4. DATA

The empirical strategy requires three kinds of data. Information on contractors and agreements is available in the administrative records of the PMGSY project, at the road level. Data on political candidates and elections are at the level of the assembly constituency. These two are linked using the population census, which allows for habitations to be matched to constituencies, as well as providing additional covariates used in the analysis.

4.1 PMGSY data

[10] Angelucci et al. (2010), and Mastrobuoni and Patacchini (2012) also uses name-based matching to study social networks

The administrative records of projects sanctioned under PMGSY are publicly available in the Online Management and Monitoring System (OMMS). The dataset used for this paper contains the agreement details of 110,185 roads serving 188,394 habitations.[11] This information includes: date of contract signing, sanctioned cost, proposed length, proposed date of completion, name of the contracting company, and – crucially for this analysis – name of the winning contractor. In addition to the agreement details, which precede road construction, the OMMS also contains later data on the physical progress of work, data on completed roads, and reports from subsequent quality inspections. These are used in section 8 to evaluate the effect of political interference on the efficiency and quality of road construction.

4.2. Assembly election data

The Election Commission of India (ECI) publishes statistical reports on assembly elections that record each candidate's name, party, gender and vote share. Since 2003, candidates have moreover been required to submit sworn affidavits to the ECI with information on their assets, liabilities, educational attainment, and any pending criminal cases. Both the election reports and affidavits are publicly available from the ECI in pdf format. This paper draws on digitised versions of this information from four separate sources. Table A3 of the appendix lists these sources – which cover different time periods and variables – and describes which variables from each source are used in the analysis. Given that the underlying primary sources are always the ECI reports, the creation of a unified dataset from four separate secondary sources should not introduce inconsistencies in measurement or definitions. The matching process is, however, complicated by discrepancies in the spelling of constituency and candidate names.[12] Where fuzzy matching of names was required, we complemented this by matching on other variables (vote shares, party, and age) in order to minimise the potential measurement error that would results from incorrect matches.[13]

[11] A single road can connect multiple habitations while multiple roads may also pass through the same habitation.

[12] Inconsistent spelling of constituencies and candidate name occurs not only across datasets but also across time within datasets.

[13] In a small number of cases, multiple constituencies within the same state have the same name. We drop all of these constituencies from our sample, to prevent false matches between election datasets and to avoid the risk of assigning roads to the wrong constituency.

Assembly elections operate on a plurality rule and the median number of candidates per election is eight. To estimate the RD we restrict attention to elections in which there are multiple PMGSY contracts issued in the term before and after the election. This gives us a sample of 8,116 candidates in 4,058 elections from 2001 to 2013 covering 2,632 constituencies. In our preferred specification we estimate on the resulting sample of 8,116 constituency-terms.

4.3. Matching roads and electoral terms using census data

The Population Census of India 2001 contains village-level data on demographic and socio-economic variables which are used as controls in the analysis. The census is also the source for habitation-level data, which is collected by the PMGSY in order to determine the prioritisation of roads. This includes information on the size of the population (the project guidelines stipulate that habitations above certain population thresholds are to be prioritised), whether or not it was connected to a road in 2001, and if so, whether this road provided all-weather access. Moreover, it reports the MLA constituency in which it each habitation was situated in 2001.

Using this information, it is possible to match PMGSY roads (at the habitation-level) to the assembly election data described in the previous sub-section. However, changes in the delimitation of MLA constituencies – which took effect in mid-2008 – led to changes in boundaries, the abolition of some constituencies, and the creation of new ones. For roads built in electoral terms after the new delimitation we use the coordinates of habitations and match these to GIS data on constituency boundaries.

While the census data allows for spatial matching of roads and constituencies, it is also necessary to match them in time. Road contracts are allocated to electoral terms based on the date of the agreement, as recorded in the PMGSY data. In order to precisely assign road contracts, it is necessary to set an exact date that marks the end of one term and the beginning of the next. We define this as the date on which the results of an election are announced.[14]

[14] These dates were collected from the website www.electionsinindia.com.

Our estimation sample includes the 4,058 electoral terms for which we have election data and in which PMGSY agreements were signed in the respective constituency both before and after the election. Map A1 shows the constituencies included in the sample which cover 24 of the 28 states that existed during the timeframe under analysis.[15] Our sample is drawn only from candidates who finished either first or second in these elections.

4.4 Matching politicians and contractors using surnames

In the electoral terms that preceded and followed the elections in the sample, 88,020 road agreements were signed. For each political candidate, we assess whether they share a surname with the contractors who received projects in their constituency in the term after the election. For every politician-contractor pair, we exclude all names except for each individual's final name and then look for matches among these surnames. The results are, however, robust to broader definitions of matches.[16] To account for different spellings of the same name (eg. Aggarwal and Agrawal) we implement a fuzzy matching algorithm optimised for Hindi names.[17]

Matches are aggregated at the electoral term level as follows. The variable $match_{nijt}$ takes the value of 1 if the contractor for a road agreement *n*, signed in constituency *j* in term *t*, shares a name with candidate *i*, and 0 otherwise. This variable is determined for the N road agreements signed in the constituency during an electoral term. $share_{ijt}$ is defined as the share of contracts in term *t* allocated to contractors who share a candidate's name. $share_{ijt-1}$ provides the equivalent share for contracts in the term prior to the election in which a candidate took part.

$$share_{ijt} = \frac{\sum_{n=1}^{n} match_{nijt}}{N_{jt}} \qquad share_{ijt-1} = \frac{\sum_{n=1}^{n} match_{nijt-1}}{N_{jt-1}}$$

The dependent variable in the main regressions is the difference between these two:

$$\Delta share_{ijt} = share_{ijt} - share_{ijt-1}$$

[15] We have no data for Assam and Goa.
[16] Naming conventions differ across India; it is common for Indians to have multiple surnames and the same name can appear in different positions within the list of names. This is also true of caste identifiers. The results are robust to considering all matches among individuals' names (excluding their first name) or only matches based only on the last two names.
[17] All results are robust to considering exact matches only.

A complication arises in elections where winning and losing candidates share at least one name. In these cases, $\Delta share_{ijt}$ is perfectly correlated or highly correlated. It is not possible to estimate the effect of winning an election in this situation, as candidates who lost will see their proximity to contractors evolve in parallel to that of the elected politicians. In the main regressions, we therefore exclude *all candidates* from elections where this issue arises from the main regressions.

4.5 Descriptive statistics

Table 1 reports descriptive statistics for the sample of candidates used in the main regressions. For the average term in the sample, the number of road contracts signed is 28. The average value of $share_{ijt-1}$ – which can be construed as a baseline measure of the frequency of surname-matches – is 4%. There is however, significant geographic variation in the frequency of matches, ranging from a mean of 0% in Mizoram to a mean of 13% in Andhra Pradesh (Map A2 shows this variation at the constituency-level).[18] However, these means do not distinguish between winning and losing candidates – the variation exploited in the empirical strategy below.

[TABLE 1 here]

5. EMPIRICAL STRATEGY

A natural control group for elected politicians are those who aspire to the same office. If being an MLA is associated with the power to intervene in the allocation of roads in one's constituency, one would expect the share of contractors with the same name as a winning candidate to be higher than the corresponding share for losing candidates.

However, an OLS regression of $share_{ijt}$ on a dummy ($winner_{ijt}$), would not be sufficient to identify an effect of coming to power on the allocation of contracts. For example, it would be biased if individuals from the majority caste in a constituency were more likely to be elected, and simultaneously more likely to share a name with anyone in the constituency (including road contractors). In order to identify whether

[18] It is likely that these baseline frequencies lead to heterogeneous treatment effects. In states or constituencies, where the distribution of names is such that matches are relatively rare, a politician who is elected may not have many potential contractors of the same name to allocate roads to.

there is a causal relationship between the election of politicians and the allocation of road contracts in their constituencies, we employ a difference-in-differences approach (DiD) combined with a regression discontinuity design (RD) exploiting the fact that in close elections, the assignment of victory can be considered conditionally independent of subsequent contracting patterns.

Taking the first difference of $share_{ijt}$ should remove unobservable, time-invariant characteristics of an individual candidate that may be correlated with the number of matches with contractors. In our context, this is a way of controlling for specificities that individual names may have within certain constituencies. Some candidates' names will be more common than others. Some may be more prevalent among certain professions (e.g. contractors) for historical reasons. Under the assumption that winning and losing candidates had a common trend in their share of matches with contractors, a simple DiD approach would be sufficient for identification. However, given that winners are likely to be systematically different from losing candidates in many respects, it is possible that they may face divergent trends in $share_{ijt}$ that are not determined by election outcomes. This suggests the use of an RD estimator.

Lee (2001) was the first to apply an RD design to electoral outcomes. In the context of Indian state-level legislative elections, this approach has recently been employed by Asher and Novosad (2015a). The underlying assumption is that candidates who won an election by a very small margin are comparable to those who narrowly lost.[19] We evaluate whether this assumption holds in our sample by running balance checks on observable characteristics (see below). In order to determine how close elections were, we define the variable $margin_{ijt}$:

$$margin_{ijt} = \begin{cases} vote\ share(winner)_{jt} - vote\ share(runnerup)_{jt} & if\ winner_{ijt} = 1 \\ vote\ share(runnerup)_{jt} - vote\ share(winner)_{jt} & if\ winner_{ijt} = 0 \end{cases}$$

We estimate equation (1) in a non-parametric RD for a range of bandwidths μ, controlling for the assignment variable $margin_{ijt}$ and its interaction with $winner_{ijt}$ to

[19] Lee (2008) formalizes the conditions under which vote shares in close elections provide quasi-random treatment assignment. The key assumptions are: (i) that there is a stochastic component to candidates' vote shares and (ii) that the probability density function of each individual's vote share is continuously distributed.

allow for a different relationship between $\Delta share_{ijt}$ and $margin_{ijt}$ among winning and losing candidates:

$$\Delta share_{ijt} = \alpha + \beta winner_{ijt} + \delta margin_{ijt} + \rho winner * margin_{ijt} + \varepsilon_{ijt}$$
$$\forall\ i\ where\ margin_{ijt} \in [-\mu, \mu] \qquad (1)$$

In order to improve the efficiency of the estimates, we introduce constituency-level controls, individual-level controls, state fixed-effects,[20] and year fixed-effects in most specifications although these are not required for identification. Because we have the top-two candidates in each election we cluster standard errors at the election level.

A feature of non-parametric RD designs is the trade-off between bias and efficiency inherent in the choice of bandwidth (Lee and Lemieux 2010). Restricting the sample to a very narrow margin around the cut-off reduces the potential for bias, at the expense of precision. While we primarily report results for bandwidths of 6.2% (derived from the optimal bandwidth choice rule of Imbens and Kalyanaraman 2012), 5% and 2.5%, the results are consistent across a wide range of bandwidths (figure 1 below). As a robustness check we also estimate a parametric RD on the whole sample:

$$\Delta share_{ijt} = \alpha + \beta winner_{ijt} + f(margin_{ijt}) + g(margin_{ijt}) * winner_{ijt}$$
$$+ \gamma X_{ijt} + \varepsilon_{ijt} \qquad (2)$$

6. Results

6.1 Regression discontinuity results

Our identification strategy is based on the premise that restricting the sample to close elections ensures that the treatment and control groups are comparable. Table 2 presents the results of a randomization test for the optimal bandwidth of 6.2%. None of the MLA characteristics display a discontinuity when the vote margin exceeds one.

[TABLES 2, 3 here]

[20] Legislative assembly terms are not synchronised across Indian states. In each year in our sample window, there were elections in multiple states.

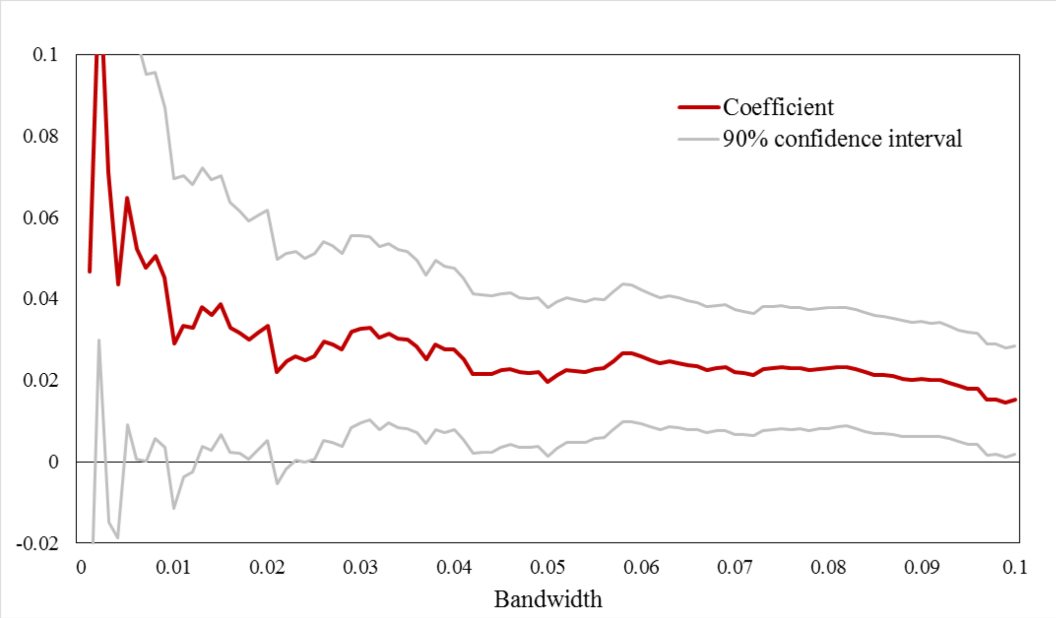
The results of local linear regression RD estimation are presented in Table 3. For each bandwidth there are two columns. The first corresponds to the basic RD in equation (1). The second adds state fixed effects, year fixed effects and additional controls. These include whether or not a constituency is reserved for candidates from scheduled castes (SC) or scheduled tribes (ST), characteristics of the PMGSY roads built in the constituency prior to the election, and candidate-level controls. The latter set of variables includes a candidate's vote share, their age, gender, and whether they were an incumbent. Standard errors are clustered at the election-level to account for the likely correlation in the error term between candidates.

For the 6.2% bandwidth, the effect of winning an election on the change in $share_{ijt}$ is consistently positive and significant. The coefficient is around 0.024 in our preferred specification including fixed effects and the full set of controls (column 4). Relative to the baseline, pre-election level of matches, the latter estimate implies that the effect of a candidate coming to power is a 63% increase in the share of roads allocated to contractors who share their surname.[21]

The results of non-parametric RD estimations can be sensitive to the choice of bandwidth. Figure 1 plots the coefficient on $winner_{ijt}$ for the main specification including fixed effects as well as constituency- and candidate-level controls. As the samples get smaller the estimates are less precise but the coefficient is relatively stable for all but very small bandwidths (less than 1%).

Figure 1: Main effect by bandwidth

[21] In appendix table A2 we report results using the level of the share of same name contractors (rather than the difference) as the main outcome. In these results, the key coefficient is slightly smaller and less consistent in magnitude across bandwidths, but it remains significant in all but one specifications.



Note: The coefficient plotted is for winner in Table 3 (columns 2, 4, 6, and 8).

Relative to the total number of roads – most of which are allocated to contractors whose name does not match the MLA's – the absolute value of the coefficient implies a small effect. Yet as explained in section 3.3, these estimates can be considered a lower bound on MLAs' true intervention in PMGSY contract allocation. If the results are interpreted as evidence of improper political involvement in the assignment of roads, it raises the question whether this improper involvement *only* occurs on behalf of individuals with the same surname. In this sense the sign and significance of the coefficient might be seen as more important than the magnitude. Secondly, given the scale of PMGSY, even a relatively small fraction can translate into what can be considered a sizeable number of affected roads and substantial financial expenditure. This is illustrated by the following, back-of-the-envelope calculation. In our dataset (including the first electoral term), 4,127 road projects were allocated to contractors sharing a name with the MLA. The total sanctioned cost of these projects was 56 billion INR, or around 1.2 billion USD.[22] Applying our preferred RD estimate (6.2% bandwidth) to the full sample, would imply that MLAs had intervened in the allocation of roughly 1,600 road

[22] Applying the average exchange rate over the period (December 2000 to May 2012): 1 INR=0.0219 USD.

contracts worth around 470 million USD.[23] Of course these estimates rely on an extrapolation from a LATE. Still, they serve to illustrate the economic significance of even proportionately small misallocations in PMGSY contracts.

The results of this section lend support to qualitative accounts on favouritism in the allocation of PMGSY contracts. Only recently, BJP leader Munna Singh Chauhan accused the Uttarkhand State Government of such misallocations:[24]

> "There is a huge scam in tender allotment in Pradhan Mantri Gram Sadak Yojana (PMGSY) in Bahuguna government. Of a total of 113 mega road construction projects, 75 contracts were awarded to chosen ones close to the echelons of power on a single bid basis. […] Coincidentally, one of the contractors awarded the project is also the brother-in-law of state rural development minister Pritam Singh," (Quoted in Zee News, 30 August 2013).

Our analysis suggests that episodes of suspected favouritism in particular states, like the one quoted above, match a wider pattern of corruption that shows up in our sample covering the whole of India. However, we do confirm earlier findings by Fisman et al. (2015) and Prakash et al. (2015) which respectively suggest that the returns to private office and the social costs of criminal politicians are larger in a sub-set of states that are known to suffer from poor institutional quality, the so-called BIMAROU states.[25]

7. VALIDITY OF THE RD APPROACH

The RD design requires that no variables other than the dependent variable exhibit discontinuities at the cut-off. The randomization test in Table 2 provided the first evidence that observable characteristics are comparable on either side of the cut-off.

[23] The estimated impact in the RD with a full set of controls on a 6.2% bandwidth is a 63% increase. This implies that 38.6% of roads allocated to contractors with the same name as the politicians would otherwise have gone to another contractor.
[24] See footnote 6 for references to similar newspaper articles.
[25] Results reported in tables A.3 and A.4. While the definition of BIMAROU is loose, the broadest set includes Bihar, Madhya Pradesh, Rajasthan, Orissa, and Uttar Pradesh, as well as the new states that have been carved out from these historical states: Chhattisgarh, Jharkhand, and Uttarkhand.

Close elections can only be considered to provide quasi-random treatment assignment when the probability density function of candidates' vote shares is continuous (Lee 2008). This will not be the case if candidates are able to strategically manipulate their vote share.[26] The standard test for strategic manipulation of the running variable in a RD design was formulated by McCrary (2008). Applying the McCrary test to the assignment variable in this analysis ($margin_{ijt}$), would not make sense because the density is continuous by construction. For every winner with a positive $margin_{ijt}$, there is a runner-up with the equivalent negative value of $margin_{ijt}$. We therefore test for manipulation in the vote share based on an alternative variable: the margin of victory/defeat for the candidate in the constituency with the higher value of $share_{ijt-1}$. The McCrary test does not reject the continuity of this variable at the threshold. Figure A2 in the appendix presents a graphical depiction of the test.

[TABLE 4 here]

Table 4 presents the results of the parametric RD, estimated on the full sample with quadratic and cubic polynomials of the winning and losing candidates' vote shares. In practice, these terms are almost all insignificant, suggesting that vote shares do not systematically affect candidates' proximity to contractors except at the cut-off that determines victory. Reassuringly, the results of the polynomial estimates are quite similar to those of local linear regression estimation above. The effect of winning an election on the share of roads awarded to contractors of the same name, is always positive and significant.

8. SOCIAL COSTS OF MISALLOCATION

Theoretical work has contended that corruption could be socially beneficial (Leff 1964). In the case of political connections, proximity may be associated with better information ex-ante or greater sanctioning power ex-post, and is therefore desirable in

[26] Using data on close US house races, Caughey and Sekhon (2011) provide evidence of such strategic sorting. Eggers et al. (2015) examine over 40,000 close elections from a range of countries (including India) and find no other country that exhibits sorting.

contexts of adverse selection or moral hazard. Distinguishing between outright corruption and this efficiency motive is a challenge that is faced by many empirical studies on political connections. We analyse PMGSY projects at the road level, in order to evaluate how MLA's interventions affect the cost, timeliness, and quality of road construction.[27]

Are roads built by contractors who are connected to politicians better or worse than other roads? We again employ an RD-approach that exploits close elections to identify the impact of political interference on the efficiency and quality of road construction.[28] We drop all roads from the sample that were not built either by a contractor who shares a name with the current MLA, or by a contractor who shares a name with the runner-up in the most recent election. Once the sample is restricted to close elections, the latter set of roads can be considered a more appropriate control group as it will be similar to the 'treated' roads.[29] Once again we control for the vote shares of winning and losing candidates. The equation for this non-parametric RD is given by:

$$Road\ Characteristic_{nsy} = \alpha + \beta * MLAsamename_{nsy} + \delta margin_{ijt} + \rho winner * margin_{ijt} + \gamma X_{nsy} + \theta_s + \vartheta_y + \varepsilon_{nsy},\ margin_{ijt} \in [-\mu, \mu] \qquad (3)$$

The first outcome we consider is the cost of road construction. If rent-seeking politicians are putting pressure on bureaucrats to reject the lowest bidder in favour of their preferred contractor, we would expect to see a rise in costs. Table 5 shows that roads built by contractors who share a name with an elected official are more expensive (per kilometre). This result is significant for the bandwidths we consider with the coefficient rising as the bandwidth declines.

[27] The efficient selection theory is perhaps most applicable when contractors are connected directly to the official or local politician making the decision. This paper focusses on connections to politicians with no official role in the choice of contractors, which raises the question of whether MLAs' concern for the efficient use of (largely federal) government funds would be sufficient to prompt intervention in bureaucratic allocations. One potential source of motivation would be the desire to ensure that constituents receive roads quickly and that the standard of construction is high.

[28] One way to approach this question empirically would be to run regressions of road characteristics on a dummy variable that takes the value of one if the MLA and the contractor for road have the same name. However, this approach would fail to control for two important sources of unobserved variation. Firstly, contractors who have the same name as politicians may have systematically different characteristics from other contractors. Secondly, the locations where contractors of the same name as the MLA operate could be systematically different from other areas targeted by PMGSY.

[29] Assuming as above, that the names of politicians who just win elections are not systematically different from the names of candidates who just lost.

[TABLE 5 here]

It should be noted that, while the RD-design is likely to be an improvement on a naïve OLS approach, it may still be insufficient to identify a causal effect. To the extent that politicians only intervene on behalf of their network for some roads, and this selective intervention is not random, the ex-ante characteristics of the roads in the treatment group may differ from those in the control group. For example, politicians might try to ensure that more difficult projects are allocated to contractors from their network whom they trust. Given that the road-level outcomes we observe were predominantly determined ex-post – at the time of the contract or during construction – this possibility cannot easily be evaluated. We control for observable variables that might affect the cost of a road project and a politician's desire to intervene in its allocation: characteristics of the terrain (altitude and ruggedness) and whether the project involved the construction of a bridge). The resulting estimates allow us to measure the bias in observable characteristics. Table A5 of the appendix, shows that the coefficient remains unchanged when these additional controls are added.[30] As a result, in order to fully explain the estimates of Table 5, the bias in unobservable characteristics would have to very large relative to the bias in observable characteristics (Altonji et al., 2005).

A rise in costs might not be socially detrimental if it were offset by improved quality. Table 6 therefore analyzes three additional measures of quality using the same RD approach: (i) the number of days between the completion date specified in the contract and the actual date of completion; (ii) the ratio between the actual cost of the project and the cost sanctioned in the agreement; and (iii) a dummy variable for whether a road was deemed "unsatisfactory" or "in need of improvement" in either the latest state quality inspection or the latest national quality inspection.[31] For delays and cost discrepancies we find no significant difference between roads constructed by

[30] In order to ensure comparability, we restrict the sample to roads for which we have information on altitude, ruggedness, and bridges in all the regressions of Table A2.

[31] The quality data available on the OMMS has some shortcomings for the purpose of this analysis. Data is available on national and state quality inspections, and a single road may have multiple inspections in each category. However, only the grade assigned in the latest inspection is provided (for each category). The data therefore do not allow us to distinguish between roads that were satisfactory at the outset, and roads that initially did not pass inspection but were improved prior to subsequent inspections. Moreover, only a fraction of the roads in our sample appear in the quality data, and many of these only had one of the two inspection types (national or state). Pooling the two inspections is not ideal, but it provides the best available measure of initial road quality.

contractors whose name matches the MLA's and those whose name matches the runner-up. However, Table 6 suggests that roads allocated to connected contractors were more likely to fail subsequent quality inspections. This result is significant for the 5% bandwidth, but is weaker and less precisely estimated in smaller bandwidths. To the extent that inferences can be drawn from this incomplete set of indicators, preferential allocation appears to reflect costly corruption with no mitigating improvements in the efficiency of road construction. Indeed, we find suggestive evidence that political intervention leads to roads that are not only more expensive but also of poorer quality.

[TABLE 6 here]

9. UNDERSTANDING THE ROLE OF KINSHIP NETWORKS

We implicitly assumes that kinship ties to politicians are relevant connections in the structure of local political corruption in India, and our results appear to validate this assumption. But why should patronage be targeted along caste or familial lines? The literature offers two main explanations: vote-buying and particularised trust. In this section we attempt to shed light on which is more applicable to corruption in PMGSY

[TABLE 7 here]

If road contracts are awarded in exchange for political contributions or political support, one would expect the bias towards connected contractors to increase in election periods. To test for this we construct more disaggregated measures of proximity: the share of contractors with a candidate's name in the first 12 months after an election ($start\ of\ term_{ijt}$.), the equivalent share for the last 12 months before the subsequent election ($end\ of\ term_{ijt}$), and finally the share for the intermediate, mid-term, period. Table 7 shows the results of applying our main estimation approach to this disaggregated sample and interacting dummies for $start\ of\ term_{ijt}$ and $end\ of\ term_{ijt}$ with $winner_{ijt}$. The overall effect of winning the election is comparable to the term-level results, and we find no differential effects in election years.

[TABLE 8 here]

Although the bias towards connected contractors does not increase in election periods, there could be different patterns for the within-term variation on the cost margin. Politicians might need to extract rents, buy support, or reward supporters were higher before or after elections. Including the interactions between $MLAsamename_{njt}$ and $start\ of\ term_{ijt}$ and $end\ of\ term_{ijt}$ in the road-level regressions, we again find no evidence of a political cycle in which election periods see increased corruption (Table 8). The observed negative effect for the end of term, which is significant in some specifications, is more consistent with increased scrutiny in the run-up to elections acting as a deterrent to corruption.

[TABLE 9 here]

$winner_{ijt} * politically\ irrelevant_{ijt} * post\ announcement_{ijt}$ Changes to the delimitation of parliamentary constituencies allow for an additional test of the vote-buying hypothesis. The changes proposed by the delimitation commission of 2002 were approved in February 2008. Subsequent assembly elections, starting with Karnataka in May 2008, were carried out under the new delimitation. After the reform had been announced and approved, the majority of MLAs elected under the old delimitation continued to hold office for several years until the next election. In constituencies where the boundaries were redrawn, this meant that only some areas would remain part of the constituency at the next election, while others would be of no consequence to the MLA's chances of re-election. We identify such areas with a dummy variable $politically\ irrelevant_{ijt}$ and also disaggregate temporally, splitting the applicable electoral term into the period before the announcement, and the period between February 2008 and the next election (the variable $post\ announcement_{ijt}$ denotes the latter). Given that the boundaries were defined by an independent commission following objective pre-set guidelines, the reform could provide plausibly exogenous variation in the incentive for vote-buying.[32] Table 9 presents the results of our main

[32] According to the Electoral Commission of India's Guidelines and Methodology for Delimitation, "the delimitation of the constituencies in a district shall be done starting from North to North-West and then proceeding in a zig-zag manner to end at the Southern side." Constituencies were to have equal populations, as far as possible, with maximum deviations of 10% from the State average, based on the 2001 Census.

specification for the disaggregated sample and interaction terms. The coefficient of interest is the triple interaction term:

$$winner_{ijt} * politically\ irrelevant_{ijt} * post\ announcement_{ijt}$$

A negative and significant coefficient would suggest that political corruption is weaker in areas where politicians have no incentive to buy votes. As shown in Table 9, the coefficient on the triple interaction is consistently insignificant.

In the absence of clear evidence for vote buying, it is possible that corruption arises within kinship networks because these provide the “particularised trust” needed to engage in risky collusive behaviour (Tonoyan, 2003). While we are unable to test this explanation explicitly, it fits the context of PMGSY and is consistent with the overall pattern of our results. The trust argument assumes that participants in corruption face a positive probability of detection and need to rely on each other’s discretion in an environment where their behaviour could be monitored. Our results are indicative of corruption in the areas for which wrongdoing would be hardest to detect ex-post: the allocation of roads and the size of the initial contract. In areas where corruption would be easier to detect (over-runs, delays, and quality) we find no or only weak evidence of inefficiencies. This pattern is consistent with a setting in which politicians and contractors are constrained by monitoring and operate on the least risky margins – allocation and total costs – and with the least risky collaborators – the members of their family or caste network.

10. CONCLUSION

This paper provides direct empirical evidence that local politicians in India abuse their power to benefit members of their own network. We exploit the variation in political leadership due to the electoral cycle, to identify systematic distortions in the allocation of contracts for a major rural road construction programme (PMGSY). By matching contractors’ and political candidates’ surnames, we generate a measure of proximity which evolves as the pool of contractors changes. A regression discontinuity design based on close elections, suggests that the causal impact of a politician coming to power is a 63% increase in the share of roads allocated to contractors who share their surname. This result withstands a series of alternative specifications and robustness checks. Further regression discontinuity estimates at the road level, indicate that political interference in the allocation of roads raises the cost of construction, without providing any offsetting benefits in terms of efficiency or quality. Corruption is therefore welfare-reducing in this context.

A distinguishing feature of our analysis, is that we identify the effect of political connections to state-level legislators who have no official involvement in the road construction programme. Our results therefore not only indicate preferential treatment of the politically connected, they also provide indirect evidence that local politicians' power over purportedly neutral bureaucrats is sufficient to coerce them into corruption. From a policy perspective, these findings indicate that more could be done to insulate the officials implementing government programmes at the local level, including those involved in PMGSY.

While this paper is primarily about the measurement of corruption, its findings have significance beyond the potential number of misallocated roads or the amount of misdirected money. If corrupt arrangements were made based on random matching between individuals, the empirical strategy would have revealed nothing. Our results provide further evidence for the role of networks in facilitating corruption and point towards theories in which kinship networks facilitate corruption through trust or the ability to impose social sanctions. The irony is, that the setting for the analysis – PMGSY – is conceptually a profoundly inclusive programme, facilitating the integration of over 100 million people into the Indian economy (Aggarwal 2015). This paper suggests that allowing them to compete equally for jobs, permits, licenses, or government procurement contracts, may require building more than roads.

REFERENCES

Acemoglu, Daron, Tarek Hassan, and Ahmed Tahoun. "The Power of the Streets: Evidence from Egypt's Arab Spring." Mimeo MIT (2014).

Aggarwal, Shilpa. "Do rural roads create pathways out of poverty? Evidence from India." *University of California, Santa Cruz, unpublished* (2014).

Amore, Mario Daniele, and Morten Bennedsen. "The value of local political connections in a low-corruption environment." *Journal of Financial Economics*110, no. 2 (2013): 387-402.

Angelucci, Manuela, Giacomo De Giorgi, Marcos A. Rangel, and Imran Rasul. "Family networks and school enrolment: Evidence from a randomized social experiment." *Journal of Public Economics* 94, no. 3 (2010): 197-221.

Asher, Sam, and Paul Novosad. "Politics and Local Economic Growth: Evidence from India." *Cambridge, MA: Harvard University, unpublished* (2015).

Asher, Sam, and Paul Novosad. "The Employment Effects of Road Construction in Rural India." *unpublished* (2014)

Banerjee, Abhijit and Rohini Pande. "Parochial Politics Ethnic Preferences and Politician Corruption." mimeo, Harvard (2009)

Banerjee, Abhijit, Sendhil Mullainathan, and Rema Hanna. *Corruption*. No. w17968. National Bureau of Economic Research, (2012).

Banerjee, Abhijit, Esther Duflo, Clement Imbert, Santosh Matthew, and Rohini Pande. "Can e-governance reduce capture of public programs? Experimental evidence from a financial reform of India's employment guarantee." Mimeo. (2014a).

Banerjee, Abhijit, Donald P. Green, Jeffery McManus, and Rohini Pande. "Are poor voters indifferent to whether elected leaders are criminal or corrupt? A vignette experiment in rural India." *Political Communication* 31, no. 3 (2014b): 391-407.

Bardhan, Pranab. "Corruption and development: a review of issues." *Journal of economic literature* (1997): 1320-1346.

Becker, Gary S., and George J. Stigler. "Law enforcement, malfeasance, and compensation of enforcers." *The Journal of Legal Studies* (1974): 1-18.

Bertrand, Marianne, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan. "Obtaining a driver's license in India: an experimental approach to studying corruption." *The Quarterly Journal of Economics* (2007): 1639-1676.

Bussell, Jennifer. "Clients or Constituents? Distribution Between the Votes in India." mimeo, University of California Berkeley (2015)

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel. 2014. "The Value of Democracy: Evidence from Road Building in Kenya", *American Economic Review (forthcoming).*

Cadot, Olivier. "Corruption as a Gamble." *Journal of Public Economics* 33, no. 2 (1987): 223-244.

Caughey, Devin, and Jasjeet S. Sekhon. "Elections and the regression discontinuity design: Lessons from close us house races, 1942–2008." *Political Analysis* 19, no. 4 (2011): 385-408.

Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein. "The value of school facility investments: Evidence from a dynamic regression discontinuity design." *The Quarterly Journal of Economics* 125, no. 1 (2010): 215-261.

Chandra, Kanchan. "Why Ethnic Parties Suceed." *Patronage and ethnic head counts in India* (2004).

Charron, Nicholas. "The correlates of corruption in India: Analysis and evidence from the states." *Asian Journal of Political Science* 18, no. 2 (2010): 177-194.

Chattopadhyay, Raghabendra, and Esther Duflo. "Women as policy makers: Evidence from a randomized policy experiment in India." *Econometrica* 72, no. 5 (2004): 1409-1443.

Chopra, Vir K. *Marginal players in marginal assemblies: The Indian MLA*. Orient Longman, 1996.

Do, Quoc-Anh, Yen Teik Lee, and Bang Dang Nguyen. "Political connections and firm value: evidence from the regression discontinuity design of close gubernatorial elections." *Available at SSRN 2190372* (2013).

Dollar, David, Raymond Fisman, and Roberta Gatti. "Are women really the "fairer" sex? Corruption and women in government." *Journal of Economic Behavior & Organization* 46, no. 4 (2001): 423-429.

Eggers, Andrew C., Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder. "On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races."*American Journal of Political Science* 59, no. 1 (2015): 259-274.

Ferraz, Claudio, and Federico Finan. "Exposing corrupt politicians: the effects of Brazilians publicly released audits on electoral outcomes." *Quarterly Journal of Economics* 123, no. 3 (2008): 703.

Ferraz, Claudio, and Frederico Finan. "Electoral Accountability and Corruption: Evidence from the Audits of Local Government." *American Economic Review* 101 (2011): 1274.

Field, Erica, Matthew Levinson, Rohini Pande, and Sujata Visaria. "Segregation, Rent Control, and Riots: The Economics of Religious Conflict in an Indian City." *The American Economic Review* (2008): 505-510.

Fisman, Raymond. "Estimating the value of political connections." *American Economic Review* (2001): 1095-1102.

Fisman, David A., Raymond J. Fisman, Julia Galef, Rakesh Khurana, and Yongxiang Wang. "Estimating the value of connections to Vice-President Cheney." *The BE Journal of Economic Analysis & Policy* 12, no. 3 (2012).

Fisman, Raymond, Daniel Paravisini, and Vikrant Vig. "Cultural proximity and loan outcomes". NBER Working Paper, No. 18096., 2012.

Fisma, Raymond, Florian Schulz, and Vikrant Vig. “Private Returns to Public Office”. *Journal of Political Economy* (2015).

Golden, Miriam A., and Lucio Picci. "Proposal for a new measure of corruption, illustrated with Italian data." *Economics & Politics* 17, no. 1 (2005): 37-75.

Goldman, Eitan, Jörg Rocholl, and Jongil So. "Do politically connected boards affect firm value?." *Review of Financial Studies* 22, no. 6 (2009): 2331-2360.

Hoff, Karla, and Priyanka Pandey. "Opportunity is not everything." *Economics of Transition* 13, no. 3 (2005): 445-472.

Horowitz, Donald L. *Ethnic groups in conflict*. Univ of California Press, 1985.

Huntington, S. "Modernisation and Corruption ‘in Political Order in Changing Societies, Yale University Press, New Havern." (1968).

Imbens, Guido, and Karthik Kalyanaraman. "Optimal bandwidth choice for the regression discontinuity estimator." *The Review of Economic Studies* (2011).

Iyer, Lakshmi, and Anandi Mani. "Traveling agents: political change and bureaucratic turnover in India." *Review of Economics and Statistics* 94, no. 3 (2012): 723-739.

Jensenius, Francesca Refsum. "Power, performance and bias: Evaluating the electoral quotas for scheduled castes in India." (2013).

Knack, Stephen, and Philip Keefer. "Institutions and economic performance: Cross-country tests using alternative institutional indicators." (1995): 207-228.

Krueger, Anne O. "The political economy of the rent-seeking society." *The American economic review* (1974): 291-303.

Lambsdorff, Johann Graf. "Corruption and rent-seeking." *Public choice* 113, no. 1-2 (2002): 97-125.

Larreguy, Horacio A., John Marshall, and James M. Snyder Jr. *Revealing Malfeasance: How Local Media Facilitates Electoral Sanctioning of Mayors in Mexico*. No. w20697. National Bureau of Economic Research, 2014.

Lee, David S. *The Electoral Advantage to Incumbency and Voters' Valuation of Politicians' Experience: A Regression Discontinuity Analysis of Elections to the US.*. No. w8441. National bureau of economic research, 2001.

Lee, David S. "Randomized experiments from non-random selection in US House elections." *Journal of Econometrics* 142, no. 2 (2008): 675-697.

Lee, David S., and Thomas Lemieux. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2010): 281-355.

Leff, Nathaniel H. "Economic development through bureaucratic corruption."*American behavioral scientist* 8, no. 3 (1964): 8-14.

Lindridge, Andrew, and Sally Dibb. "Is ‘culture’a justifiable variable for market segmentation? A cross- cultural example." *Journal of Consumer Behaviour* 2, no. 3 (2003): 269-286.

Lui, Francis T. "An equilibrium queuing model of bribery." *The journal of political economy* (1985): 760-781.

Mastrobuoni, Giovanni, and Eleonora Patacchini. "Organized crime networks: An application of network analysis techniques to the American mafia." *Review of Network Economics* 11, no. 3 (2012).

Mauro, Paolo. "Corruption and growth." *The Quarterly Journal of Economics* (1995): 681-712.

McCrary, Justin. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142, no. 2 (2008): 698-714.

Melo, Marcus André, Carlos Pereira, and Carlos Mauricio Figueiredo. "Political and institutional checks on corruption: Explaining the performance of Brazilian audit institutions." *Comparative Political Studies* (2009).

Mironov, Maxim and Zhuravskaya, Ekaterina. “Corruption in Procurement and the Political Cycle in Tunneling: Evidence from Financial Transactions Data” (2014), mimeo PSE.

Olken, Benjamin A., and Patrick Barron. "The simple economics of extortion: evidence from trucking in Aceh." (2007).

Olken, Benjamin A. "Corruption perceptions vs. corruption reality." *Journal of Public economics* 93, no. 7 (2009): 950-964.

Kramon, Eric and Daniel Posner. 2016. “Ethnic Favoritism in Education in Kenya.”, *Quarterly Journal of Political Science*, 11(1).

Prakash, Nishith, Marc Rockmore, and Yogesh Uppal. 2015. “Do criminally accused politicians affect economic outcomes? Evidence from India”, *Mimeo.*

Reinikka, Ritva, and Jakob Svensson. "Local capture: evidence from a central government transfer program in Uganda." *The Quarterly Journal of Economics*(2004): 679-705.

Rose-Ackerman, Susan. "The economics of corruption." *Journal of public economics* 4, no. 2 (1975): 187-203.

Rose-Ackerman, Susan. *Corruption and government: Causes, consequences, and reform*. Cambridge university press, 1999.

Schumacker, Randall E. "Regression discontinuity models and the variance inflation factor." *Multiple Linear Regression Viewpoints* 34, no. 1 (2008): 13.

Shleifer, Andrei, and Robert W. Vishny. "Corruption.” *The Quarterly Journal of Economics* 108, no. 3 (1993): 599-617

Tonoyan, Vartuhí. "Corruption and Entrepreneurship: Impact of Trust and Civic Society; a Cross Cultural Comparison of 21 Countries; Paper Presented at the International Conference on" Trust and Entrepreneurial Behavior in East and West European Economies: Concepts, Developments and Comparative Aspects", University of Bremen 2003." ifm, 2003.

The World Bank, “Curbing Fraud, Corruption, and Collusion in the Roads Sector," Technical Report, The World Bank, Washington, DC 2011.

Treisman, Daniel. "The causes of corruption: a cross-national study." *Journal of public economics* 76, no. 3 (2000): 399-457.

Vissa, Balagopal. "A matching theory of entrepreneurs' tie formation intentions and initiation of economic exchange." *Academy of Management Journal* 54, no. 1 (2011): 137-158.

APPENDIX: MAIN TABLES

Table 1: Descriptive Statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
Roads allocated to contractors of the same name					
Share_{t-1}	8116	0.037	0.141	0.000	1.000
Share_{t}	8116	0.035	0.134	0.000	1.000
ΔShare	8116	-0.002	0.151	-1.000	1.000
Candidate characteristics					
Vote share	8036	0.279	0.103	0.020	0.837
Margin	8116	0.000	0.102	-0.695	0.695
Incumbent	8116	0.282	0.450	0.000	1.000
Runner-up previous election	8116	0.147	0.354	0.000	1.000
Age	7357	49.164	10.207	23	87
Female candidate	8116	0.062	0.242	0.000	1.000
Candidate with criminal charges	4434	0.182	0.386	0.000	1.000
Total assets (1000000s of INR)	5010	106	4740	0.000	300000
Liabilities (1000000s of INR)	5286	1.841	17.800	0.000	644
University graduate	5286	0.596	0.491	0.000	1.000
Postgraduate degree	5286	0.192	0.394	0.000	1.000
Congress candidate	7239	0.295	0.456	0.000	1.000
BJP candidate	7239	0.203	0.403	0.000	1.000
Named Kumar	8116	0.058	0.234	0.000	1.000
Named Lal	8116	0.022	0.145	0.000	1.000
Named Patel	8116	0.009	0.094	0.000	1.000
Named Ram	8116	0.018	0.133	0.000	1.000
Named Reddy	8116	0.016	0.124	0.000	1.000
Named Singh	8116	0.112	0.316	0.000	1.000
Named Yadav	8116	0.014	0.117	0.000	1.000
Constituency characteristics					
Reserved seat	8116	0.335	1.349	0.000	84.092
Road count_{t}	8116	27.691	30.822	1.000	479
Road count_{t-1}	8116	22.086	25.744	1.000	388
Mean road length_{t}	8116	5.833	3.999	0.350	42.654
Mean road length_{t-1}	8116	4.963	3.838	0.410	53.985
Mean population	7822	961.697	633.986	30.000	7230
Mean SC/ST population	7822	244.078	193.401	0.000	2283
Mean connectivity	7822	0.561	0.308	0.000	1.000

Note: The number of observations varies due to missing values. Reserved seat refers to constituency reserved for MLAs from scheduled castes or scheduled tribes. Road count_{t} is computed at the term-level by counting the number of road contracts signed in a constituency within a term. Mean road length is the average length of roads (in km) built in a constituency and term. Mean population and mean SC/ST population are averages of 2001 census data for all of a constituency's villages. Mean connectivity is the share of a constituency's villages that had all-weather road access at the time of the 2001 census.

Table 2: Randomization for local linear regression at 6.2% bandwidth

	Observations	Winner	Standard error
Candidate characteristics			
Share of same name contractors$_{t-1}$	4,396	-0.0072	(0.0084)
Incumbent	4,396	-0.0366	(0.0296)
Runner-up in previous election	4,396	-0.0119	(0.0214)
Age	4,036	0.2874	(0.6014)
Female candidate	4,396	-0.0004	(0.0129)
Candidate with criminal charge	2,536	-0.0421	(0.0293)
Total assets (1000000s of INR)	2,770	261.119	(718.93)
Liabilities (1000000s of INR)	3,049	0.1260	(0.5146)
Candidate with university degree	3,049	-0.0139	(0.0328)
Candidate with post-grad. degree	3,049	-0.0047	(0.0271)
BJP candidate	3,940	0.0104	(0.0274)
Congress candidate	3,940	-0.0194	(0.0312)
Pre-trend: share of roads built by contractors of same name in term prior to election			
Share 5 years before election	2,502	0.0078	(0.0111)
Share 4 years before election	2,898	-0.0151	(0.0124)
Share 3 years before election	2,634	-0.0096	(0.0116)
Share 2 years before election	1,688	0.0037	(0.0133)
Share 1 year before election	1,866	-0.0131	(0.0152)
Prevalence of most common names			
Named Kumar	4,396	0.0119	(0.0136)
Named Lal	4,396	-0.0019	(0.0084)
Named Patel	4,396	0.0026	(0.0061)
Named Ram	4,396	-0.0021	(0.0076)
Named Reddy	4,396	0.0074	(0.0054)
Named Singh	4,396	0.0195	(0.0171)
Named Yadav	4,396	0.0052	(0.0076)

Note: Coefficients are estimated by regressing the row variables on winner, the vote margin, and the vote margin interacted with winner in OLS regressions Standard errors are clustered at the election level. The bandwidth of 6.2% is derived from the optimal bandwidth choice rule of Imbens and Kalyanaraman (2012). *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Local linear regression RD

ΔShare of same name contractors$_t$	Whole Sample		Margin of Victory <6.2%		Margin of Victory <5%		Margin of Victory <2.5%	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Winner	0.0094*	0.0094*	0.0252***	0.0243**	0.0202**	0.0196*	0.0267*	0.0259*
	(0.0052)	(0.0055)	(0.0092)	(0.0099)	(0.0102)	(0.0111)	(0.0141)	(0.0153)
Margin	-0.0001	0.0271	-0.3153*	-0.2710	-0.2949	-0.2591	-1.0043	-1.1322
	(0.0299)	(0.0345)	(0.1856)	(0.1886)	(0.2389)	(0.2556)	(0.6660)	(0.7430)
Margin*winner	-0.0059	-0.0504	0.0835	0.0562	0.3363	0.3259	1.4266	1.7642*
	(0.0414)	(0.0456)	(0.2724)	(0.2734)	(0.3507)	(0.3720)	(0.8778)	(0.9802)
Incumbent		-0.0014		0.0031		0.0036		-0.0108
		(0.0046)		(0.0065)		(0.0072)		(0.0093)
Runner-up in previous election		0.0067		0.0081		0.0081		0.0052
		(0.0056)		(0.0079)		(0.0085)		(0.0125)
Female candidate		-0.0011		-0.0135		-0.0113		-0.0054
		(0.0066)		(0.0107)		(0.0097)		(0.0138)
Age		0.0002		-0.0001		0.0000		0.0003
		(0.0002)		(0.0003)		(0.0003)		(0.0004)
AC controls		X		X		X		X
State fixed effects		X		X		X		X
Election year fixed effects		X		X		X		X
N	8,116	7,068	4,396	3,921	3,760	3,365	2,104	1,880

Note: Local linear regression estimates. Standard errors are clustered at the election-level. Variables are defined either in the text or in the note for table 1. AC controls include: Reserved seat, Road count$_{t-1}$, Mean population, Mean SC/ST population, Mean connectivity, Mean road length$_{t-1}$. The bandwidth of 6.2% is derived from the optimal bandwidth choice rule of Imbens and Kalyanaraman (2012). *** p<0.01, ** p<0.05, * p<0.1

Table 4: Parametric regression discontinuity estimated on full sample

$\Delta Share_t$	Linear		Quadratic Polynomials		Cubic Polynomials	
	(1)	(2)	(3)	(4)	(5)	(6)
Winner	0.0094*	0.0094*	0.0138**	0.0138**	0.0190**	0.0193**
	(0.0052)	(0.0055)	(0.0062)	(0.0067)	(0.0075)	(0.0082)
Margin	-0.0001	0.0271	-0.0226	-0.0058	-0.1004	-0.1562
	(0.0299)	(0.0345)	(0.0609)	(0.0719)	(0.1209)	(0.1506)
Margin*Winner	-0.0059	-0.0504	-0.0740	-0.0998	-0.1247	-0.0253
	(0.0414)	(0.0456)	(0.0870)	(0.0994)	(0.1604)	(0.1890)
Margin^2			-0.0764	-0.1130	-0.6165	-1.2326
			(0.1540)	(0.1921)	(0.6727)	(0.9282)
Margin*Winner ^2			0.3834	0.3849	1.8161*	2.0317
			(0.2689)	(0.2869)	(1.0541)	(1.2413)
Margin ^3					-0.8224	-1.8595
					(0.8418)	(1.3418)
Margin*Winner ^3					-0.5365	1.0619
					(1.0901)	(1.5242)
Constituency-level controls		X		X		X
Candidate-level controls		X		X		X
State fixed effects		X		X		X
Election year fixed effects		X		X		X
N	8,116	7,068	8,116	7,068	8,116	7,068

Note: Estimated by OLS. Standard errors are clustered at the election-level. Variables are defined either in the text or in the note for table 1. The candidate and constituency level controls in columns 2, 4, 6 and 8 are the same as in column (2) of Table 3. These controls and state and election year fixed effects are not reported. All regressions include a constant. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Road-level regression discontinuity - estimates for cost

Ln(sanctioned cost/km)	Margin of victory <5%		Margin of victory <2.7%		Margin of victory <2.5%	
	(1)	(2)	(3)	(4)	(5)	(6)
MLAsamename	0.2328***	0.0683*	0.3122***	0.1238**	0.2735**	0.1320**
	(0.0895)	(0.0376)	(0.1166)	(0.0522)	(0.1170)	(0.0561)
Margin	-3.5799	-1.8487*	-9.0103*	-6.9359**	-10.3331*	-8.0188**
	(2.3060)	(1.0848)	(5.4485)	(2.8075)	(5.8739)	(3.1738)
Margin* MLAsamename	-1.5417	2.6023*	4.3953	7.5140**	10.7941	8.6252**
	(3.5541)	(1.5201)	(7.9768)	(3.6619)	(8.6207)	(3.9456)
Months since election	0.0045***	-0.0008	0.0054***	-0.0010	0.0056***	-0.0008
	(0.0012)	(0.0006)	(0.0013)	(0.0007)	(0.0013)	(0.0008)
Ln(length)	0.0171	-0.0554**	-0.0407	-0.0718**	-0.0478	-0.0695*
	(0.0357)	(0.0257)	(0.0478)	(0.0349)	(0.0500)	(0.0366)
Reserved seat	-0.0155	0.0491	-0.1598	0.0556	-0.2165	0.0562
	(0.0936)	(0.0402)	(0.1275)	(0.0627)	(0.1353)	(0.0719)
Mean population of habitations	-0.0299	-0.0129	-0.0401	-0.0225	-0.0502	-0.0230
	(0.0302)	(0.0111)	(0.0355)	(0.0158)	(0.0369)	(0.0165)
SCST share of habitations	-0.1258	0.0061	-0.1210	0.0352	-0.0082	0.0524
	(0.0874)	(0.0442)	(0.1144)	(0.0640)	(0.1089)	(0.0661)
State fixed effects		X		X		X
Agreement year fixed effects		X		X		X
N	2,418	2,418	1,542	1,542	1,435	1,435

Note: Standard errors clustered at the contractor level to account for intra-contractor correlation of the error term at the road level. We use Ln(length) to account for non-linear relationship between cost and distance. The bandwidth of 2.7% is derived from the optimal bandwidth choice rule of Imbens and Kalyanaraman (2012). *** p<0.01, ** p<0.05, * p<0.1

Table 6: Road-level regression discontinuity - estimates for quality

Dependent variable:	Days overrun			Ratio: actual cost to sanctioned cost			Failed inspection		
Margin of victory:	<5%	<3%	<2.5%	<5%	<3.5%	<2.5%	<5%	<4.3%	<2.5%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MLAsamename	-66.4	-40.7	-76.4	-0.0308	-0.0322	-0.0207	0.1470**	0.1377*	0.0913
	(65.5)	(86.6)	(97.2)	(0.0340)	(0.0398)	(0.0518)	(0.0721)	(0.0771)	(0.1012)
Margin	-271.6	928.9	531.5	0.9589	2.0515	1.3936	-0.9374	-0.2468	-0.2114
	(1,792.1)	(3,929.1)	(4,058.0)	(1.2139)	(1.8813)	(3.2588)	(2.1666)	(2.7267)	(5.0508)
Margin*MLAsamename	1,364.8	-2,541.6	1,794.2	-0.8923	-2.8526	-1.8681	-4.5103	-5.1430	-2.8051
	(2,690.5)	(5,122.6)	(6,198.5)	(1.4314)	(2.2585)	(4.0687)	(2.9817)	(3.7194)	(6.5823)
Months since election	1.8	2.4*	3.7**	0.0007*	0.0004	0.0010*	0.0016	0.0013	-0.0007
	(1.1505)	(1.3)	(1.5024)	(0.0004)	(0.0004)	(0.0006)	(0.0014)	(0.0015)	(0.0018)
Ln(length)	36.6**	44.6**	53.7**	0.0181*	0.0164	0.0156	0.0120	0.0142	0.0101
	(17.8)	(20.1)	(21.2)	(0.0092)	(0.0104)	(0.0140)	(0.0231)	(0.0241)	(0.0300)
Reserved seat	127.0*	190.4**	188.4*	0.0140	0.0281	0.0512	0.0123	0.0056	0.1024
	(65.9)	(74.5)	(99.1)	(0.0437)	(0.0529)	(0.0852)	(0.0509)	(0.0529)	(0.0716)
Mean population of habitations	-22.1	-28.6	-17.2	-0.0020	-0.0026	0.0043	0.0209	0.0237	-0.0245
	(15.3)	(18.1)	(19.2)	(0.0045)	(0.0047)	(0.0061)	(0.0236)	(0.0245)	(0.0358)
SCST share of habitations	-24.0	-54.4	-115.4*	0.0012	-0.0007	-0.0083	-0.1562**	-0.1468*	-0.0471
	(61.3)	(64.7)	(67.4)	(0.0214)	(0.0259)	(0.0368)	(0.0770)	(0.0797)	(0.1140)
State fixed effects	X	X	X	X	X	X	X	X	X
Agreement year FE	X	X	X	X	X	X	X	X	X
N	1,604	1,139	940	1,933	1,570	1,139	820	758	482

Note: Standard errors clustered at the contractor level to account for intra-contractor correlation of the error term at the road level. The optimal bandwidths of 3% (for days overrun), 3.5% (for cost overruns) and 4.3% (for failed inspections) are derived from the optimal bandwidth choice rule of Imbens and Kalyanaraman (2012). All regressions include a constant. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: RD test for electoral cycles in preferential allocation

$\Delta Share_t$	Start of term heterogeneity		End of term heterogeneity		Start and end of term heterogeneity	
	(1)	(2)	(3)	(4)	(5)	(6)
Winner	0.0249***	0.0288***	0.0264***	0.0246**	0.0260***	0.0236**
	(0.0081)	(0.0089)	(0.0084)	(0.0108)	(0.0087)	(0.0113)
Margin	-0.1926	-0.1764	-0.1891	0.0323	-0.1898	0.0282
	(0.1952)	(0.2095)	(0.1953)	(0.2605)	(0.1952)	(0.2603)
Margin*winner	-0.2357	-0.3650	-0.2411	-0.4818	-0.2400	-0.4740
	(0.3363)	(0.3590)	(0.3363)	(0.4445)	(0.3363)	(0.4442)
Start of term	-0.0058	-0.0060			-0.0035	-0.0083
	(0.0053)	(0.0057)			(0.0052)	(0.0066)
Start of term* winner	0.0021	0.0021			0.0010	0.0023
	(0.0069)	(0.0072)			(0.0070)	(0.0084)
End of term			0.0078	0.0174**	0.0063	0.0138
			(0.0058)	(0.0087)	(0.0057)	(0.0086)
End of term* winner			-0.0034	-0.0103	-0.0030	-0.0093
			(0.0069)	(0.0092)	(0.0070)	(0.0094)
Constituency Controls		X		X		X
Candidate Controls		X		X		X
State fixed effects		X		X		X
Agreement year fixed effects		X		X		X
N	6,266	5,572	6,266	4,188	6,266	4,188

Note: Standard errors clustered at the election level. All estimates conducted on 5% bandwidth. All regressions include a constant. For this analysis there are potentially three observations per electoral term: the value of $Share_t$ for the first 12 months after an election, the value of $Share_t$ for the last 12 months before the next election, and the value of $Share_t$ over the remaining term. For constituencies where no roads were built in one of these periods, the number of observations will be less than three. *** p<0.01, ** p<0.05, * p<0.1

Table 8: RD test for electoral cycles in cost

Ln(sanctioned cost/km)	Start of term heterogeneity		End of term heterogeneity		Start and end of term heterogeneity	
	(1)	(2)	(3)	(4)	(5)	(6)
MLAsamename	0.311***	0.137***	0.298***	0.146***	0.332***	0.163***
	(0.110)	(0.053)	(0.111)	(0.054)	(0.116)	(0.056)
Margin	-9.933**	-7.107**	-10.489**	-7.053**	-10.135**	-7.063**
	(4.921)	(2.797)	(5.054)	(2.808)	(4.897)	(2.776)
Margin*MLAsamename	7.022	7.866**	7.601	7.486**	7.462	7.682**
	(6.935)	(3.654)	(7.034)	(3.651)	(6.918)	(3.631)
Start of term	0.176*	0.133**			0.222**	0.184***
	(0.102)	(0.054)			(0.107)	(0.062)
Start of term* MLAsamename	-0.237**	-0.092*			-0.256**	-0.111**
	(0.104)	(0.054)			(0.109)	(0.054)
End of term			-0.019	0.032	-0.009	-0.000
			(0.098)	(0.082)	(0.104)	(0.085)
End of term* MLAsamename			-0.106	-0.134**	-0.151	-0.156***
			(0.105)	(0.057)	(0.110)	(0.059)
Road level controls	X	X	X	X	X	X
State fixed effects		X		X		X
Agreement year fixed effects		X		X		X
N	1,542	1,542	1,542	1,542	1,542	1,542

Note: Standard errors clustered at the contractor level. All estimates conducted on 5% bandwidth. The set of road-level controls is the same as in Table 5 and Table 6. All regressions include a constant. *** p<0.01, ** p<0.05, * p<0.1

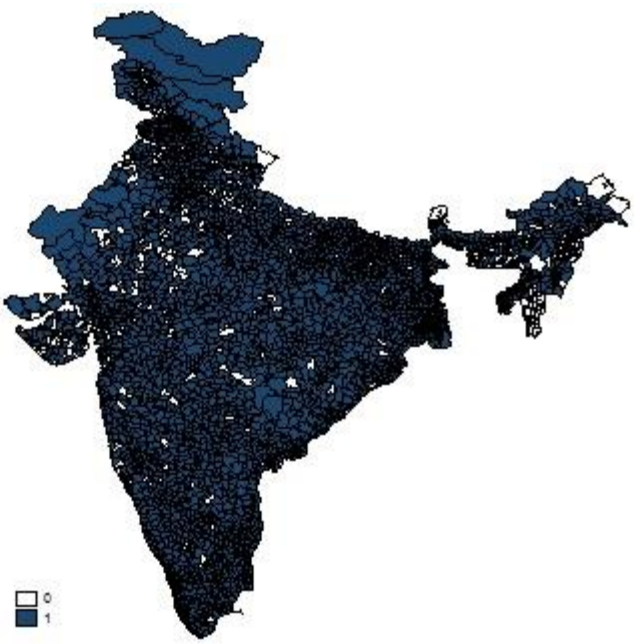
Table 9: RD test for heterogeneity based on "political relevance"

Δ $Share_t$	Full sample with interactions		Margin of victory <6.2%		Margin of victory <5%		Margin of victory <2.5%	
	(1)	(2)					(3)	(4)
Winner	0.0114**	0.0138**	0.0300***	0.0279**	0.0267**	0.0297**	0.0226	0.0200
	(0.0057)	(0.0061)	(0.0101)	(0.0109)	(0.0106)	(0.0116)	(0.0140)	(0.0151)
Margin	0.0139	0.0124	-0.1183	-0.0961	-0.1533	-0.1654	-0.2873	-0.2439
	(0.0282)	(0.0305)	(0.1729)	(0.1882)	(0.2215)	(0.2394)	(0.6242)	(0.7029)
Margin*winner	-0.0138	-0.0187	-0.3593	-0.2924	-0.0746	-0.1029	0.2786	0.6273
	(0.0445)	(0.0485)	(0.2589)	(0.2692)	(0.3233)	(0.3526)	(0.8525)	(0.9580)
Politically irrelevant	0.0028	0.0078	-0.0060	0.0034	-0.0030	0.0106	0.0045	0.0157
	(0.0052)	(0.0067)	(0.0079)	(0.0102)	(0.0086)	(0.0112)	(0.0107)	(0.0145)
Post announcement	0.0047	0.0062	0.0045	0.0077	0.0067	0.0120	-0.0033	-0.0008
	(0.0070)	(0.0085)	(0.0097)	(0.0114)	(0.0107)	(0.0125)	(0.0171)	(0.0196)
Politically irrelevant*winner	-0.0069	-0.0103	-0.0109	-0.0151	-0.0043	-0.0160	-0.0054	-0.0142
	(0.0082)	(0.0088)	(0.0119)	(0.0129)	(0.0124)	(0.0136)	(0.0162)	(0.0177)
Post announcement* winner	-0.0151	-0.0141	-0.0165	-0.0131	-0.0150	-0.0136	0.0051	0.0072
	(0.0103)	(0.0108)	(0.0145)	(0.0150)	(0.0158)	(0.0163)	(0.0230)	(0.0238)
Politically irrelevant*post announcement	-0.0032	-0.0062	0.0103	0.0016	0.0097	-0.0025	0.0123	0.0034
	(0.0099)	(0.0114)	(0.0143)	(0.0165)	(0.0158)	(0.0181)	(0.0220)	(0.0249)
Politically irrelevant*post announcement* winner	0.0063	0.0060	0.0110	0.0144	-0.0049	0.0070	-0.0168	-0.0080
	(0.0147)	(0.0156)	(0.0206)	(0.0218)	(0.0219)	(0.0231)	(0.0295)	(0.0315)
Constituency and candidate level controls		X		X		X		X
State fixed effects		X		X		X		X
Agreement year fixed effects		X		X		X		X
N	9,774	8,462	5,234	4,658	4,532	4,063	2,580	2,327

Note: Observations at the MLA level. Standard errors are clustered at the election-level. As in Table 7, the term-level sample is disaggregated, allowing for multiple observations per term. The sample for Columns 3 and 4 is restricted (i) to areas that did not remain part of the same constituency after delimitation and (ii) to the time period between the announcement of the delimitation reform and the first election under the new delimitation. *** p<0.01, ** p<0.05, * p<0.1.

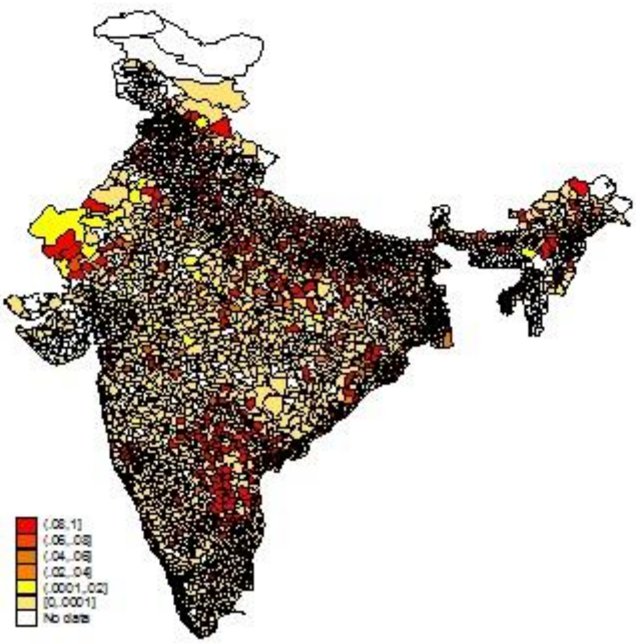
APPENDIX: MAPS

MAP A.1: CONSTITUENCIES IN SAMPLE



Note: Constituencies shaded blue are in the sample (pre-2008 delimitation)

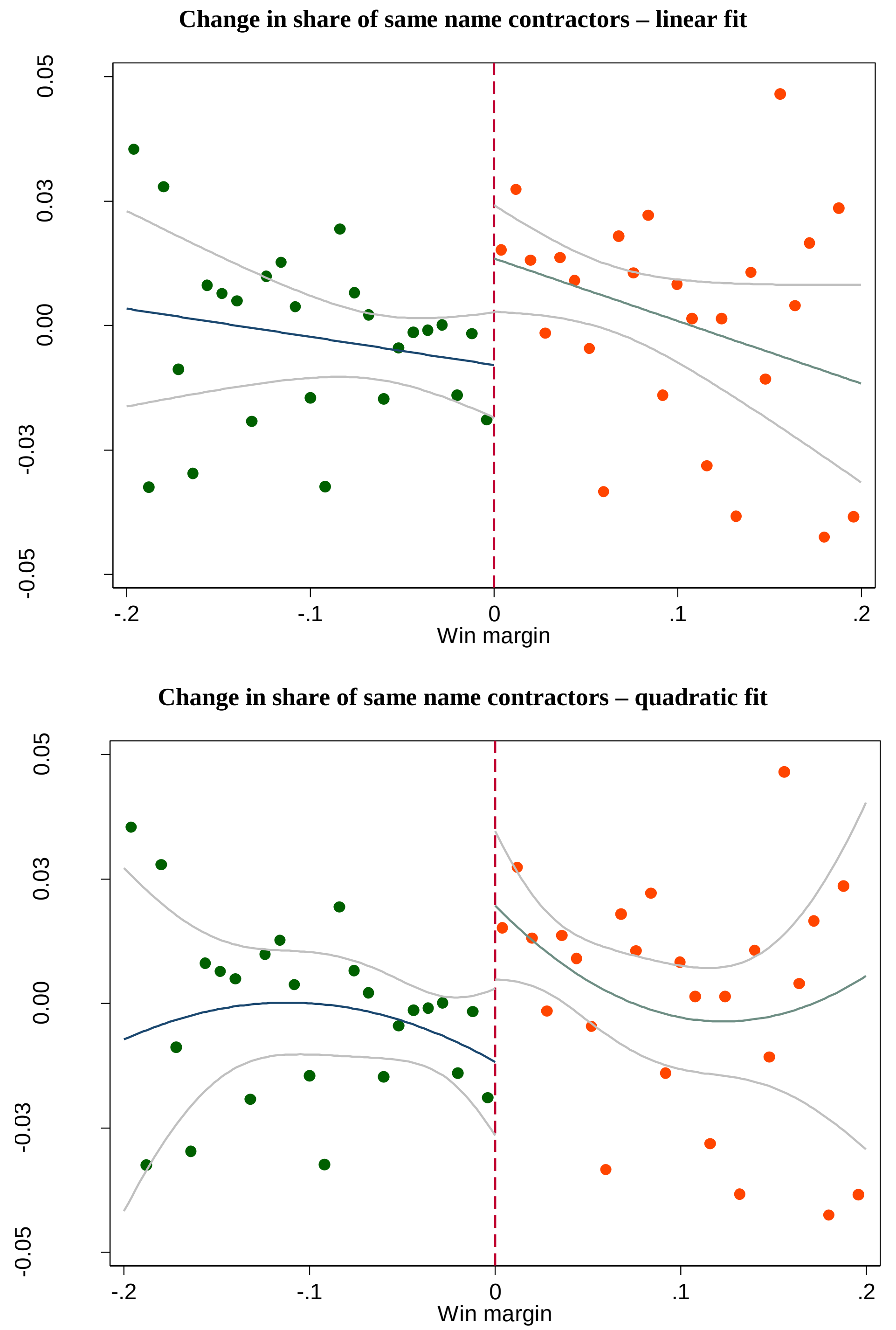
MAP A.2: VARIATION IN THE BASELINE SHARE OF SAME NAME CONTRACTORS



Note: Map plots $share_{ijt}$ for all constituencies in the sample. Darker shades indicate a higher value of $share_{ijt}$.

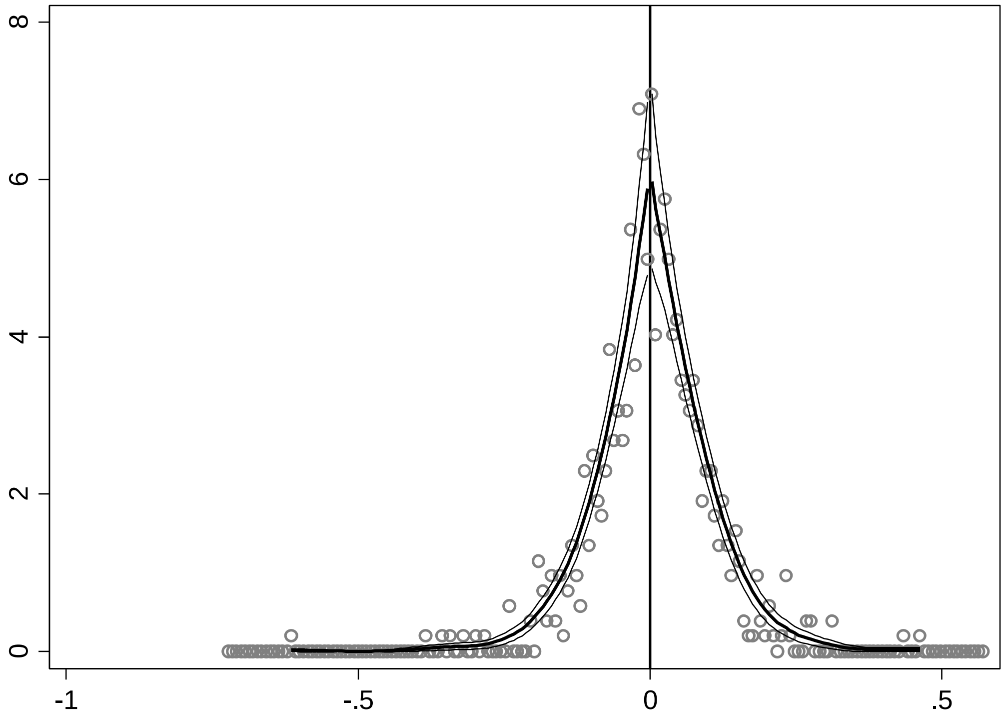
APPENDIX: FIGURES

FIGURE A.1: GRAPHICAL DEPICTION OF RD



Note: Lines fitted separately on the samples left and right of the cut-off. 5% confidence intervals plotted in grey. Each marker represents a bin of 100 observations.

FIGURE A.2: GRAPHICAL DEPICTION OF THE MCCRARY TEST



Note: This figure plots the McCrary test. The running variable in this analysis is continuously distributed by construction. The test is performed on an alternative version of the margin variable: the margin of victory for the candidate with the higher level of $share_{ijt-1}$.

APPENDIX: ADDITIONAL TABLES

Table A1: Sources of election data

Source	Years covered	No. of elections	No. of candidates	Candidate-level variables used in the sample
ECI digitised data	2005-2014	7,328	80,323	name, vote share, gender, party
Bhavani (2012)	1977-2012	31,422	300,087	name, vote share, gender, party
Empowering India	1951-2015	19,715	196,935	assets, education, age
National Election Watch	2004-2015	8,944	73,200	assets, liabilities, education, criminal charges

Note: ECI digitised data refers to a subset of the ECI data that are available online at eci.nic.in. Bhavani (2012) is a dataset kindly made public by Rikhil R. Bhavnani. Empowering India and National Election Watch are NGOs. Their data is accessible at www.empoweringindia.com and myneta.info respectively. Where a variable is listed twice in the fifth column, this is due to incomplete time series or missing values that are filled in by drawing on multiple datasets.

Table A2: Local linear regression RD - levels

Share of same name contractors$_t$	Whole Sample		Margin of Victory <6.2%		Margin of Victory <5%		Margin of Victory <2.5%	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Winner	0.0085***	0.0116**	0.0121**	0.0161**	0.0097*	0.0082	0.0149*	0.0187*
	(0.0032)	(0.0047)	(0.0054)	(0.0079)	(0.0059)	(0.0089)	(0.0077)	(0.0111)
Margin	0.0119	-0.0084	-0.0891	-0.1013	-0.0298	0.0437	-0.7048*	-0.8124*
	(0.0212)	(0.0250)	(0.1120)	(0.1443)	(0.1443)	(0.1889)	(0.4017)	(0.4440)
Margin*winner	-0.0364	0.0114	0.0632	0.1070	0.0770	0.2981	1.0360	1.1712
	(0.0349)	(0.0435)	(0.1802)	(0.2385)	(0.2476)	(0.3435)	(0.6630)	(0.7479)
Incumbent		0.0085**		0.0140**		0.0155**		0.0106
		(0.0040)		(0.0056)		(0.0062)		(0.0079)
Runner-up in previous election		0.0067		0.0020		0.0043		0.0026
		(0.0048)		(0.0067)		(0.0073)		(0.0103)
Female candidate		-0.0192***		-0.0174**		-0.0207***		-0.0115
		(0.0049)		(0.0071)		(0.0070)		(0.0118)
Age		0.0000		-0.0001		-0.0001		0.0000
		(0.0002)		(0.0002)		(0.0002)		(0.0003)
AC controls		X		X		X		X
State fixed effects		X		X		X		X
Election year fixed effects		X		X		X		X
N	15,208	7,068	8,202	3,921	6,964	3,365	3,816	1,880

Note: Local linear regression estimates. Standard errors are clustered at the election-level. Variables are defined either in the text or in the note for table 1. AC controls include: Reserved seat, Road count$_{t-1}$, Mean population, Mean SC/ST population, Mean connectivity, Mean road length$_{t-1}$. *** p<0.01, ** p<0.05, * p<0.1

Table A3: Local linear regression RD – BIMAROU states

ΔShare of same name contractors$_t$	Whole Sample		Margin of Victory <6.2%		Margin of Victory <5%		Margin of Victory <2.5%	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Winner	0.0166*	0.0151	0.0325**	0.0317**	0.0334**	0.0314*	0.0507**	0.0503**
	(0.0086)	(0.0092)	(0.0143)	(0.0154)	(0.0162)	(0.0174)	(0.0235)	(0.0255)
Margin	-0.0257	-0.0220	-0.4830*	-0.4748	-0.8118**	-0.8145**	-2.3617**	-2.6942**
	(0.0746)	(0.0839)	(0.2804)	(0.2924)	(0.3605)	(0.3851)	(1.1230)	(1.2365)
Margin*winner	-0.0358	-0.0263	0.1807	0.1354	0.7914	0.8428	2.7213*	3.3305**
	(0.1044)	(0.1201)	(0.3909)	(0.4058)	(0.5180)	(0.5353)	(1.4589)	(1.5414)
Incumbent		-0.0031		0.0017		0.0007		-0.0094
		(0.0074)		(0.0099)		(0.0110)		(0.0152)
Runner-up in previous election		0.0067		0.0131		0.0152		0.0172
		(0.0083)		(0.0111)		(0.0125)		(0.0173)
Female candidate		-0.0261***		-0.0464***		-0.0372***		-0.0211
		(0.0100)		(0.0157)		(0.0142)		(0.0171)
Age		0.0001		-0.0001		-0.0002		0.0003
		(0.0003)		(0.0004)		(0.0004)		(0.0005)
AC controls		X		X		X		X
State fixed effects		X		X		X		X
Election year fixed effects		X		X		X		X
N	3,386	3,143	2,114	1,967	1,818	1,700	1,064	993

Note: Local linear regression estimates. Standard errors are clustered at the election-level. Variables are defined either in the text or in the note for table 1. AC controls include: Reserved seat, Road count$_{t-1}$, Mean population, Mean SC/ST population, Mean connectivity, Mean road length$_{t-1}$. *** p<0.01, ** p<0.05, * p<0.1

Table A4: Road-level regression discontinuity estimates for cost: BIMAROU states

Ln(sanctioned cost/km)	Margin of victory <5%		Margin of victory <2.7%		Margin of victory <2.5%	
	(1)	(2)	(3)	(4)	(5)	(6)
MLAsamename	0.206**	0.060*	0.424***	0.136***	0.389***	0.131***
	(0.088)	(0.032)	(0.110)	(0.043)	(0.110)	(0.045)
Margin	-3.139	-2.255**	-16.140***	-8.277***	-14.842***	-7.589***
	(2.308)	(0.961)	(5.214)	(2.001)	(5.545)	(2.133)
Margin* MLAsamename	0.650	2.946**	8.211	8.197***	9.319	7.670**
	(3.185)	(1.313)	(7.069)	(2.874)	(7.434)	(3.155)
Months since election	0.003***	-0.001	0.004***	-0.001	0.005***	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(length)	0.008	-0.025**	-0.013	-0.016	-0.009	-0.012
	(0.022)	(0.011)	(0.025)	(0.013)	(0.026)	(0.013)
Reserved seat	0.126	0.021	0.142	0.028	0.049	-0.006
	(0.084)	(0.029)	(0.104)	(0.035)	(0.108)	(0.037)
Mean population of habitations	0.070***	-0.001	0.060***	-0.004	0.063***	-0.003
	(0.019)	(0.008)	(0.022)	(0.011)	(0.023)	(0.011)
SCST share of habitations	-0.399***	-0.042	-0.396***	-0.052	-0.321***	-0.021
	(0.075)	(0.033)	(0.092)	(0.042)	(0.093)	(0.041)
State fixed effects		X		X		X
Agreement year fixed effects		X		X		X
N	1,696	1,696	1,149	1,149	1,078	1,078

Note: Standard errors clustered at the contractor level to account for intra-contractor correlation of the error term at the road level. We use Ln(length) to account for non-linear relationship between cost and distance. *** p<0.01, ** p<0.05, * p<0.1

Table A5: Cost estimates with additional controls (2.7% bandwidth)

Dependent variable: Ln(sanctioned cost/km)	(1)	(2)	(3)	(4)	(5)
MLAsamename	0.083**	0.084**	0.078*	0.083*	0.079*
	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)
Margin	-5.298***	-5.276***	-4.970**	-5.349***	-4.998**
	(1.957)	(1.955)	(1.968)	(1.964)	(1.972)
Margin* MLAsamename	5.562**	5.536*	5.310*	5.596**	5.316*
	(2.830)	(2.828)	(2.771)	(2.817)	(2.750)
Bridge		0.258			0.267
		(0.272)			(0.265)
Altitude			-0.068**		-0.069**
			(0.033)		(0.034)
Ruggedness				14.798	15.849
				(13.753)	(12.941)
Road level controls	X	X	X	X	X
State fixed effects	X	X	X	X	X
Agreement year fixed effects	X	X	X	X	X
N	1,394	1,394	1,394	1,394	1,394

Note: Standard errors clustered at the contractor level to account for intra-contractor correlation of the error term at the road level. The bandwidth of 2.7% is derived from the optimal bandwidth choice rule of Imbens and Kalyanaraman (2012) All regressions include the same set of road level controls as Table 6. *** p<0.01, ** p<0.05, * p<0.1