ABSTRACT

Title of Dissertation: Essays on Institutions, Governance and Economic Growth

Kartikeya Batra
Doctor of Philosophy, 2024

Committee Members: Prof Sebastian Galiani (Chair, UMD), Prof Ethan Kaplan (UMD), Prof Sam Asher (Imperial College London), Prof Kenneth Leonard (UMD), Prof Sonalde Desai (UMD)

Economic development and growth are impacted by several factors. Among these, existing social institutions, and quality of governance are important determinants. These factors become especially relevant in the context of low and middle-income countries. Such nations are home to a large share of the world's population, and aspire to grow their economies at high rates. Understanding constraints to their socio-economic development and prescribing policy solutions is, therefore, an important area of research. In the three chapters of this dissertation, I explore three different issues that impact social institutions and governance, which, in turn, impact socio-economic development. I do so in the context of India, which is home to approximately 20% of the world's total population.

In the first chapter, I explore whether historical land policies impact long-run socio-economic outcomes, including the persistent institution of the caste system and stereotypes associated with it. I find that lower land concentration does lead to improved socio-economic outcomes, especially for the socially marginalized landless communities. In the second chapter, I test whether enhanced state capacity by means of better public infrastructure improves the performance of bureaucrats in rural India. I find that better roads enhance bureaucratic performance, possibly due to improved monitoring by higher officials whose mobility is positively impacted.

Finally, in the third chapter, I examine whether the size of a political party impacts its decisions to field wealthy candidates. I find that a smaller political party is likely to field a wealthier candidate than a bigger political party, possibly due to fewer avenues to mobilize resources. This is important, for the wealth profile of a candidate, in turn, has the potential to impact governance outcomes in their area.

The three chapters are aimed at understanding causal relationships pertaining to important questions in the context of India's society, political economy and economic development. My results provide novel contributions to relevant strands literature, and also allow me to provide relevant policy prescriptions.

ESSAYS ON INSTITUTIONS, GOVERNANCE AND ECONOMIC GROWTH

by

Kartikeya Batra

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Committee Members:

- Professor Sebastian Galiani (Chair, University of Maryland)
- Professor Ethan Kaplan (University of Maryland)
- Professor Sam Asher (Imperial College London)
- Professor Sonalde Desai (University of Maryland)
- Professor Kenneth Leonard (University of Maryland)

DEDICATION

To those who watch us from the heavens, and to those who make this world heavenly.

Thank you, Nana ji, Nani ma, and Dadi ma.

ACKNOWLEDGMENTS

The successful conclusion of a long and challenging journey is not possible without contributions from a lot of people. I would like to start by thanking my Committee's Chair, Professor Sebastian Galiani. His encouragement and confidence in my abilities has allowed me to grow as a researcher. Over the last four years, he was always available to give feedback with utmost clarity. During this period, he created a very healthy environment for me to interact with him without fearing any kind of judgment. The journey would have been very different without him.

I am also fortunate that I have Professor Ethan Kaplan on my committee. He is among the most knowledgeable people that I know of. His knowledge bank is not limited to economics, which has allowed for several fascinating conversations. He has always treated me with respect, and has shown a keen interest in my academic development. During this journey, he ensured that our relationship was extremely healthy, and was always accessible to answer the silliest of my queries. I thank him for making my life easy over the last few years.

It is hard to imagine a mentor who is also one's friend. That is how I define the role of Professor Sam Asher in this journey. Over the last four years, he has been a constant source of support and encouragement. Apart from sharing his technical knowledge, he has paved way for me to foster collaborations with different kinds of people. He has also been there in times of stress, and has been extremely successful in reviving my spirits whenever they were down. I thank him with all sincerity.

I would also like to thank Professor Sonalde Desai for being a wonderful mentor, and an even better boss during my summer fellowship at the National Council for Applied Economic Research in New Delhi. I also extend my sincere gratitude to the research group at the University of Maryland's Agricultural and Resource Economics department for their constructive feedback in the evolution of my research. In particular, I am thankful to Professor Kenneth Leonard, Professor M.R. Sharan, and Professor Cory Smith. In addition, I am also grateful to senior faculty members in the Economics department -Professor John Shea, Professor Guido Kuersteiner, Professor Alan Drazen, and Professor Jessica Goldberg- for their constant support during this journey. Furthermore, I thank Professor Neeraj Prasad and Professor Priya Mukherjee for their valuable inputs.

I also wish to extend my gratitude to the fantastic people who manage the Economics department at the University of Maryland. I thank Ms. Vickie Fletcher, Ms. Wanda Hauser, Ms. Orpha Jewell, Ms. Heather M Nalley, Mr. Brian McLoughlin, Mr. Evan Gray Schroeder, and Mr. Mark S. Wilkerson for facilitating this journey. Their efforts ensured that a lot of obstacles were overcome with relative ease. I am grateful to my friends in the department who helped me both academically as well as cheered me up whenever required. I thank Dr. Dheeraj Chaudhry, Dr. Shanjukta Nath, Mr. Alvaro Silva, Mr. Dario Salcedo, Mr. Juan Angula Santacruz, Mr. Jorge Enrique Perilla Garcia, Dr. Daniel Kolliner, Ms. Rachel Nesbit, and Ms. Julia Brown.

Although I started my doctoral program in 2018, my development as an economic researcher began under the able guidance of Professor Rohini Pande, Dr. Charity Moore, Professor Yusuf Neggers and Eric Dodge. I am also grateful to my colleagues and managers from that period: Ms. Neha Hridaya, Ms. Vartika Singh, Ms. Tamara Mathai, and Ms. Ramita Taneja. I am also thankful to my teachers from high school, the University of Delhi and Tufts University for their immense contributions in my development as an academician. I thank Dr. Supriya Sharma, Ms. Deepa Bahadur, Ms. Tanu Singh, Ms. Maninder Oberoi, Mr. Naveen Verma, Dr. Shalini Prakash, Professor Michael Klein, Professor Lawrence Krohn, Professor Jenny Aker, Professor Carsten Kowalczyk, and Professor Patrick Schena for being such wonderful mentors.

During this journey, I have also met new people from different walks of life. I thank all of them for their kind support and contributions: Mr. Ashutosh, Mr. Yashwant Deshmukh, Miss Barkha Dutt, Miss Shiralie Chaturvedi, Mr. Vijai Trivedi, Mr. Sanket Upadhyay, Mr. Vinod Agnihotri, Mr. Sheetal P Singh, Dr. Mukesh Kumar, Mr. Prashant Mittal, Mr. Deepak Dulani, Mr. Owais Ahmed, Mr. Sidharth Tripathi, Miss Reshma Roshania, Miss Ruchi Jain, and Mr. Piyush Pratap Singh.

Finally, no such journey can ever be complete without the endless contributions of friends and family. My journey is no exception to this rule. I thank my parents (Mrs. Geeta Aneja Batra and Mr. Naval Batra), and my sister Ms. Kriti Batra Mehra for standing by me through thick and thin. I am certain that my prolonged absence took a psychological toll on them, but I was never made to realize the same. I also thank my brother-in-law, Mr. Ashish Mehra, for his support. My friends, who are no less than family, have stood by me through testing times. I thank Ms. Ria Kukreja (also my partner), Mr. Deborshi Barat, Mr. Abhinav Jain, Mr. Tanmay Wadhwa, Ms. Jayati Grover, Ms. Tanvi Aggarwal, Mr. Shakti Sehra, Mr. Rahul Dwivedi, Ms. Shikhty Sunny, Ms. Anam Elahi, Ms. Deepika Yadav, and Ms. Anisha Baghudana for their unflinching moral support. A number of people ensured that I seldom felt lonely while living in the DMV area. I thank Ms. Ridhika Batra, Ms. Avantika Prabhakar, Ms. Adyasha Mohanty, Ms. Manshi Chaudhry, Ms. Aprita Tiwari, Mr. Neerad Deshmukh, Mr. Keshav Kaul (and his family), Mr. Kenneth Pinto, Mr. Ivan Kadiam, Mr. Nomie Hamid, Mr. Srihari R., and Mr. Siddharth RS for their presence.

I am sure I am missing a lot of other names. My apologies and sincere gratitude to them as well. I thank god that I was able to embark on this journey, and complete it with reasonable success.

Contents

List of Tables

List of Figures

List of Abbreviations

1. **2SLS:** Two stage least squares
2. **BJP:** Bharatiya Janata Party
3. **ECI:** Election Commission of India
4. **EIC:** East India Company
5. **FY:** Financial Year
6. **GDP:** Gross Domestic Product
7. **GP:** Gram Panchayat
8. **GRS:** Gram Rozgar Sewak
9. **GIS:** Geographic Information System
10. **INC:** Indian National Congress
11. **INR:** Indian Rupee
12. **IV:** Instrumental Variable
13. **MGNREGA:** Mahatma Gandhi National Rural Employment Guarantee Act
14. **MIS:** Management Information System
15. **NCP:** Nationalist Congress Party
16. **NIC:** National Informatics Center
17. **NRLP:** National River Linking Project
18. **NSDP:** Net State Domestic Product
19. **OBC:** Other Backward Castes

20. **OLS:** Ordinary Least Squares
21. **PMGSY:** Pradhan Mantri Gram Sadak Yojana
22. **RDD:** Regression Discontinuity Design
23. **SC:** Scheduled Caste
24. **SHRUG:** The Socioeconomic High-resolution Rural-Urban Geographic Platform for India
25. **ST:** Scheduled Tribe
26. **TSC:** Total Sanitation Campaign
27. **UDISE+:** United District Information System for Education Plus
28. **UP:** Uttar Pradesh
29. **USD:** US Dollars

1 CHAPTER 1: Long-Run Effects of Land Redistribution: Evidence from India[1]

How does inequality in land ownership impact long-run economic development? I study the effects of a land redistribution program implemented in north India in the 1820s. I exploit intra-state variation provided by a colonial policy that led to the recognition of property rights among village communities across 72% of the state. In the remaining area, a smaller number of pre-colonial revenue farmers were recognized as landowners. Using present day village-level microdata, I estimate treatment effects by means of a spatial RD design. Results suggest that redistribution led to an overall increase in durable asset ownership, nonfarm employment and years of schooling, and that positive effects were also realized by descendants of those lower caste households that did not receive land. Using data from a field survey, I find that across reformed areas, such households are less likely to comply with restrictive socio-economic norms, which may be a channel of persistence. These findings are consistent with a dynamic political economy model, which predicts that landowners facing higher competition are less likely to resist the dilution of such norms. The results have important implications for understanding non-economic channels through which landowning patterns may impact development.

[1]I would like to express my sincere gratitude to Sebastian Galiani, Ethan Kaplan, Sam Asher, Jessica Goldberg, M.R. Sharan, Cory Smith and Allan Drazen for their invaluable insights and comments. I would also like to thank Dheeraj Chaudhary, Avantika Prabhakar, members of the Development Economics Research Group, the Political Economy Research Group and students and faculty at the AREC department at the University of Maryland, College Park for their feedback. I am also thankful to participants at the Applied Young Economist Webinar, the Fourth SANEM-World Bank North America Discussion Forum 2023, the 7th European Workshop on Political Macroeconomics, the Annual Meeting of the Economic History Association (2023), and The North East Universities Development Consortium (NEUDC) Conference 2023 for their constructive feedback. In addition, I would like to extend my gratitude to Yusuf Neggers for his continuous guidance. The fieldwork associated with this paper would not have been possible without financial support from the Allan G. Gruchy Fellowship and funding from the Department of Economics, University of Maryland (College Park).

1.1 Introduction

Historical institutions can drive persistent spatial disparities in socio-economic development (Engerman and Sokoloff, 1997; Banerjee and Iyer, 2005; Dell and Olken, 2020). Those pertaining to distribution of land ownership are especially relevant. Historically, land was an important factor of production in a global economy dominated by agriculture. However, the utility of land ownership was not limited to economic returns. For instance, apart from being an important factor of production, land ownership cemented the social elite status of landowning communities. This was true for communities across different parts of the world, such as the aristocrats in Europe and upper caste households in the Indian subcontinent. A combination of these factors allowed the landowning social elite to control the majority non-elite, while impeding the latter's socio-economic development. It is, therefore, possible that policies that affect distribution of land ownership may impact long-run development by altering traditional social hierarchies.

Traditionally, across India, land was concentrated in the hands of households belonging to the social elite, or the upper caste. Land ownership, in turn, reinforced the latter's elite status in the social hierarchy. It allowed landowning upper caste households to control the majority lower (non-upper) caste community by enforcing caste-based norms, which restricted the latter's socio-economic mobility. Intermediate caste households, for instance, were only allowed to take up jobs involving manual labor. Lowest caste households were condemned to tasks such as scavenging. Facilitated by high levels of land concentration, the enforcement of these norms may have played a role in suppressing socio-economic development across lower caste households. The latter, as per estimates, make up for over 65% of the country's population, but a majority of them continue to lag behind in terms of socio-economic indicators. The context makes it interesting to study the long-run impact of historical policies that may have altered land ownership patterns in India.

In this paper, I study the long-run effects of a 200-year old land redistribution policy implemented in north India through its impact on the functioning of the caste system. I estimate the policy's aggregate effects on indicators of long-run development. I estimate its impact on those lower caste households whose ancestors did not receive any land under the policy, but resided in areas where it was implemented. Finally, I test whether some of the long-run effects of redistribution can be attributed to changes in the functioning of the caste system. Identification of the effects of such a policy faces empirical challenges related to endogeneity. I overcome this challenge by exploiting plausibly exogenous intra-state variation provided by a colonial policy implemented in the 1820s in India's most populous state, Uttar Pradesh.

The British East India Company (EIC) entered India during the 17^{th} century. Over time, it transformed into a political entity. By the end of the 18^{th} century, the EIC had started gaining control over the north Indian state of Uttar Pradesh. By the 1820s, the EIC controlled roughly 72% of the state's total area.[2] Due to ideological changes in London, in this portion of Uttar Pradesh, the EIC abandoned the system of appointing revenue farmers for land revenue collection, which had been implemented in eastern India. Instead, the Company identified village communities with land ownership claims, and granted them modern proprietary titles. These communities were also made jointly responsible for land revenue payments.

When the rest of the state, i.e. the central region called Oudh, was acquired in the 1850s, the EIC intended to implement similar reforms. However, proprietary titles of large parcels of land were granted to a much smaller group of pre-colonial revenue farmers (*taluqdars*). The decision was taken to ensure the *taluqdars'* loyalty in the wake of a nationwide mutiny against the Company in 1857. These two disparate systems, wherein the average number of landlords per 1000 rural people in the reformed districts was thrice the number in unreformed districts, con-

[2] I exclude the districts of present day Uttarakhand (a state that was carved out of Uttar Pradesh in the year 2000), and the districts of Lalitpur, Jalaun and Jhansi, for these were governed under different provisions.

tinued until India's independence in 1947, after which policies were enacted to bring uniformity in land ownership patterns across the state. Evidence suggests that despite the colonial policy, at least until independence, in both regions, a majority of the land was owned by households belonging to the traditional social elite, i.e. the Hindu upper caste and the Muslims.[3]

Intra-state spatial variation in the concentration of land ownership provides a setup for a spatial regression discontinuity (RD) design, wherein I compare villages within a narrow bandwidth around an arbitrary border that separated the unreformed area of Oudh from the rest of the state. I merge village-level microdata from three different sources, and estimate long-run aggregate effects of land redistribution on three outcomes representative of different indicators: an index of all durable assets listed in the Population Census of 2011 (wealth), number of people in a village employed in nonfarm establishments relative to its population (structural transformation) and average years of schooling per capita (human capital). When estimating effects for households belonging to the lowest castes, I use the same variables, except one. Owing to unavailability of data, I replace the number of people in a village employed in nonfarm establishments with the number of such establishments owned by members of these castes. Overall, I compile data for over 107,000 villages across the state. In the benchmark RD specification, I use approximately 10,000 of these observations.

Village-level data on asset ownership across all households comes from the Population Census of 2011, whereas that for lowest caste households comes from the publicly unavailable Socio-Economic Caste Census of 2011-12. Indicators related to nonfarm employment are obtained from the Economic Census 2013. For both the overall population as well as lowest caste households, data on average years of schooling comes from the Socio-Economic Caste Census of 2011-12. Finally, to the best of my knowledge, there is no publicly available data related to the

[3]For instance, in the sample, 86-90% of the land is owned by these groups.

functioning of the caste system at the village level. I overcome this challenge by conducting a field survey across 189 villages.

My estimates suggest that two hundred years after the reform, areas where the policy was implemented (reformed) continue to witness better performance across all three indicators. While the average effect of redistribution on asset ownership and years of schooling is in the range of 6-7% of the control mean, that on number of nonfarm workers per 100 people in the village is around 50% (on a control mean of 2.63). Results are robust to the choice of a range of bandwidths, usage of different kernels, removal of areas around a major river, different levels of clustering, and the usage of Conley standard errors. I also find evidence of similar positive effects on the traditionally landless lowest caste households. Given that the ancestors of these households were unlikely to have received proprietary titles, these results point to positive effects through channels other than direct land ownership.

Acknowledging that these results may be driven by historical pre-treatment confounders, I compile a novel dataset with pre-treatment (16^{th} century) variables at the sub-district level. These include tax demands, presence of infantry and cavalry, charitable grants, as well as the sub-district's social composition during the pre-treatment period. These variables are balanced between the two areas, and their inclusion in the benchmark specification does not affect estimates. I further dispel concerns related to pre-treatment confounders by showing that for the majority of their recorded history (250 BC-1795 AD), both the unreformed and reformed areas within my sample were ruled by the same rulers.

Differences in concentration of land ownership may impact landless groups through several institutional mechanisms, such as changes in local politics (Anderson et al., 2015), government's policies (Dell, 2010) and delivery of public services (Pandey, 2010). Historically, across

India, land has been held by households belonging to the upper caste. Through its norms, the caste system promotes stereotypes that impede the socio-economic mobility of the traditionally landless lower caste households. In the reformed areas, land redistribution led to the creation of a bigger community of landed upper caste households with smaller landholdings per landowner. I hypothesize that this may have resulted in the reduction of influence of individual landowners, hence impacting the enforcement of caste-based norms among lower caste households. For example, historical qualitative evidence (Baden-Powell, 1892; Ahmed, 1952) points to the fact that reformed areas witnessed lower caste-based residential segregation, pointing to dilution of caste-based norms. I corroborate these findings by showing a negative causal impact of redistribution on the number of caste-based hamlets across reformed villages (as of 2018). The result hints at persistent effects of land redistribution on the implementation of a caste-based social system.

In addition, I estimate the long-run effects of redistribution on compliance with caste-based norms by using data from a primary survey conducted among 2,038 households across 189 villages in the year 2022. Results suggest that lower caste households across reformed areas are less likely to adhere to historical socio-economic stereotypes (12-29%), which prescribe lower education and low-productivity jobs for them. This is relevant, for literature suggests that aspirations may impact long-run development (Genicot and Ray, 2017). The causal chain, however, may be confounded by intermediate outcomes of redistribution that may impact compliance with caste-based norms.

For instance, redistribution may have led to higher agricultural income during the 19^{th} century, which may have impacted compliance with norms. I provide suggestive evidence that does not support this hypothesis. Using archival district-level agricultural productivity data from Donaldson (2018), I show that during 1884-1930, that average real income per acre across reformed districts was either no different or lower than that in the unreformed districts. Similarly, redistribution may have impacted colonial governance, which may have impacted compliance with norms. To test this hypothesis, I compile novel archival village-level data from colonial gazettes. I find casual evidence that within the bandwidth, the probability of placement of a school across reformed villages was no different from that in the unreformed areas. As another measure of colonial governance, I compile novel data at the level of the sub-district. I find suggestive evidence that within the bandwidth, colonial tax demands in the reformed areas were no different from those in the unreformed areas. These results provide new insights into channels of persistence of colonial land policies in India (Banerjee and Iyer, 2005; Banerjee et al., 2005).

I argue that these results can be explained by a dynamic political economy model of elite behavior formulated by Acemoglu and Robinson (2006). Given higher levels of competition among landed upper caste households and lower levels of political rents per landed upper caste household in reformed areas, such households had lower incentive to block the process of change in lower caste households' attitudes towards caste-based norms. Recent literature shows that traditional elite in democratic societies rely on better social relations to maintain persistent social control (Acemoglu et al., 2014; Anderson et al., 2015). Combining the model's predictions and the said empirical literature, a lower inclination to ensure social control among upper caste households in reformed areas should manifest in poorer social relations with lower caste households. Causal estimates from survey data suggest that upper caste households in reformed areas are less likely to be cordial towards lower caste households. For instance, across reformed areas, upper caste households are less likely to fetch water from the same source as lower caste households (53% on average). Unsurprisingly, results from an Ultimatum Game played with respondents show that lower caste households in reformed areas are less likely to trust upper caste households (15.3-16.7% on average).

The rest of the paper is organized as follows. In Section 1.2, I highlight the paper's contributions to existing literature. In Section 1.3, I lay out the context. In Section 1.4 and Section 1.5, I explain the empirical strategy and data sources respectively. In Section 1.6, I show the main results, followed by a discussion of channels of persistence in Section 1.7. Section 1.8 provides a potential explanation for these results, while Section 1.9 contains the final conclusions.

1.2 Contribution to Literature

There is a broad literature on the persistent effects of historical institutions (Acemoglu et al., 2001; Nunn, 2008; Dell, 2010; Dell, 2018; Dell and Olken, 2020). Within this broader strand, this paper contributes to literature related to land-related institutions. Engerman and Sokoloff (1997), Banerjee and Iyer (2005) and Pandey (2010) provide evidence of negative long-run effects of land concentration on overall economic development and public goods. On the other hand, Acemoglu et al. (2008) and Dell (2010) show that in weakly institutionalized settings such as parts of Latin America before 20^{th} century, inequality in land ownership led to better long-run socio-economic outcomes, for it allowed powerful landowners to negate the potential negative effects of policies implemented by other stakeholders such as the political elite. To the best of my knowledge, the context in this paper does not resemble the aforementioned weakly institutionalized settings of Latin America. For instance, the counterfactual to the unreformed group in this setting is an area with smaller landholders who received proprietary titles from the colonial authorities. The fact that I find results contrary to those found by Acemoglu et al. (2008) and Dell (2010) is, therefore, not surprising.

Another relevant strand of literature looks at the long-run effects of economic policies established by colonial authorities across different countries. While most papers find negative long-run effects of such policies (Nunn, 2008; Lowes and Montero, 2021), Dell and Olken (2020) find positive long-run effects of the sugar cultivation system established by Dutch colonizers in Indonesia. This paper provides another example of a colonial policy, which led to positive socio-economic outcomes in the long-run. It also adds to the literature by providing evidence on the impact of such policies on compliance with social and cultural norms, which may be a potential channel of persistence. Dell and Olken (2020), on the other hand, focus on the Dutch colonizers' emphasis on establishing manufacturing hubs, as well as setting up transportation infrastructure.

The relationship between land ownership and the caste system has been explored in the past. Concentration of land ownership among specific caste groups can impact quality of public services (Pandey, 2010), inter-community trade (Anderson, 2011) and local politics (Anderson et al., 2015). To the best of my knowledge, this paper is the among the first to document changes in compliance with caste-based norms as a response to reduction in land concentration. Another strand of literature closely associated with this paper deals with the long-run effects of land-related institutions set up by the British in colonial India (Banerjee and Iyer, 2005; Pandey, 2010). Specifically, exploiting inter-district variation, Banerjee and Iyer (2005) estimate the overall impact of different colonial land tenure systems on agricultural outcomes and productivity during the period 1956-1987. Using an IV strategy, they find that areas that used to be under large landlords lagged behind in agricultural outcomes. The primary contribution of this paper to that literature is the exploration of changes to the caste system as a channel of persistence. I also contribute to the literature by looking at a wider range of socio-economic outcomes that may not be as closely linked with land ownership patterns as agriculture.[4] Methodologically, I use a cleaner identification strategy, although my estimates are less externally valid than those in Banerjee and Iyer (2005). Given the dearth of intra-state variation in land tenure systems, the baseline empirical specifications in that paper can not control for regional (state or district) fixed effects. I overcome that limitation by means of a spatial regression discontinuity design,

[4]For instance, Foster and Rosenzweig (2004) suggest that nonfarm growth in India is not predicated on growth in agricultural productivity.

where I compare approximately 10,000 villages within the same state.

Compared to the literature on other land reforms, the body of literature on land redistribution is much smaller. One can assess the impact of land redistribution on two broad sets of indicators: first-order effects on variables related to the local economy (Mendola and Simtowe, 2015), and second-order effects that operate via changes in social and political institutions (Faguet et al., 2020). I contribute to the literature concerning the latter by focusing on the impact of land redistribution on long-run outcomes through its effect on social and cultural institutions. The paper also contributes to the economic literature surrounding trust and social cooperation. Nunn and Wantchekon (2011) show that extractive institutions such as slave trade led to lower levels of trust in Africa. However, I show that seemingly extractive institutions (such as higher land concentration in unreformed areas) may actually lead to higher trust between elite and non-elite if the former have adequate incentive in this direction.

Finally, an extensive literature focuses on the positive effects of trust on economic growth. Knack and Keefer (1997) use cross-country analysis to show the positive association between trust and civic cooperation, and economic performance. Algan and Cahuc (2010) find causal evidence of the positive impact of trust on economic growth. However, in the context of relations between traditional elite and non-elite, Acemoglu et al. (2014) and Anderson et al. (2015) show negative correlation between social capital and economic growth in the context of Sierre Leone and India respectively. Similar to Acemoglu et al. (2014) and Anderson et al. (2015), I find evidence of a negative association between trust and cooperation, and socio-economic outcomes.

1.3 Context

1.3.1 Colonial Land Tenure Systems and Property Rights

The British Crown started administering Indian affairs in 1858. However, the British formally entered India with the entry of the East India Company (EIC) during the 17$^{\text{th}}$ century. Although the EIC had entered India for trade and commerce, it turned into a political force during the middle of the 18$^{\text{th}}$ century with military victories in eastern parts of the country. Thereafter, using different methods (battles, treaties, etc.), the Company started acquiring large pieces of territories from local rulers. Land revenue had been the major source of income even for the rulers before the British, i.e. the Mughals. The same trend continued once the EIC annexed different parts of the country.

Upon its arrival, the Company set up different systems of land revenue collection, or land tenure systems. By the middle of the 19$^{\text{th}}$ century, they had set up three different land tenure systems across different parts of India. Under the first system (*raiyatwari*), land revenue settlements were made directly with the cultivators. Under the second (*mahalwari*), these were made with local village bodies. An important feature of these two types of settlements was that the colonial authorities assigned ownership to those with whom the settlement was made, i.e. cultivators and local village bodies. Finally, under the third system (*zamindari*), the settlement was made with large landlords, who collected revenue from farmers working on their land. In effect, even under this system, the landlord owned property rights, for they could bequeath, sell or buy their rights to collect revenue (Kumar, 1982).

As discussed in Banerjee and Iyer (2005), there is little variation in the type of land tenure system within regions/states. Archival documents suggest that even within these three categories, there was spatial variation in implementation.[5] A detailed explanation of each of these systems across India, however, is not relevant. Instead, in Section 1.3.3, I explain historical variation in concentration of land ownership within the state of Uttar Pradesh, i.e. the setting of the policy experiment that I exploit in this paper.

1.3.2 About Uttar Pradesh

India's most populous state, Uttar Pradesh, is home to over 200 million people- close to the population of countries such as Brazil and Nigeria. It lies in the northern part of the country, bordering the nation of Nepal along its northern side. As of 2018-19, it had the second lowest per capita Net State Domestic Product (NSDP) among all Indian states. Furthermore, like most big administrative regions, different regions of the state exhibit different levels of economic prosperity (Figure 1).

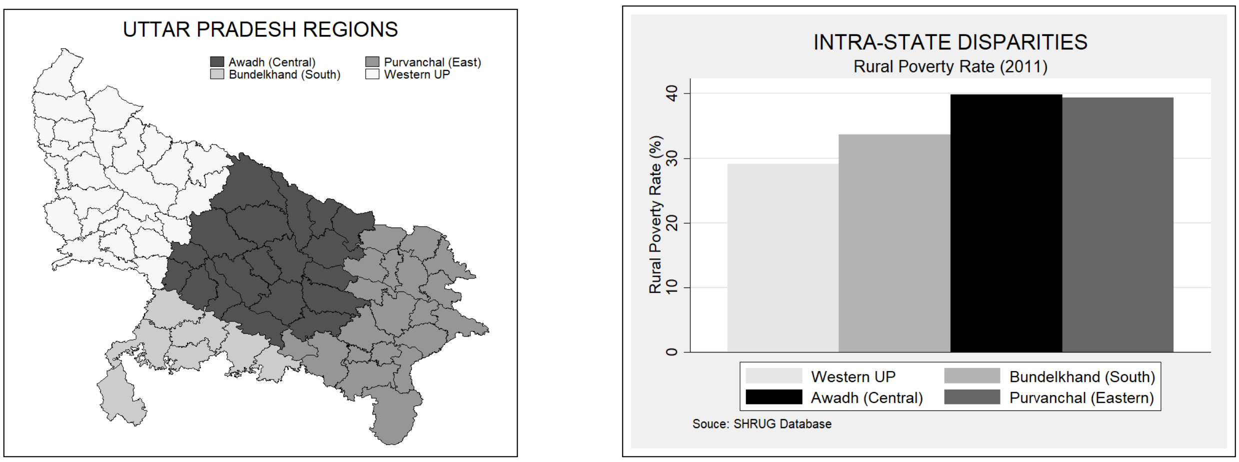


Figure 1: Regional Disparities Within Uttar Pradesh

The state's size is also responsible for its political clout in national politics. With the largest state legislature (403 members) in the country, the state sends the maximum number of national legislators- around 14% of the combined strength of the Upper and the Lower House- to

[5] For instance, as pointed by Baden-Powell (1892), despite being categorized as *mahalwari*, the system implemented in the central territories of India was very similar to the *zamindari* system.

the National Parliament. The state is also fragmented along social lines with Varshney (2019) describing it as a "20-20-20-40" state, wherein Hindu upper castes, Hindu lowest castes (Scheduled Castes/SC) and Muslims each make up for roughly 20% of the population, and Hindu intermediate castes (Other Backward Castes/OBC) make up for the remaining 40%.[6]

1.3.3 Variation in Land Ownership within Uttar Pradesh

Within Uttar Pradesh, the British followed a staggered annexation policy. In Figure 2, the controlling authority of a region is represented by its flag. Until 1801, most of the state was under local rulers, although the British flag in the south-eastern region of the state represents the territory of Benares acquired by the British in 1770s. Subsequently, large parts of the state were either ceded by local rulers or conquered by the EIC during the period 1801-1803. When the British acquired these territories, they inherited an aristocratic agrarian society, with a few politically powerful overlords (or *taluqdars*) in control of land and revenue farming. In almost all cases, these *taluqdars* were either upper caste Hindus, or Muslims.

Under the influence of growing utilitarianism in London, a new land policy was implemented in the newly acquired areas in the 1820s. The same policy was implemented in other areas of the state that were subsequently acquired until 1856 (Stokes, 1959). Under this policy, British authorities identified local communities who had been the original owners of land. These communities were then awarded "modern proprietary" titles (Stokes, 1983). They were also made jointly liable to pay land revenue to the British authorities. The arrangement, called the *Mahalwari* system, was a departure from the old system under the Mughal ruler, where revenue

[6]The caste system has determined families' social status for centuries across India. Castes are fixed at the household level and do not change across generations. Traditionally, upper caste households were the elite (warriors, priests, merchants) and the primary landowners. Intermediate caste households (OBCs) were the traditional manual laborers. Finally, the lowest castes (SCs) were considered as untouchable and were treated as outcasts. They were condemned to tasks such as scavenging and handling carcasses. Over time, even though the caste-profession linkage has weakened, the relevance of caste as a social indicator continues, especially across rural areas.

farmers were responsible for revenue collection. Like the British, the Mughals were conquerors and Baden-Powell (1892) notes that under their regime, the administration "cared nothing for landed rights". In fact, in some areas, such as the north-western part of Uttar Pradesh, all vestiges of proprietary rights of local village bodies had been "stamped out" (Baden-Powell, 1892).

Beginning 1822, the British authorities started assessing proprietary claims of local villages bodies. Subsequently, they granted land titles to village bodies whose claims were recognized. The said village bodies could be in the form of an individual/family, or a body of co-owners linked by ancestry (*Pattidari*). They could also be in the form of a body of co-owners not linked by ancestry who divided the land equitably amongst themselves (*Bhaichara*). In Figure 2, the *mahalwari* system is represented by circles. Notably, even after the process of redistribution and titling, land was predominantly owned by upper caste (traditional social elite) households (Appendix A.1). These areas under the newly formed system will henceforth be referred to as the reformed areas.

As shown in Figure 2, the central part of the state (Oudh) was not annexed until 1856. Oudh's annexation was succeeded by the famous Indian Revolt of 1857, which was a consequence of several factors. Once the revolt was brought under control, in March 1858, the colonial authorities decided to first confiscate and then reconfer land titles to Oudh's *taluqdars* as proprietors (Baden-Powell, 1892; Chanana, 2013). This was done to ensure their loyalty (Banerjee and Iyer, 2005). In the absence of the revolt, the British would have implemented the *mahalwari* system in Oudh (Pandey, 2010). Instead, they granted proprietary rights of a majority of the land to the *taluqdars*. In the economics literature, the process of granting proprietary titles to the *taluqdars* in these regions has, therefore, been considered as exogenous (Banerjee and Iyer, 2005; Pandey, 2010). These areas will henceforth be referred to as the unreformed areas.

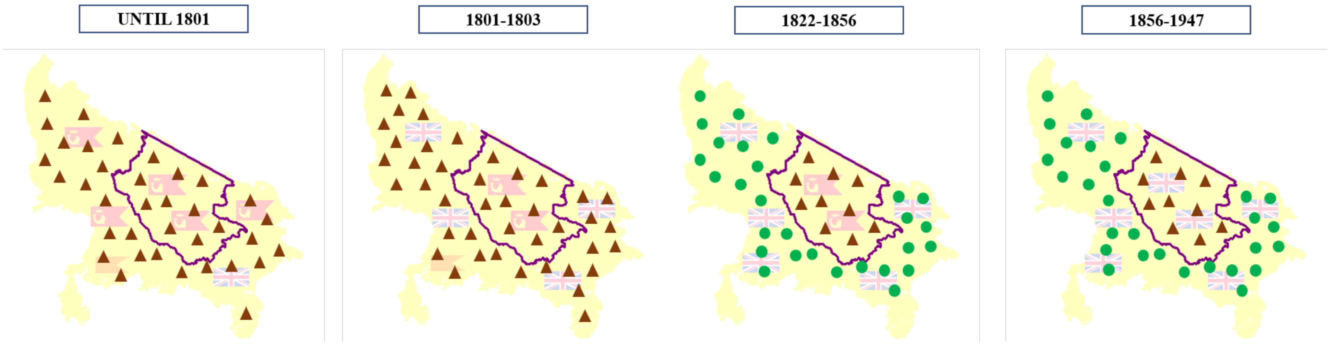


Figure 2: Land Reform Process (Triangles= Taluqdari; Circles= Mahalwari)

Across the unreformed areas, *taluqdars* owned 60% of all villages. In these villages, the *taluqdars* acted as "lords of the manor", while also farming revenue for the British authorities (Baden-Powell, 1892).[7] These villages were further divided into two kinds. In villages within the first category, "tenants or...under-proprietors with very weak rights" cultivated the land (Baden-Powell, 1892). In villages within the second category, some autonomy was given to village proprietary bodies, which functioned under the auspices of the *taluqdar*. In the remaining 40% villages, land revenue settlements were made with independent village proprietors. In Figure 3, I use modern day district boundaries (an approximation) to show the percentage of villages directly held by *taluqdars* across different districts. As of the first decade of the 20^{th} century, the average number of *taluqdars* across each unreformed district was 33 and varied in the range of 20-62 individuals/families.

Given the usurpation of old proprietary rights by Mughal administrators, the colonial policy involved the expropriation of land from big landlords (pre-colonial taluqdars) and distribution

[7]The presence of taluqdars was associated, to varying degrees, with subservience of the local cultivators. For instance, in the northern parts of the reformed region, Taluqdars played a major role in turning wastelands into rural ecosystems and hence, acquired subordination of the cultivators. Generally speaking, villages under Taluqdars, over time, witnessed obliteration of their original structure and turned into mere "aggregates of cultivating holdings" (Baden-Powell, 1892). Unsurprisingly, Taluqdars were often found acting as de facto power centers. For instance, if the Taluqdar happened to be from the family of a traditional ruler of the area, they would participate in processes such as administering justice and deciding disputes (Baden-Powell, 1892).

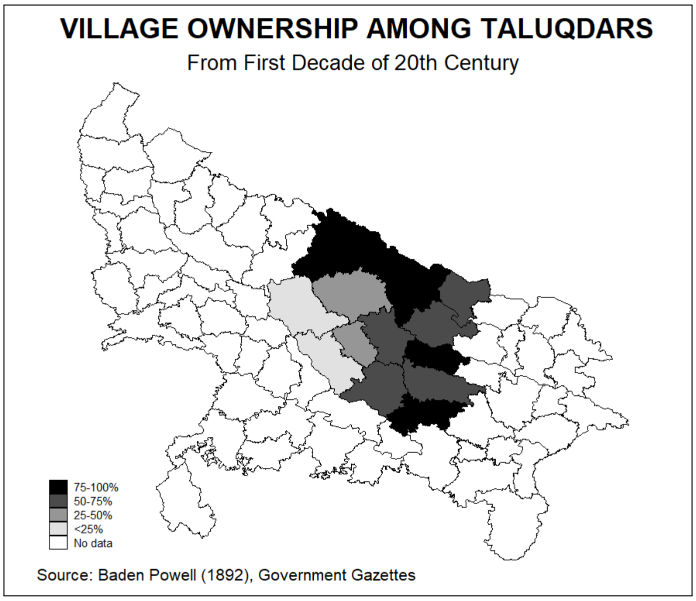


Figure 3: Percentage of Villages Owned by Taluqdars of Oudh (using modern district boundaries)

among smaller village bodies across reformed areas. These bodies' ownership claims were recognized through modern proprietary titles. Across unreformed areas, such titles were awarded to pre-colonial taluqdars. At the same time, the reform process also led to two different land tenure systems. In the *mahalwari* (reformed) areas, the colonial authorities made revenue settlements with local village bodies, whereas in the *taluqdari* (unreformed) areas, this was done with large landlords (*taluqdars*). In summary then, the impact of reforms (or treatment) is a combination of a different land tenure system as well as smaller landholdings in the *mahalwari* areas. At the time of independence, the recorded number of landlords per 1000 rural people across the reformed districts was 49, i.e. almost three times the number in the unreformed areas (17), suggesting the creation of a larger landowning elite group with lower land ownership per landed elite in the reformed areas. After India's independence in 1947, the state government enacted laws to reduce inequality in land ownership, which were applied uniformly across the state.

1.4 Empirical Strategy

A naive OLS regression to estimate effects of land reforms on village v is shown in Equation (1). Outcomes Y in village v will be regressed on a binary variable $ReformedArea_v$, which equals

1 if the village lies in the reformed area (and 0 otherwise), while controlling for village-level covariates X_v. In this case, β is the coefficient of interest. However, estimates of β may be biased owing to unobservable confounders, such as differential treatment by pre-colonial rulers.

$$Y_v = \alpha + \beta ReformedArea_v + X'_{v,z,s}\lambda + \epsilon_v \tag{1}$$

In order to estimate treatment effects of land redistribution, I therefore, exploit the discontinuity on either side of the border that separates reformed and unreformed areas. For this exercise, I compare villages in the unreformed districts with those in neighboring reformed districts, most of which were ceded by the Mughal ruler in 1801, as per the Treaty of Oudh (Figure 4).[8] The Treaty of Oudh (1801) led to the transfer of the highlighted districts into the hands of the British. The cession was based on fiduciary negotiations between the British authorities and the Mughal ruler of the kingdom, who used to contribute towards the maintenance of British troops. As Mahajan (2020) points out, after 1798, the British raised their monetary demands with the intent of gaining territories. The objective was to burden the ruler, and then forego demands in return for territories. Thus, in 1801, a settlement was reached wherein areas whose gross revenue was equal to the amount foregone were handed over to the British. As a result, the ruler gave up territories along the periphery of his kingdom (which were to become a part of the reformed areas), while retaining the central parts of his kingdom (the unreformed areas of Oudh). The latter were annexed by the British only in 1856 and are referred to as the unreformed areas, for unlike in the reformed areas, proprietary titles in these areas were granted to pre-colonial taluqdars. Therefore, to the best of my knowledge, existing historical evidence suggests that apart from the fact that the unreformed region was home to the Mughal ruler's capital (Lucknow), the placement of the boundary was not influenced by any observable

[8]Some of these districts, as highlighted in a lighter shade of gray, had been ceded back in 1775.

factors.[9]

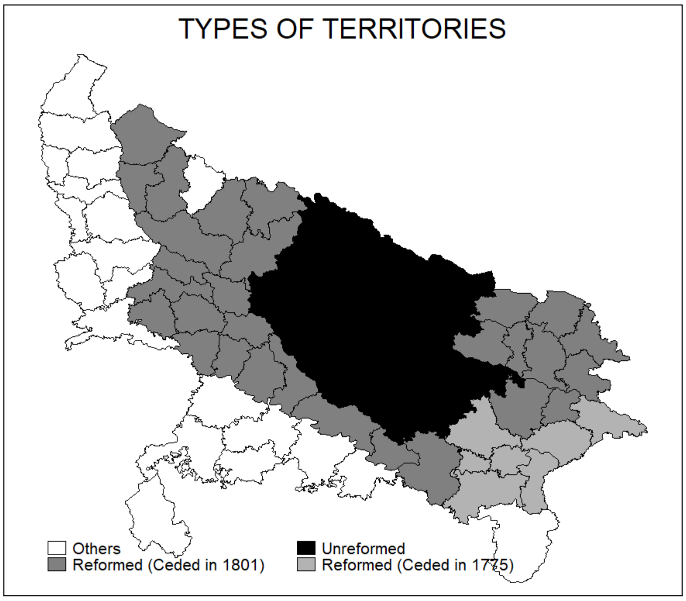


Figure 4: Reformed (Ceded) and Unreformed Districts

Following Dell (2010), dividing the border into 25 segments of equal length, I estimate β using a spatial regression discontinuity (RD) design. As shown in Appendix A.2, in my benchmark specification, I compare villages within a narrow bandwidth of 10 kilometers (6.25 miles) on either side of the border. Like Dell (2010), a village is assigned to that segment which is closest to its centroid. Equation (2) provides the main specification. Outcome $y_{v,z,s}$ for village v in agro-climatic zone z along segment s is regressed on a local linear polynomial in distance to border $f(Location_{v,z,s})$, allowing for different slopes on either side of the border (Cattaneo et al., 2019). Binary variable $ReformedArea_{v,z,s}$ equals 1 if the village lies in the reformed area, and 0 if it lies in the unreformed area, which makes β the main coefficient of interest. In the main specification, I use three outcomes of interest, which are broadly representative of a village's socio-economic well-being. To estimate effects on household wealth, I construct an index of durable asset ownership at the village level, combining all assets for which data is available in the Population Census of 2011. To estimate the impact on the nature of economic activity and

[9] I provide additional justifications regarding the exogeneity of the boundary's placement in Section 1.6.1.

structural transformation in a village, I use nonfarm employment per 100 people as the outcome variable. Finally, I estimate the impact on human capital using average years of schooling per individual.

$$y_{v,z,s} = \alpha + \beta ReformedArea_{v,z,s} + f(Location_{v,z,s}) + \gamma \sum_{i=1}^{4} DistHistCities^{i}_{v,z,s} + \eta_s + \theta_z + X'_{v,z,s}\Psi + \epsilon_{v,z,s} \tag{2}$$

Following Dell et al. (2018), who explicitly control for distance of villages to the region's largest urban area in the Vietnamese context, $DistImpCity^{i}_{v,z,s}$ is the distance of the village centroid to four historically important cities in the state (indexed i = 1 to 4). One of these cities (Lucknow) is the current capital city of the state (and also a former capital), while two are former capitals (Agra and Allahabad). A fourth city, Kanpur, is a historically important commercial hub. $X_{v,z,s}$ is a vector of rich village-level controls, including historical climatic conditions (1970-2000), soil quality (2015), village v's elevation and its distance to the nearest river. These controls are included to enhance the precision of the point estimates. η_s and θ_z denote segment and agro-climatic zone fixed effects respectively. Segment and agro-climatic zone fixed effects are required for identification as well as precision. Given its shape, a uniform bandwidth on either side of the border implies an uneven number of villages on either side in one of the three agro-climatic zones (and the segments within it) in the sample. The presence of an international border with Nepal along the state's northern border is the primary reason for the same. Agro-climatic zone and segment fixed effects help control for this imbalance. $\epsilon_{v,z,s}$ is the error term. In accordance with common practice, I use triangular kernels in the main specification.

Abadie et al. (2017) suggest that standard errors should be clustered at the same level as the unit of randomization. In this case, this requires a deeper understanding of the nature of the decision to hand over territories to the British (which later became the reformed areas). Specifically, if territories were handed over in the form of entire administrative divisions, then standard errors should be clustered at those levels. If, on the other hand, individual villages were identified for the purpose of transfer, clustering may not be required. Figure 5 shows a part of the Treaty of Oudh of 1801 (under which districts around Oudh's boundary were handed over to the British authorities), listing the territories transferred to the British authorities. As is visible, there was no uniform pattern among the units that were used to transfer territories. For instance, while in some cases entire districts (*Chucklas*) were transferred, in others, individual locations and areas around them are mentioned. Accordingly, while in the main specification, standard errors are clustered at the district-segment level, in Section 1.6.3, I discuss that my results are robust to the choice of different clustering methods. Finally, while I use a bandwidth of 10 kilometers in the main specification, in Section 1.6.3, I discuss that results are robust to the usage of a variety of bandwidths.

1.5 Data

1.5.1 Outcomes

Village-level data on asset ownership comes from India's Population Census of 2011. The Census lists the percentage of households in each village owning the following assets: radio, bicycle, car, scooter/motorcycle, mobile phone, fixed line phone and computer. Novel data from the Socio-Economic Caste Census 2011-12 provides similar data on a slightly different (and smaller) basket of assets by caste group. Although this first of its kind census was conducted in 2011-12,

Article 1.

His Excellency the Nabob Vizier hereby cedes to the Honourable the East India Company, in perpetual sovereignty, the undermentioned portions of his territorial possessions, amounting in the gross revenue to one crore and thirty-five lakhs of rupees, including expenses of collections, in commutation of the subsidy, of the expenses attendant on the additional troops, and of the Benares and Furruckabad pensions.

Statement of the Jumma.

Chuckla Korah, Kurrah, and Chuckla Etawah					55,48,577	11	9
Kehr and others					5,33,374	0	6
Furrackabad and others					4,50,001	0	0
Khairaghur and others					2,10,001	0	0
Azimghur and others, Azimghur, Mownaut Bunjun					6,95,624	7	6
Goruckpore and others and Butwul	Goruckpore, etc.	5,09,853	8	0			
	Butwal	40,001	0	0			
					5,49,854	8	0
Soubah of Allahabad and others					9,34,903	1	3
Chuckla Bareilly, Asophabad, and Kelpoory					43,13,457	11	3
Nabob Gunge, Kehly, and others					1,19,242	12	0
Mohoul and others, with the exception of the Talook of Arwul					1,48,378	4	0
Total Jumma, Lucknow Sa. Rs.					1,35,23,474	8	3

Figure 5: A Snapshot from the Treaty of Oudh 1801

granular village-level data has not yet been made available for public consumption. In the main specification, I estimate treatment effects on an aggregated asset index using assets from the Population Census of 2011. This index is a weighted average of the percentage of households owning each of the said assets at the village level, with the average price of the asset at the intra-state regional level recorded from the National Sample Survey 2011-12. In Appendix A.3, I explain the process of constructing this index. The index can be treated as a proxy for average household wealth in a village.

Data on economic establishments, their sector, total employment, and the caste of owners in each village comes from the Economic Census of 2013. I calculate the number of people employed in the nonfarm sector per 100 people in a village by dividing the total number of people employed in nonfarm establishments operating in the village by that village's population (and multiplying by 100). This variable is an indicator of structural transformation within a village. When estimating treatment effects by caste groups, due to unavailability of data on the caste of workers, I use the number of establishments owned by members of different caste groups as an outcome. Finally, I obtain average years of schooling at the village level (both aggregated and by caste group) from the aforementioned Socio-Economic Caste Census of 2011-12. This variable represents human capital per individual residing in a village.

1.5.2 Geographical Variables

Modern-day GIS data at the village level (shape file) comes from Meiyappan et al. (2018), while historical maps come from Baden-Powell (1892, 1894). Data for covariates comes from different sources. Fick and Hijmans (2017) have prepared a high-resolution (1 km^2) dataset with information on historical bioclimatic conditions (1970-2000), which I match into the village polygons. Similarly, high-resolution soil quality data (1 km^2 resolution) is available from the International Soil Reference and Information Centre (2015). Data on distance to the nearest river comes from shape files provided by the NRLP (National River Linking Project).

1.5.3 Archival Data: Social and Economic Indicators

In order to confirm that my main effects are not being driven by historical (pre-treatment) confounders, I digitize data from a popular book written during the 16^{th} century, the $Ain - i - Akbari$. This book contains data regarding pre-treatment tax demands and other variables such as population's social composition (caste and religion), number of infantry, cavalry, and amount of charitable grants at the sub-district level (see Appendix A.4). The digitization process, therefore, involved manually mapping modern day sub-districts to colonial sub-districts, and further mapping those to sub-districts during the pre-colonial (Mughal) period in the 16^{th} century (see Appendix A.5). I compiled this dataset for a total of 94 sub-districts. Within the bandwidth, I end up obtaining data for 59 out of the 70 sub-districts.

Finally, for validity checks and the exploration of potential mechanisms, I use archival data at the district level from different sources: average agricultural income per acre (Donaldson, 2018), inter-district migration (Population Census 1881) and literacy levels (Population Census 1881). I also compile novel datasets on colonial tax demands (sub-district level) and presence of schools (village-level) from colonial gazettes published during the early 20^{th} century. [10]

1.5.4 Primary Data and Field Survey

Due to absence of secondary data, the exploration of mechanisms related to compliance with caste-based norms within the bandwidth is challenging. As a result, during the Summer of 2022, I conducted a field survey. For the same, I implemented a three-stage sampling design, which involved randomly selecting 200 villages within the bandwidth, followed by random selection of hamlets in villages with multiple hamlets. In each village, I randomly chose 12 households. The three-step sampling process is explained below.

1. **Choice of Villages:** Given that I estimate treatment effects within a bandwidth of 10 kilometers along a boundary, which I divide into 25 segments of equal length, I created four bins around each segment based on combinations of treatment status (reformed and unreformed) and two equidistant bins of 5 kilometers each (0-5 kilometers and 5-10 kilometers). As shown in Figure 6, around each segment, I randomly selected 2 villages in each of the following bins: reformed side within 0-5 kilometers, reformed side within 5-10 kilometers, unreformed side within 0-5 kilometers and unreformed side within 5-10 kilometers. My target sample, therefore, consisted of 200 villages (25 segments and 8 villages per segment). Owing to logistical constraints, however, I was able to reach 189 of these

[10] For further details on all secondary data sources, please refer to Appendix A.6.

villages. In Appendix A.7, I show that these villages are broadly representative of all villages within the bandwidth with regard to bioclimatic, soil and socio-economic variables.

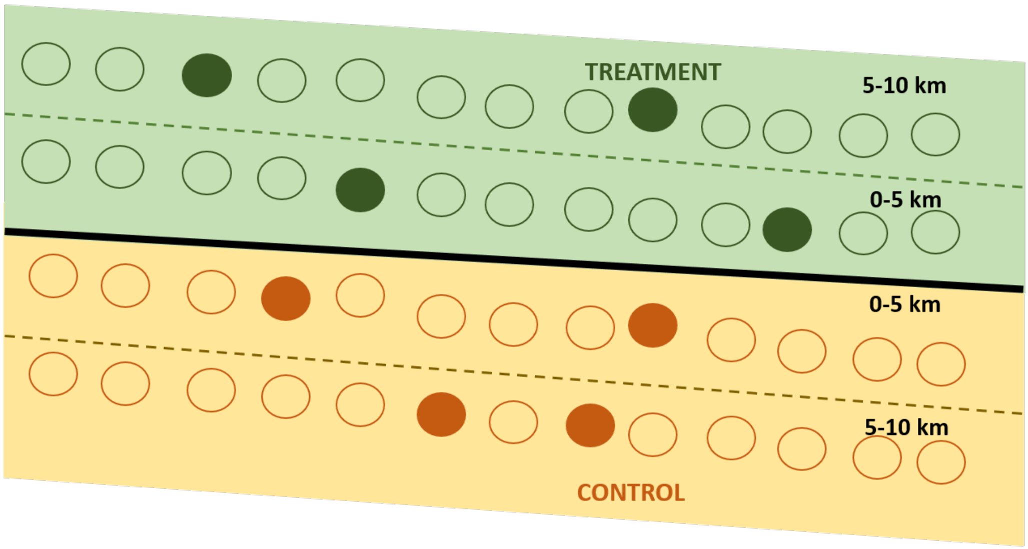


Figure 6: Sampling Design: Snapshot of One Segment

2. **Choice of Hamlets:** Indian villages are segregated into smaller habitations, or hamlets, which are often based on the caste of the local population. Approximately 44% of the 189 villages in my final sample consist of only one habitation. Hence, this second step did not apply to these villages. The remaining villages contain multiple habitations. I obtained population data along caste lines (lowest caste/SC and the rest) from the Ministry of Water Resources. Within each village, I, therefore, divided these hamlets into two categories: SC dominated (>=50% SC population) and non-SC dominated (<50% population). This is relevant, for households belonging to different castes may behave differently based on the caste composition of their habitation. After this stratification, I randomly chose one habitation per village in each category: one SC-dominated habitation and one non-SC dominated habitation per village. In villages where all habitations belong to only one of these categories, I randomly selected one habitation.

3. **Choice of Households:** The requirement from each village was a random sample of 12

households. I divided these households into two categories: 6 households from the SC category and 6 from the non-SC category. In the absence of census data that would allow random sampling at the household level, I followed the standard "right-hand" rule. Enumerators entered a habitation and counted the number of households. If the requirement from that habitation was a random sample of 3 households, they divided the number of houses by 3 and got the quotient x. They then started from the beginning of the habitation and sampled every x^{th} household to their right.[11] Each surveyor only interviewed the head of the sampled household. Detailed protocols were provided to the surveyors in order to deal with issues of absence of household members and/or closure of the house. Following these protocols, the survey team was able to survey 2,038 households, i.e. 85% of the target sample (2,400 households) across 189 villages.

1.6 Results

1.6.1 Validity of RD

Balance Checks

One of the two standard assumptions in regression discontinuity designs is that apart from treatment status, no other independent variable, which may impact outcomes of interest, changes discontinuously across the border. In order to test the same, using Equation (2), I estimate the coefficient on the treatment variable ($ReformedArea_{v,z,s}$) for a variety of geographical and bioclimatic variables, which are used as covariates in the main specification. While running the balance check for a covariate, I control for distance to historically important cities, agro-climatic zone fixed effects and segment fixed effects. Results from this exercise are shown in Appendix A.8.

[11] As an example, in a habitation with 20 houses, if the requirement was a random sample of 3 households, the surveyor divided 20 by 3 and got 6 as the quotient. They then surveyed the 6^{th}, 12^{th} and 18^{th} household on their right side.

Of the 40 covariates tested, the coefficient on the treatment variable is significant at the 10% level for one covariate (precipitation of the driest quarter) and at the 5% level for one covariate (precipitation seasonality). These results are not surprising, for even if the true coefficient on the treatment variable equals zero, one expects estimates for 4 out 40 variables to be statistically significant at the 10% level by chance. Similarly, even if the true coefficient on the treatment variable equals zero, 5% of 40, i.e. estimates of treatment effect for two out of the 40 variables may be statistically significant at the 5% level by chance. The magnitudes of the point estimates as a proportion of the control means are as follows:, i.e. -0.8% and 0.1% for precipitation of the driest quarter and precipitation seasonality respectively.[12] In the benchmark specification, I control for each of the 40 covariates.

Despite apparent balance on observable variables between the two areas, there may be other unobservable differences that may have come about as a result of pre-treatment institutional differences such as different rulers at different points of time. If that is the case, estimates of treatment effects may be picking up the impact of these pre-treatment differences. I am able to partly dispel these concerns by looking at the political evolution of the area in my sample. Joppen (1914) documents the evolution of different national and regional kingdoms and empires across India over time. Among all the pre-19$^{\text{th}}$ century snapshots in time that Joppen (1914) records (fourteen snapshots during the period 250 BC-1795 AD), only one raises concern. As shown in Figure 7, the southern boundary of the Kingdom of Jaunpur in 1398 (pink shade) overlaps with the southern boundary of the unreformed area. Keeping the same in mind, in Section 1.6.3, I conduct robustness checks to show that even if one removes observations along the southern boundary from the sample, estimates of treatment effects do not change.

[12] A simple binomial test suggests that the likelihood of one out of 40 results being statistically significant at the 5% level by chance is 87%. Similarly, the likelihood of two out of 40 results being statistically significant at the 10% level by chance is 92%

Furthermore, I compile pre-treatment data on tax demands, number of infantry and cavalry, demand for charitable grants, and regional caste composition at the sub-district level from the 16^{th} century. In Appendix A.9, I show that these variables are balanced at the sub-district level. In Section 1.6.3, I show that inclusion of these covariates does not affect estimates.

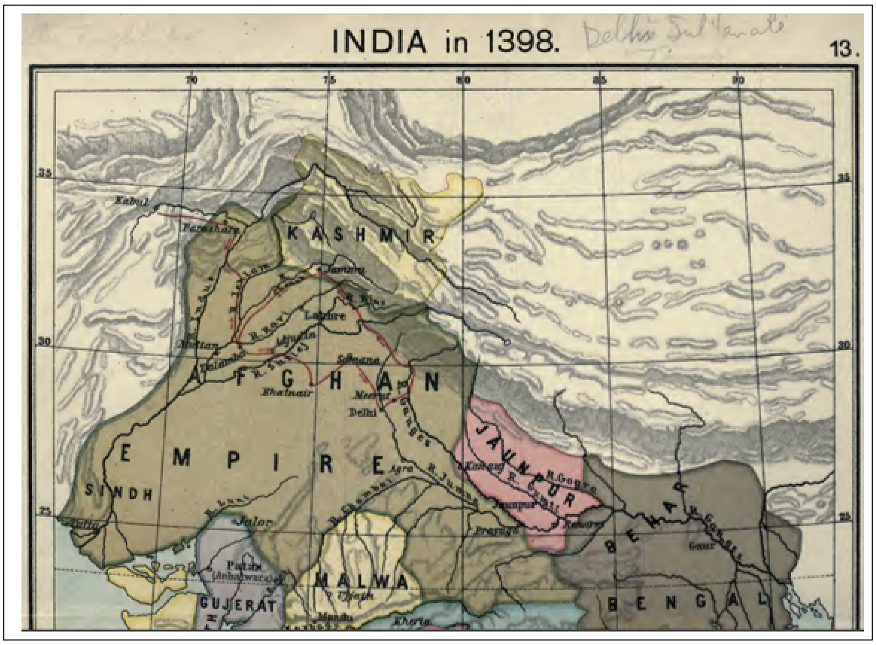


Figure 7: Kingdom of Jaunpur (1398)

Sorting Around Boundary

The other threat to identification may potentially come from sorting around the boundary. In this context, that will imply that in anticipation of the reforms, people may have sorted between the two areas. The reforms took place during the 1820s. However, the first official national census was conducted only in 1881. The Population Census of 1881 records the number of people in each district by the district of their birth. Using this information, in Appendix A.10, I show that as of 1881, average inter-regional migration across the border of Oudh for the districts within the bandwidth is statistically indistinguishable (p-value of t-test equals 0.76) for the two groups.

Even in the field survey, households were asked about their family's migration history. On average, overall migrant population makes up for only 1.2% of the population across sampled villages in reformed region, and 1.6% in the unreformed region. These are statistically indistinguishable.

1.6.2 Main Results

In Table 1, I show estimates of treatment effects obtained from the benchmark specification. These estimates suggest that colonial land reforms have had persistent positive effects across reformed areas. All three estimates are statistically significant at conventional levels, i.e. unadjusted p-values suggest that estimates are significant at the 1% level for asset ownership and nonfarm workers per 100 people, and at the 10% for average years of schooling. The Bonferroni-adjusted p-values suggest that while point estimates for asset ownership and nonfarm employment are significant at the 5% and 1% level respectively, the estimate for average years of schooling turns insignificant (with a p-value of 0.225). Given the conservative nature of the Bonferroni correction method, I also show the sharpened False Discovery Rate (FDR) q-values as described in Anderson (2008). As shown in Table 1, the sharpened FDR q-values for the estimates suggest statistical significance at the 5% level for asset ownership and average years of schooling, and at the 1% for nonfarm workers per 100 people.[13]

[13] In Appendix A.11, I show that estimates of treatment effects without controls are similar to the benchmark estimates shown in Table 1. However, they are estimated with lower precision.

Table 1: Village Level Socio-Economic Indicators (2011)

	(1) Average Asset Ownership Index	(2) Nonfarm Workers Per 100 People	(3) Log Avg Years of Schooling
Reformed Area=1	0.307***	1.333***	0.060*
	(0.116)	(0.238)	(0.035)
Sharpened q-value	[0.011]**	[0.001]***	[0.031]**
Bonferroni corrected p-value	[0.030]**	[0.000]***	[0.264]
Control Mean	4.94	2.63	4.37
Bandwidth	10 km	10 km	10 km
N	10412	10364	9862

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

On a control mean of 4.94, the treatment effect on the index of asset ownership is 0.307 (around 6.2% of control mean). To put things in perspective, a village will have a value of 100 on this index if all its resident households own at least one of all of the following: television, radio, bicycle, motorcycle/scooter, computer, car, fixed line phone and mobile phone. On average, across villages in the unreformed areas, the average distribution of asset ownership along the extensive margin at the household level is as follows: 17.8% own a television, 30.8% own a radio, 15.5% own a scooter/motorcycle, 82% own a bicycle, 2.3% own a car, 62.3% own only a mobile phone, 2.3% own only a fixed line phone, 1.5% own both types of phones, and 5.7% own a computer (which makes the average value of the index in such villages equal 4.94). As an example, keeping all else constant, a treatment effect of 0.31 on the index implies that owing to colonial land reforms, when compared to the average unreformed village, in the average reformed village, the fraction of households with at least one television, computer and mobile phone (only) is higher by 100%, 25% and 20.3% respectively.

The treatment effect on the number of nonfarm workers per 100 people is 1.33 over a control mean of 2.63, i.e. 50% of the control mean. Finally, average years of schooling in reformed areas are 6.3% higher than in the unreformed areas. To put this in context, as per the Barro-Lee dataset from 2010, education attainment across the unreformed areas is close to countries such as Rwanda (4.36 years). Owing to colonial land reforms, this increases to roughly the level of Côte d'Ivoire (4.65 years).

1.6.3 Robustness Checks

Different Bandwidths

The above results are robust to the choice of different bandwidths. In Appendix A.12.1, I estimate the benchmark specification using 16 different bandwidths (5-25 kilometers in intervals of 1 kilometer). Across bandwidths, sign of the estimate remains the same, statistical significance is maintained and 95% confidence intervals remain statistically indistinguishable from the benchmark estimate.

Donut Hole RD Estimates

It is possible that observations close to the border may be driving results. This may be especially relevant because I use triangular kernel weights in the benchmark specification. While this may not be a concern in itself, it may be problematic for the empirical strategy's internal validity if the treatment status of villages close to the border is wrongly assigned due to measurement error. Villages on either side that are very close to the border and hence, each other, may also be impacted by spillover effects. Therefore, I show estimates using the donut hole approach. For all three outcomes, I remove observations close to the border, i.e. within a distance of 2-4 kilometers on either side, in a sequential manner. In Table 24, Table 25 and

Table 26 in Appendix A.12.2, I show that all results obtained after eliminating observations from the border have the same sign as the benchmark estimate. As shown in Figure 40 and Figure 42 (Appendix A.12.2), the magnitude of point estimates for asset ownership and average years of schooling remain stable and within the benchmark confidence interval. As shown in Figure 41, however, the magnitude of point estimates for nonfarm employment, shrinks in size from 50% (benchmark) to 16% (donut hole of 4 kilometers) of the control mean.

Inclusion of Pre-Treatment Covariates

As discussed above, it is possible that pre-treatment differences in administration may be driving treatment effects. It is also possible that benchmark estimates may be driven by pre-treatment differences in the social composition of the reformed and unreformed areas. However, any analysis that can account for such factors faces data constraints. Colonial authorities maintained written records of different areas through various documents such as district gazettes, land settlement reports and official censuses. A lot of these records are accessible in archives, and many of them have been digitized. However, availability of such records from the pre-colonial period is limited.

A rare source from the pre-colonial era often cited in colonial documents is the *Ain-i-Akbari*, a book written in the 16^{th} century by an official of the Mughal emperor Akbar's court. As shown in Appendix A.4,this book contains records on land revenue demanded by the Mughal administration, presence of important social groups, number of infantry and cavalry, and charitable grants (Suyarghul) at the sub-district level. I start by mapping modern day sub-districts within the bandwidth to colonial and pre-colonial sub-districts (Appendix A.5). I then digitize data for the pre-colonial sub-districts mapped to modern day sub-districts within the bandwidth. At the end of this exercise, I have pre-colonial data for 59 of the 70 modern sub-districts within the bandwidth. I then assign the values compiled for a historical sub-district to all modern day villages mapped to it. I re-estimate the benchmark specification with and without these controls. In Appendix A.12.3, I show that estimates are robust to the inclusion of these covariates.

Other Historical Confounders

The nature of the southern part of the border separating the reformed and the unreformed regions may pose a threat to identification for two distinct reasons. One of them, as discussed in Section 1.6.1, is the overlap between the southern boundary of the region and the Kingdom of Jaunpur in 1398. Second, a geographically and culturally important river, the Ganga, flows along the southern boundary. This may be a potential confounder, for rivers may restrict physical mobility. I, therefore, re-estimate results after omitting this region. As shown in Appendix A.12.4, all point estimates obtained after the removal of villages along the southern boundary are in the same direction and within the confidence intervals of the benchmark estimates. The only statistically insignificant result, i.e. effect on average years of schooling is positive with a p-value of 0.28. A potential reason behind this insignificance could be the reduction in sample size by as much as 25%.

Across reformed areas, land revenue was revised periodically. However, one of the districts in this region (Jaunpur) had been annexed in the 1770s and a permanent settlement, i.e. fixed revenue settlement had been put in place. It is possible that benchmark results may be driven by this area. I, therefore, re-estimate the benchmark specification after excluding the said region. As shown in Appendix A.12.5, results are robust to the exclusion of these observations. The sign and size of the point estimates remain stable. For asset ownership and nonfarm employment per 100 people, estimates remain statistically significant at the 5% and 1% level respectively. The estimate for treatment effect on average years of schooling turns statistically insignificant with a p-value of 0.115.

Finally, as mentioned above, historical maps obtained from Joppen (1914) suggest that areas within a bandwidth of 10 kilometers on either side of the border seem to have been governed by the same rulers across all time periods, except during the 14^{th} century. However, for a few decades during the 18^{th} century, reformed areas along the western part of the border were held by a different Muslim community called the Rohillas. The Rohillas had been awarded the region by Aurangzeb, the last Mughal ruler with pan-India influence. Hence, in the maps presented by Joppen (1914), they are shown no differently from other remnants of the Mughal empire. It is possible that treatment effects are be driven by observations within these areas (called Rohilkhand). As shown in Appendix A.12.6, estimates are robust to the exclusion of observations along that part of the border.

Different Kernels

In the main specification, I use triangular kernels. In Appendix A.12.7, I show that estimates obtained using uniform and Epanechnikov kernels are statistically indistinguishable from benchmark estimates.

Different Levels of Clustering

As mentioned above, Abadie et al. (2017) suggest that standard errors should be clustered at the level of randomization. However, given the evidence in Figure 5, it is difficult to identify the conceptually accurate level of randomization. In the benchmark specification, I cluster standard errors at the level of district-segment. In Appendix A.12.8, I show that for two of the three outcomes, i.e. asset ownership and nonfarm employment per 100 people, estimates remain statistically significant at conventional levels when standard errors are clustered at the following levels: village (no clustering), segment, district, sub-district and development block (which is similar to a sub-district). However, for log of average years of schooling, estimates become insignificant when standard errors are clustered at the segment (p-value=0.13) and district level (p-value=0.16).

Conley Standard Errors

Although I cluster my standard errors at the district-segment level and show robustness to different levels of clustering in the previous sub-section, recent literature related to spatial analysis has taken additional note of potential autocorrelation across clusters (Conley, 2010; Kelly, 2019). In this section, I show the robustness of my results to the usage of Conley standard errors that correct for this correlation. In the absence of specific algorithms to decide the cut-off for spatial correlation, I use a range of 20 kilometers. Also, given the limited progress in this subject area, I am unable to estimate coefficients with Conley standard errors using triangular kernel weights. For this comparison, therefore, I compare benchmark estimates using rectangular kernels and estimates with Conley standard errors without any weights. As shown in Appendix A.12.9, results are robust.

1.6.4 Treatment Effects by Caste Groups

The colonial policy in reformed areas led to the recognition of formal proprietary rights for smaller village communities. A majority of these landowners, however, belonged to upper castes (Appendix A.1). Existing literature shows that ownership of land has the potential to significantly impact a household's socio-economic outcomes (Finan et al., 2005; Keswell and Carter, 2014). At the same time, the creation of a larger landowning class could have impacted lower (non-upper) caste households, most of whom did not receive land, but resided in reformed areas. Archival data suggests that households belonging to the lowest castes remained landless even in the reformed areas. Members of intermediate caste groups were primarily landless, although certain sub-groups held land in some of the reformed areas. In this section, I report estimates of the long-run effects of land redistribution on the descendants of households belonging to the lowest caste group. I use the benchmark empirical specification, comparing outcomes for lowest caste households across villages on either side of the border within the same bandwidth.

Results reported in Table 2 point to positive long-run effects of redistribution on these households. As shown in Column (1), although imprecisely estimated (unadjusted p-value=0.17), the effect on durable asset ownership is 11.7% of the control mean.[14] The effect on average years of schooling is 6%, which is almost the same as the aggregate effect. Finally, the Economic Census does not report the caste of employees across different establishments, but provides the caste of their owners. Hence, in Column (2), I estimate the effect of redistribution on ownership of nonfarm establishments across lowest caste households. The treatment effect is positive, statistically significant and roughly 86% of the control mean. Unfortunately, due to data constraints, I cannot estimate treatment effects on asset ownership and average years of schooling for intermediate caste groups. However, the Economic Census provides disaggregated data for these castes as well. Although not reported in Table 2, estimate of treatment effect on ownership of nonfarm establishments is positive and statistically significant even for this group of households. These estimates along caste lines are an addition to existing literature on the long-run effects of different colonial land policies in India, which primarily focuses on aggregate effects (Banerjee and Iyer, 2005; Banerjee et al., 2005).

[14] Asset ownership data by caste groups comes from the Socio-Economic Caste Census, which is publicly unavailable. The basket of assets reported in this Census is different from that reported in the Population Census, which is used for the estimation of the aggregate effects. The SECC basket is smaller, and leaves out ownership of radios, bicycles, televisions and computers, while including refrigerators.

Table 2: Treatment Effects for Lowest Caste HH

	(1) Average Asset Ownership Index	(2) Nonfarm Establishments Owned Per 100 People	(3) Log Avg Years of Schooling
Reformed Area=1	0.252	0.612***	0.059*
	(0.182)	(0.114)	(0.032)
Sharpened q-value	[0.110]	[0.001]***	[0.071]*
Bonferroni corrected p-value	[0.507]	[0.000]***	[0.198]
Control Mean	2.15	.71	3.48
Sample	Lowest Caste	Lowest Caste	Lowest Caste
Bandwidth	10 km	10 km	10 km
N	8888	9584	8814

Robust standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

1.7 Potential Channel of Persistence

Results reported in Section 1.6 point to positive long-run effects on socio-economic indicators for the entire population, as well as for those lower caste households whose ancestors did not receive any land under the colonial redistribution policy. This is noteworthy, given that since India's independence in 1947, several federal and state-specific redistributive policies have been implemented to ensure convergence among different socio-economic groups. Lowest caste groups, for instance, have benefited from affirmative action in legislatures, public sector jobs and educational institutions. It, therefore, seems that land redistribution led to persistent changes that have positively impacted the lives of lower caste households in the reformed areas. In the following sub-sections, I explore one such potential channel of persistence.

1.7.1 Compliance with Caste-Based Norms

The caste system prescribes different ways of life for different caste groups. Traditionally, elite castes were associated with higher levels of education and putative professions such as warriors, priests and traders. At the same time, intermediate castes provided manual labor in the fields, and lowest castes took on tasks related to "organic, waste, filth, ritual pollution, death, evil spirits or several menial tasks" (Deliege, 1999). Similarly, members of the intermediate and lowest caste groups were not expected to be educated. Given the pro-elite nature of these caste-based norms, it was in the interest of the landed upper caste households to retain the status quo. Ownership of land gave such households adequate power to ensure the enforcement of such norms. It is, therefore, not implausible to think that a policy shock to landowning patterns could have impacted the enforcement of these norms by lowering the ability of the landed upper caste households to enforce them. In fact, Stokes (1983) documents the struggle of non-cultivating upper caste landlords impacted by constant divisions in their property, and that this was more pronounced in the reformed areas.

Qualitative archival evidence also suggests that a first-order effect of land redistribution may have been a reduction in caste-based residential segregation within villages (Baden-Powell, 1892; Ahmed, 1952). Using present day data from the Ministry of Water Resources, I find that these effects persist. Causal estimates reported in Table 3 suggest that redistribution led to smaller villages with fewer number of caste-based habitations per village. Because most official government data captures population by two broad groups: lowest caste (SC/ST) and the rest, I show that across reformed areas, the policy led to a reduction in the number of SC/ST-dominated habitations (>50% population) and those with no SC/STs, suggesting lower residential segregation and higher mixing, which suggests a weakening of caste-based norms.

Table 3: RD Estimate: Village Structure (2018)

	(1) Log Village Area	(2) Log Population Density	(3) Log No. of All Habitations	(4) No. of Habitations with no SC/ST	(5) No. of SC/ST Dominated Habitations
Reformed Area=1	-0.232***	0.005	-0.302***	-0.424*	-0.172**
	(0.053)	(0.052)	(0.086)	(0.232)	(0.074)
Control Mean	197.21	8.78	3.83	1.48	.76
Bandwidth	10 km	10 km	10 km	10 km	10 km
N	11024	10378	9539	9539	9539

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

Given the proximate effects on caste-based residential segregation, it is plausible that land redistribution may have led to long-lasting changes in the functioning of the caste system in the reformed areas. As mentioned earlier, the caste system prescribes limited socio-economic mobility for lower caste households. Therefore, the level of compliance with such restrictions among these castes is also an outcome of interest. This is important, for compliance with such norms shapes aspirations. Literature suggests that optimal levels of aspirations among individuals are important for long-run development (Genicot and Ray, 2017). The caste groups themselves represent a sizeable population of interest. Estimates suggest that people belonging to intermediate and lowest castes make up for over 60% of the state's population (Varshney, 2019). Nationally, as per the Population Census of 2011, members of lowest caste groups (Scheduled Caste and Scheduled Tribe) account for roughly 25% of the country's overall population. The Population Census, however, does not distinguish between the upper and the intermediate castes. Estimates form other government sources, however, suggest that members of intermediate caste groups make up 40-50% of the country's population.[15]

[15] For instance, as per the United District Information System for Education Plus (UDISE+), school enrolment data suggests that intermediate caste groups make up for 45% of the country's population. This is close to the estimate provided by the National Sample Survey Organisation in 2007, i.e. 41% households belong to the intermediate category.

In order to test the long-run effect of redistribution on compliance with restrictive socio-economic norms amog lower caste households, I conducted a field survey, which has been described in Section 1.5.4. During this survey, respondents were asked several questions related to their attitudes towards these norms. For instance, the survey team read out statements related to traditional caste-based socio-economic norms and recorded the level of agreement using a 5-point Likert scale (Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree). These statements are listed in Appendix A.13. For the purpose of estimation of treatment effects, I created a composite index, which equals the mean of responses received to all these statements. Given the traditional linkage between caste and profession, respondents were also asked a question related to their aspirations for their son/grandson: What job would you like your son/grandson (<18 years) to take up when he grows up?

To estimate treatment effects on the aforementioned norms using survey data, I use a modified version of the specification mentioned in Equation (2). While everything else remains the same, the only modification is in the construction of weights used to estimate the coefficient of interest. For each respondent in a randomly selected village, I multiply that village's original weight (as calculated using triangular kernels) by the weight assigned to that household within its bin (a combination of segment, the distance bin and treatment status) based on its caste and that caste's share within the stratum. I also control for the household's caste, respondent's gender and age. In Table 4, I show estimates of treatment effects. The respondents for both the outcome variables are restricted to lower caste households (intermediate and lowest caste households).

Table 4: RD Estimates: Socio-Economic Aspirations

	(1) Index of Caste-Based Stereotypes (Scale 1-5)	(2) Low Quality Job Aspiration for Son
Reformed Area=1	-0.290*	-0.149*
	(0.161)	(0.086)
Sample	Lower Castes	Lower Castes
Control Mean	2.47	.52
Bandwidth	10 km	10 km
N	1801	1492

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Results reported in Table 4 suggest that colonial reforms have led to weaker beliefs in caste-based socio-economic norms/restrictions among lower caste households. On average, the likelihood of respondents agreeing with caste-based socio-economic restrictions for members of their caste is 11.7% lower in reformed areas. Similarly, on average, reforms have led to a 28.6% reduction in the likelihood that respondents aspire for low-quality jobs for their sons/grandsons. Both results are significant at the 10% level. Additionally, in the survey, surveyors also asked respondents about their opinions about the current functioning of of the caste system. Inspired by a question in the World Values Survey (2017-21), respondents were asked the following question.

"In this question, we will give you three basic kinds of attitudes concerning the caste system. Please choose the one which best describes your own opinion.

(a) The entire way our society is organized must be radically changed by revolutionary action

(b) Our society must be gradually improved by reforms

(c) Our present society must be valiantly defended against all subversive forces"

Consistent with findings related to economic mobility, as shown in Table 5, lower caste households in reformed areas are more disapproving of the status quo. Depending on the scale, owing to reforms, on average, lower caste households are 8.5-9% more likely to advocate for changes to the caste system. Similarly, owing to these reforms, on average, they are 71% more likely to advocate for radical changes to the caste system.

Table 5: RD Estimates (Binary Variables Unless Specified)

	Should Caste system be changed?		
	(1) Desire for Any Change (Scale 1-3)	(2) Desire for Any Change	(3) Desire for Radical Change
Reformed Area=1	0.171**	0.078*	0.093*
	(0.078)	(0.042)	(0.051)
Sample	Lower Castes	Lower Castes	Lower Castes
Control Mean	2	.87	.13
Bandwidth	10 km	10 km	10 km
N	1792	1792	1792

Robust standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

1.7.2 Competing Explanations

Although compelling in nature, the argument that indirect effects on lower caste households' socio-economic indicators flow through these attitudinal changes is questionable. It is possible that land reforms may have ushered in economic growth, which may have led to a change in these attitudes. However, anecdotal evidence suggests that reformed areas did not witness any immediate positive economic impact (Dutt, 1900; Stokes, 1983). Empirically, suggestive evidence using district-level data on real agricultural income per acre compiled by Donaldson (2018) also points in the same direction. Using data for districts within the bandwidth, I estimate Equation (3). I regress the log of real agricultural income per acre (y) of a district i during year t on the district's treatment status ($Reformed_i$), while controlling for agro-climatic zone fixed effects (z_i). In Figure 8, I report point estimates of treatment effects for each year during the period 1884-1930 along with 95% confidence intervals. Estimates suggest that during the said period, the two areas were either no different, or the reformed areas lagged behind the unreformed areas in terms of real income. When seen along with evidence on the proximate effect of reforms on caste-based norms (Table 3), results suggest that the causal chain between reforms and socio-economic outcomes for lower caste households seems to be flowing through changes in compliance with norms, and not the other way around.

$$y_{i,t} = \alpha_t + \beta_t Reformed_i + z_i + \epsilon_{i,t} \tag{3}$$

Another potential causal pathway may be linked to colonial governance. It is possible that lower caste households in unreformed areas may have been worse off if the colonial government was more extractive in those areas. Alternatively, lower caste households may have benefited due to better colonial governance in the reformed areas. These factors may have changed compliance with caste-based norms. I compile novel archival data from colonial gazettes from the early 20^{th} century to address these concerns.

To address concerns regarding the extractive nature of colonial governments, I digitized tax

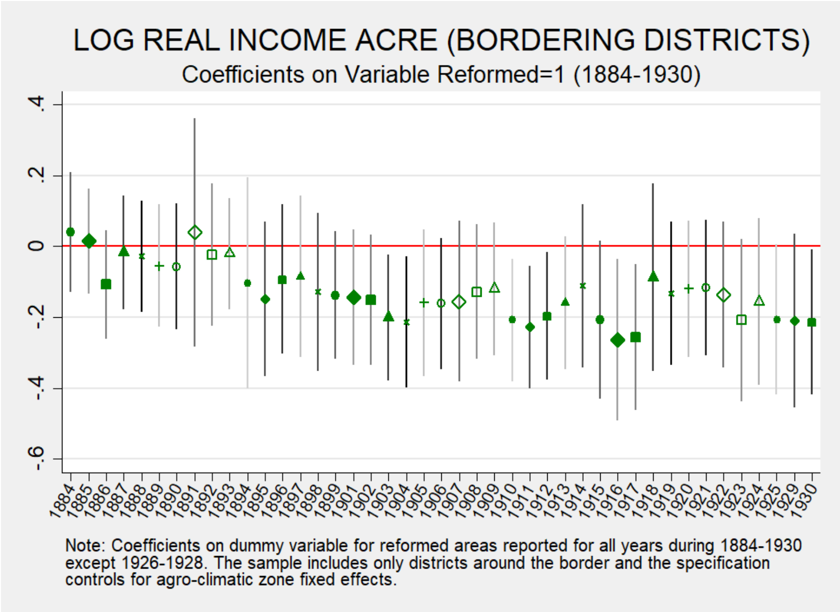


Figure 8: Suggestive Evidence on Proximate Economic Effects of Reforms

demand data at the sub-district level from colonial gazettes. In Appendix A.14.1, I show results from t-tests conducted for tax demands across sub-districts within the bandwidth. Results suggest that between the two areas, the difference in the log of average tax demand per unit cultivable land is statistically insignificant. Similarly, the difference in the log of average tax demand per unit overall land is also statistically insignificant. These gazettes also contain locality-wise information on schools. To address concerns related to colonial governance, I digitized data for 2,780 schools located in sub-districts that are part of the bandwidth. Owing to different spellings and changes in names over time, I manually mapped these schools to modern day villages and cities. At the end of the exercise, I was able to match 2,364 (85%) of the 2,780 schools to modern day areas. As reported in Appendix A.14.2, RD estimates suggest that there was no difference in probability of school placement between the two areas. Colonial governance, therefore, does not seem to be a mediator between redistribution and compliance with caste-based norms.

Existing literature on persistence of long-run effects of historical policies and institutions has focused on channels such as historical provision of infrastructure and public goods (Dell, 2010;

Dell and Olken, 2020), differences in post-colonial governance (Banerjee and Iyer, 2005, Pandey 2010) and nature of relations between elite and non-elite (Acemoglu et al., 2014; Anderson et al., 2015). Results shown in this section provide novel evidence of the persistent impact of a historical colonial policy on compliance with traditional social and cultural norms.

1.8 Conceptual Explanation

Empirical results suggest that over two centuries after their implementation, positive effects of colonial land reforms on three important indicators related to household wealth, structural transformation and human capital persist. Heterogeneity analysis suggests positive effects on the descendants of lower caste households that did not receive land. Further analysis suggests that reforms led to lower compliance with restrictive caste-based norms among lower caste households, which may be a potential channel of persistence.

Lower compliance with caste-based norms challenges the authority of upper caste households. It is, therefore, natural that upper caste households will want to resist any such change. Using a political economy model by Acemoglu and Robinson (2006), I argue that upper caste households in the reformed and unreformed areas face different incentive structures. As a result, such households in the reformed areas are less likely to resist social changes such as a reduction in compliance with caste-based norms among lower caste households.

Acemoglu and Robinson (2006) provide a theoretical framework, which models the behavior of the elite in the face of socio-economic changes that might challenge their authority. An abridged version of the model is laid out in Appendix A.15. The central idea behind the model is that when faced with an option to accept or reject a socio-economic change (such as dilution of norms), elites, such as upper caste households, face a trade-off between replacement and erosion of power. Given this central premise, the model predicts that elites are less likely to block change as political competition rises from intermediate to high levels, and as political rents reduce.[16] Relative to the unreformed areas, upper caste landowners in the reformed areas did face higher political competition. In addition, apart from creating a bigger elite group, reforms also led to lower landholdings per upper caste landowner. Give the socio-political dominance of upper caste households over lower caste households during the colonial period (Prasad 1979; Mohanty, 2001), this meant a reduction in political rents. The model, therefore, predicts that upper caste landowners in reformed areas have had lower incentive to block changes in compliance with caste-based norms among lower caste households.

The model's prediction, when tied with empirical evidence related to social control by traditional elite in democratic societies, provides another prediction that can be empirically tested. Empirical evidence suggests that with the emergence of democratic institutions in post-colonial societies, traditional elites are turning to the use of social capital and trust as a means to ensure persistent social control (Acemoglu et al., 2014; Anderson et al., 2015). These findings and the model's prediction, therefore, suggest that upper caste households in reformed areas should be less cooperative with lower caste households. As a corollary, in the unreformed areas, upper caste households have higher incentive to control society due to lower competition and higher political rents. The model, in conjunction with the literature, therefore, predicts that upper caste households in unreformed areas should be more cooperative with lower caste households. This cooperation may then be used to block socio-economic changes such as reduction in com-

[16] The model also makes the following predictions: elites are less likely block economic change as external threat from foreign powers rises and that they are less likely to block economic changes that are more likely to erode their incumbency advantage. In the context of this paper, these are not relevant, for to the best of my knowledge, any foreign threat after 1856 has been uniform between the two areas. Similarly, we are only concerned with one change, and are not comparing between different types of changes. The last prediction states that elites are less likely to block economic change as levels of human capital among non-elite rise. For instance, elites controlling higher human capital may be tempted to opt for technological/institutional change if they believe that a combination of this human capital and the change can bring about higher benefits than current economic rents. In order to test whether this factor is relevant in this scenario or not, I digitized district-level data related to education levels from the Population Census of 1881 (the first national census conducted by the British). As shown in Appendix A.16, across bordering districts in my sample, percentage of population who were either literate or under instruction as of 1881 is the same. This evidence suggests that levels of non-elite human capital should not have had any bearing on elite's response towards attitudinal changes.

pliance with caste-based norms, which could have benefited lower caste households.

I test the above prediction using two kinds of questions in my survey. First, surveyors asked lower caste households whether their caste group shares the same source of water (well) as the elite. This is an important question (also used by Anderson et al., 2015), for traditional caste relations forbade sharing water from the same source. Additionally, in order to obtain evidence on levels of trust through indirect elicitation, following the method used in Kubota et al (2013), surveyors played the Ultimatum Game (explained in Appendix A.17) with lower caste households, varying the proposer's caste and the amount offered. Results are reported in Table 6 and Table 7.

Table 6: RD Estimates: Water Relations

	(1) Well Sharing with Elite (Modal)	(2) Well Sharing with Elite (Modal)
Reformed Area=1	-0.323***	-0.341***
	(0.106)	(0.112)
Level	Village-Caste	Village-Caste
Sample	All Castes	Lower Castes
Control Mean	.68	.64
Bandwidth	10 km	10 km
N	398	353

Robust standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 7: Ultimatum Game RD Estimates (Outcome: Weighted Rejection Rate)

	Offers From		
	(1)	(2)	(3)
	All Caste	All UC	Traditional Landowning UC
Reformed Area=1	0.059	0.079*	0.092*
	(0.043)	(0.047)	(0.049)
Sample	Lower Castes	Lower Castes	Lower Castes
Control Mean	.48	.52	.54
Bandwidth	10 km	10 km	10 km
N	1801	1794	1794

Robust standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Results (Table 6) suggest that upper caste households across reformed areas are less accommodating towards lower caste households as measured by the likelihood of water-sharing among caste groups (cooperation). On average, responses from lower caste households suggest that reforms have made upper caste households 53% less likely to share water from the same source. Results from the Ultimatum Game (Table 7) point in the same direction. Existing literature suggests that despite rationality dictating acceptance of all offers during the Ultimatum Game, rejections come about due to factors such as lower trust (Dotsch et al., 2008) and negative stereotypes regarding the proposer (Dovidio et al., 1996). In Table 7, I show that lower caste households are more likely to reject offers from the traditional elite in reformed areas, thereby suggesting lower levels of trust and/or stronger negative stereotypes regarding upper caste households. The fact that these two pieces of evidence (water-sharing and Ultimatum Game) point in the same direction suggests that upper caste households in reformed areas have maintained worse relations with lower cast households than their counterparts in unreformed areas. Compared to the unreformed areas, this may have made the process of dilution of caste-based norms easier, which would have benefited the lower caste households.

Further evidence that may point to wider social acceptance of upper caste households among lower caste households in the unreformed areas comes from candidate selection in state-level elections. As shown in Figure 9, evidence from the 2012 and 2017 state-level elections (within the bandwidth) suggests that upper caste candidates, despite their caste's population being in a minority, are more likely to be fielded by the state's traditionally dominant political parties in the unreformed areas (p-value of t-test equals 0.01). This is true even for parties such as the Samajwadi Party and the Bahujan Samaj Party, which were established in the second half of the 20^{th} century as the parties of the intermediate and lowest castes respectively (p-value of t-test equals 0.02).

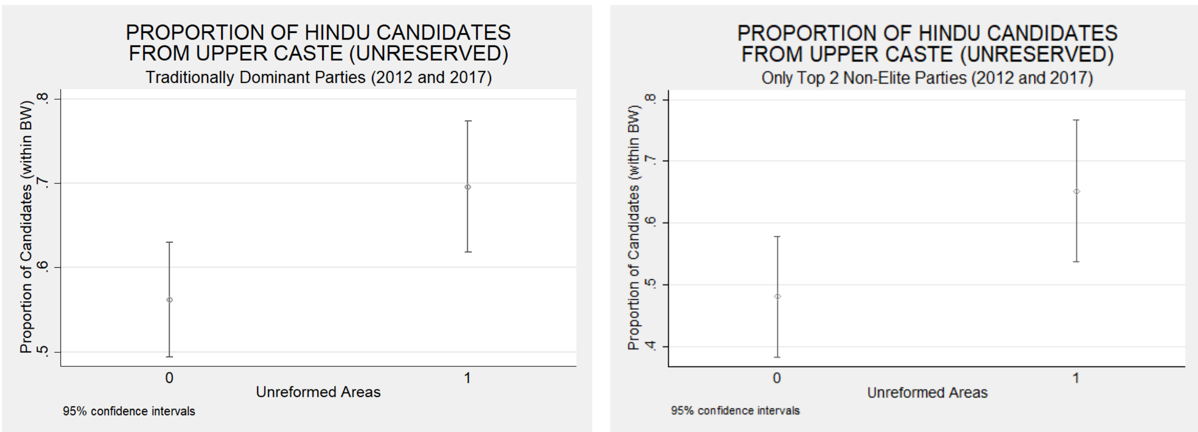


Figure 9: Candidates from Elite Castes

1.9 Conclusion

Policies impacting land ownership patterns may influence long-run socio-economic development (Engerman and Sokoloff, 1997; Dell, 2010). However, the estimation of the impact of such policies faces empirical challenges. Existing literature has focused on institutional, political and economic channels through which differences in land ownership patterns may impact long-run development. However, given the linkages between land ownership and social hierarchies across the world, polices impacting such patterns may operate through another channel, i.e., social and cultural norms.

In this paper, I estimate the long-run effects of land redistribution on socio-economic outcomes by exploiting the characteristics of a colonial policy implemented in the 1820s. Owing to the staggered annexation of India's largest state, Uttar Pradesh, the British authorities, while redistributing land among a larger set of landowners in 72% of the state (reformed areas), granted proprietary titles to a small set of landlords in the remaining area. The process eventually led to a scenario wherein the average number of landlords per 1000 rural people in the reformed areas (49) was almost three times the number in the unreformed areas (17). The policy experiment provides me with a setup for a spatial regression discontinuity design around an arbitrary border separating the two regions.

Estimates using village-level microdata suggest that even after two centuries of implementation and passage of post-independence laws regarding land redistribution across the state, average positive effects of colonial land redistribution on durable asset ownership (6.2%), nonfarm employment (50%) and average years of schooling (6.0%) persist. I also find positive treatment effects on descendants of lowest caste households that resided in the reformed areas, but did not receive land under the redistribution policy. On average, lowest caste households residing in reformed areas are more likely to own durable assets (11.7%), own a higher number of nonfarm establishments (86%), and are more educated (5.9%).

There may be several channels of persistence behind these long-run effects. In this paper, I focus on the impact of redistribution on the functioning of the caste system as a potential channel of persistence. Traditionally, upper caste households were the social elite as well the landed economic elite. In both the reformed and the unreformed areas, over two-thirds land was owned by upper caste Hindu households. A combination of these factors allowed such households to control society as per caste-based norms, which prescribed restriction on lower caste households, especially those belonging to the lowest castes. These norms restricted their socio-economic development. I hypothesize that land redistribution acted as a shock to land concentration in the reformed areas, which may have impacted the ability of upper caste households to control society, as well as lower caste households' attitudes towards caste-based restrictions.

Qualitative and quantitative evidence, for instance, suggests that reduction in caste-based residential segregation may have been a proximate impact of land redistribution in the reformed areas. Using data from a field survey conducted across 189 villages during the Summer of 2022, I find causal evidence that reforms led to higher aspirations (29%, on average) and lower adherence to pro-elite caste-based socio-economic norms among non-elite households (12%, on average), along with a higher appetite for changes in the existing institution of the caste system (8.5-9%, on average). Using novel archival data, I show evidence that it is improbable that compliance with these norms was impacted by other intermediate outcomes resulting from redistribution, such as differences in colonial governance or levels of agricultural productivity.

I argue that these results are consistent with a dynamic political economy model proposed by Acemoglu and Robinson (2006) that explains elite's behavior towards institutional changes that may threaten their hegemony. As per the model, owing to higher competition and lower political rents, landed upper caste households in reformed areas did not have adequate incentive to block the weakening of caste-based norms.

When read in conjunction with recent literature regarding the usage of social cooperation by elite as a means to control society in democratic countries, the model implies lower (higher) cooperation between upper caste and lower caste households in reformed (unreformed) areas. Analysis of survey data provides causal evidence that reforms led to upper caste households being less accommodating towards lower caste households. For instance, on average, reforms have led to lower likelihood of water-sharing between upper caste and lower caste households from the same well (53%, on average). Evidence from Ultimatum Game played with respondents point to lower trust between the two groups (15.3-16.7%, on average).

A holistic reading of these results implies that land redistribution weakened incentives among upper caste landlords to enforce the caste system across reformed areas, which ultimately allowed lower caste households to grow socially and economically. At the same time, owing to high political rents and low competition, upper caste landlords in unreformed areas may have been more inclined and therefore, more successful in controlling society as per traditional caste-based norms. This may have played a role in impeding socio-economic development among lower caste households.

2 CHAPTER 2: Road Connectivity, State Capacity and Bureaucratic Efficacy: Evidence from India[17]

Several developing countries have institutionalized social protection schemes to serve socio-economically marginalized groups. However, timely delivery of such public services remains a challenge. In this paper, I test the hypothesis that enhanced state capacity by means of improved roads can speed up delivery of public services by facilitating easier commutes for on-field bureaucrats. Using village-level data from 2014-15, I estimate the impact of rural roads on the timeliness of last-mile delivery in the context of a national public works scheme in India. Using an identification strategy inspired by fuzzy RD, exploiting exogenous variation in rural road construction provided by arbitrary population-based thresholds under a national road construction program, I find evidence that rural roads lead to a reduction in time taken to process beneficiary payments. Results from heterogeneity analysis show that these effects are driven by villages far away from local government headquarters. Spatial heterogeneity in treatment effects may point to complementarity between state capacity and bureaucrats' incentives.

[17]I would like to express my sincere gratitude to Sebastian Galiani, Jessica Goldberg, Allan Drazen and Ethan Kaplan for their invaluable insights and comments. My sincere gratitude to Deepak Dulani at the NIC, Ministry of Rural Development, Government of India. I thank Dheeraj Chaudhary, Tomohiro Hara, Elena Ramirez, Anusuya Sivaram, Matthew Clay Summers, Jorge Enrique Perilla Garcia, Danny Kolliner, Weizheng Lai and Juan Angulo Santacruz for their inputs. I am also grateful to participants at the 1st Annual World Bank Conference on Transport Economics (March 2023) for their suggestions. Finally, I am thankful to Sam Asher for his guidance on data-related issues.

2.1 Introduction

Populist welfare schemes have become an important component of economic policymaking across developing countries. The Indian context is no exception. Over the last few decades, programs aimed at rural employment, sanitation, food security, and education have been conceived and implemented by successive federal and state-level governments. However, most of these schemes suffer from issues related to last-mile delivery (Marfatia 2019). Implementation of these schemes often requires significant on-field coordination among bureaucrats and representatives elected to local governing bodies. Hence, indicators of state capacity such as quality of last-mile road connectivity may influence timeliness of service delivery. In this paper, I investigate whether an increase in state capacity, as measured by construction of feeder roads in rural areas, improves outcomes related to last-mile delivery in the context of a popular public works scheme in India.

Identification of the effects of rural road placement faces challenges related to endogeneity. For instance, roads may be constructed in areas that are better governed or are politically more important for the government. In order to address these concerns, I exploit the characteristics of a national-level rural roads program (Pradhan Mantri Gram Sadak Yojana, or PMGSY) launched in India in December 2000. The program uses arbitrary population-based thresholds to prioritize construction of paved roads across villages, which are integrated with the larger all-weather road network of the area. In the absence of manipulation in population records, the number of villages that are just eligible and just ineligible, i.e. those with population only slightly higher and lower than the thresholds, should be exogenous. This allows me to instrument the ratio of villages connected by a paved road in an area by the number of villages that are just eligible for road construction.

I use the aforementioned strategy to assess the impact of rural roads on the timeliness of last-mile delivery in the context of a national-level public works scheme (Mahatma Gandhi

National Rural Employment Guarantee Act, or MGNREGA) during the financial year (FY) 2014-15.[18] Under this scheme, beneficiaries can demand manual work from their local governing body, and should receive their payments within 15 days of work. However, at least during FY 2014-15, delays in processing beneficiary payments were a major challenge (Narayanan, Dhorajiwala and Golani 2019). During the said period, this process involved several on-field steps that required inter-personal coordination, which meant that officials had to often travel within their assigned areas, as well as to and from the local sub-district office. By facilitating easier commutes for such officials, improved rural connectivity should have ideally enhanced the state's capacity to ensure timely payments. Therefore, in this paper, I estimate the impact of rural road connectivity on the time taken to process beneficiary payments under MGNREGA.

Under MGNREGA, the payment process consists of on-field and off-field steps. As per the program's provisions, all these steps should be completed within 15 days of a beneficiary's labour. The program's official website provides data on the number of transactions and amount of beneficiary payments for which *all on-field* steps were completed within a) 15 days, b) 16-30 days, c) 31-60 days, d) 61-90 days and e) more than 90 days. I estimate treatment effects of road construction on the proportion of transactions processed within each of these intervals. My estimates suggest that road construction impacts the proportion of transactions processed within the 16-30 day and 31-60 day intervals. After controlling for a rich set of baseline covariates and fixed effects, I find that ensuring road connectivity across all just eligible villages in an area raises the proportion of transactions for which on-field steps are completed during the 16-30 day interval by 12.8 percentage points (over a control mean of 27%). Correspondingly, the treatment effect on the proportion of transactions processed within the 31-60 day interval is minus 12.4 percentage points (over a control mean of 24%). When the outcome variable changes to proportion of amount processed, these estimates change to 15.3 percentage points

[18]The Indian financial year is defined as the period between April 1 and March 31 (of the subsequent year).

over a control mean of 27% for the 16-30 day interval, and minus 14.4 percentage points over a control mean of 24% for the 31-60 day interval. Given these results, I conclude that rural road construction ensures quicker completion of on-field steps concerning payment processing under MGNREGA. On average, this implies quicker processing of annual payments worth USD 2,193-2,705 in a rural area with total annual beneficiary payments worth USD 17,684.

Results are robust to the usage of a different yet related endogenous regressor, i.e. ratio of population (and not villages) in an area connected by roads. They are also robust to widening the definition of road coverage to include all roads (PMGSY and non-PMGSY). For my benchmark specification, I use a bandwidth of +/-50 people around the mandated thresholds (500 and 1,000) to determine just eligible and just ineligible villages. I, therefore, test whether results are sensitive to choice of bandwidth. While treatment effects within alternate bandwidths are statistically indistinguishable from the benchmark estimate, I find that the magnitude of point estimates increases as bandwidths are shortened, thereby suggesting that treatment effects may be localized.

I also find that treatment effects are spatially heterogeneous, for positive treatment effects on timeliness of payments are driven by areas farther away from local administrative (sub-district) headquarters. These results rule out general improvements in efficacy of on-field bureaucrats as a potential mechanism. Hence, although suggestive, I offer an explanation based on monitoring and oversight of on-field functionaries by higher officials. Given that transportation costs can be an impediment for monitoring (Asher, Nagpal and Novosad 2018), it is possible that by making remote villages more accessible, rural roads allow officials at the sub-district level to focus more on these newly connected villages. Under MGNREGA, on-field functionaries are not rewarded for good performance, but they may be punished for poor performance by means of show-cause notices and suspensions. It is, therefore, in the interest of on-field functionaries to focus on areas where senior officials are more likely to be pro-active. These results may, therefore, be suggesting that on-field functionaries leverage higher state capacity in areas where they have higher incentives to perform better.

Larger treatment effects in remote areas may be a result of other factors. For instance, roads built in those areas may be of superior quality. However, I rule out this possibility by showing that the probability of receiving a road made of bitumen or concrete is not a function of distance to sub-district headquarters. Neither do I find any reduction in total MGNREGA transactions in such areas, which may have potentially increased timeliness of payments by reducing bureaucrats' workload.

The rest of the paper is organized as follows. In Section 2.2, I outline the contribution of this paper to existing literature. Section 2.3 describes the setup in context of the two schemes. Section 2.4 contains the empirical specifications, while Section 2.5 consists of the main results. In Section 2.6 and Section 2.7, I show results from robustness checks and explore potential mechanisms respectively. In Section 2.8, I outline the paper's conclusions.

2.2 Contribution to Literature

Across developing countries, enhanced road connectivity in rural areas has been widely associated with positive socio-economic outcomes (Jacoby 2000; Khandker, Bakht and Koolwal 2009; Mu and van de Walle 2009; Aggarwal 2018; Asher and Novosad 2020), although Faber (2014) suggests that the Chinese highway system negatively impacted local GDP in areas newly connected to metropolitan centers. Across qualitative and quantitative studies, rural roads have also been associated with improved access to public goods such as medical and education facilities among potential beneficiaries (Porter 2002; Gage and Calixte 2007; Kanuganti et al. 2015; Adukia, Asher and Novosad 2020). I contribute to these strands of literature by estimating the causal effects of rural road placement on last-mile delivery of public goods from the perspective of the local bureaucracy. This is somewhat close to a descriptive study by Bryceson, Bradbury and Bradbury (2008), who find that in the Ethiopian context, villages with better roads attract better service infrastructure and staffing.

I also contribute to a second body of literature that focuses on conditions that can possibly enhance a government's performance at the local level. Existing literature focuses on factors such as workplace diversity (Rasul and Rogger 2015), performance-linked financial and non-financial incentives (Muralidharan and Sundraraman 2011; Duflo, Hanna and Ryan 2012; Ashraf, Bandiera and Kelsey Jack 2014), electoral accountability (Besley and Burgess 2002) and formal and informal monitoring and oversight mechanisms (Tendler 1997; Davis 2004; Olken 2007; Tsai 2007; Björkman Nyqvist and Svensson 2009). To my knowledge, however, empirical evidence on the role of bureaucratic/state capacity is limited (Muralidharan, Niehaus and Sukhtankar 2016; Bozcaga 2020). Muralidharan et al. (2016) is especially relevant, for by means of a randomized controlled trial in the Indian state of Andhra Pradesh, they estimate the treatment effects of enhanced state capacity by introducing a biometrically authenticated payments infrastructure within the context of the same public works program (MGNREGA). They find that access to such a system among beneficiaries makes payments more timely and prone to less corruption. I contribute to this strand by approaching the issue of enhanced state capacity (rural roads) from the perspective of the local bureaucracy.

2.3 Context

2.3.1 About MGNREGA (Outcome)

Approximately half of India's total workforce depends on agriculture- a seasonal activity (The World Bank 2020). At the same time, as of 2015-16, the ratio of gross irrigated area to the total cropped area was less than 50%, making agriculture a risky proposition (Press Information

Bureau 2020). Acknowledging this, in the year 2005, the federal government launched a rural employment-based social insurance scheme (called MGNREGA, or the Mahatma Gandhi National Guarantee Act), which entitles each rural household to 100 days of unskilled work every year. Under this scheme, each household is issued a job card with each adult family member's name and relevant details (Figure 10). At any given point in time, any family member(s) can go up to the local scheme coordinator (called the Gram Rozgar Sevak, or the GRS) and demand work, which should be provided within a 15-day period. While the program is funded primarily by the federal government, it is administered by respective state governments by means of local government officials at the district, sub-district/block and Gram Panchayat (GP) level.[19]

Furthermore, the legislation states that beneficiary wages should be paid within 15 days of work. As per the law's provisions, if this deadline is not met, the State Government is obligated to pay delay compensation to beneficiaries and later recover the amount from functionaries responsible for the delay. Yet, payment delays have remained an impediment to the scheme's smooth implementation. For instance, after analyzing 9 million transactions across 10 states for the financial year 2016-17, Narayanan et al. (2019) show that only 21% of all payments were made on time, i.e. within the mandated 15-day period of work. They also find major flaws in the calculation of delay compensation.

During the period FY 2014-15, the payment process was split into two broad components.[20] Stage 1 (which this paper focuses on) consists of all the on-field steps. While different states might have had different operating procedures, a broad summary of the on-field steps is as follows. Once beneficiaries submit their demand for MGNREGA work, official documentation required to record beneficiaries' attendance is generated through the official Management In-

[19]As shown in Figure 11, each Indian state is divided into districts, which are, in turn, divided into sub-districts/development blocks. Each sub-district/block has several Gram Panchayats (GPs). Each GP may contain one or more villages. The GP is the smallest unit of administration for the purpose of MGNREGA.

[20]The payment process steps keep changing as procedures and facilities available to local officials change. However, the mentioned steps were relevant to the period of this study, i.e. FY 2014-15.

formation System (MIS). This usually happens at the sub-district/block office, which caters to several GPs. Once this document is generated, beneficiaries work in and around their village for a period of 1-2 weeks under the supervision of the GRS, after which an engineer is supposed to visit the site to inspect the work done and calculate the group wage. Their assessment is physically submitted to the sub-district/block office after approval from the local GP-level elected representatives and another GP-level bureaucrat (called the Sachiv, or the Secretary). Once the documents are sent to the sub-district/block office, they are processed and submitted onto an online server for digital payment processing. After the completion of these steps, Stage 2 begins. During this phase, a state-level centralized system processes the uploaded documents and transfers the due amount into the beneficiaries' bank accounts. As noted during field visits, all of the involved on-field officials receive a flat monthly salary and are unlikely to receive any positive monetary incentives for good performance, although poor performance, whenever noticed by senior officials, can elicit negative consequences such as show-cause notices and suspension orders (Press Information Bureau 2011). Figure 12 summarizes time taken to complete all steps of Stage 1 for all states in my sample during FY 2014-15.[21] As shown, despite the stipulated time limit for the entire process (Steps 1 and 2) being 15 days, Step 1 for for only 26% of all transactions was completed within that interval.

Intuitively, construction of rural roads can augment state capacity by facilitating easier commutes for government officials working in the field. For instance, while GRSs commute within their assigned GPs (across villages and work sites), each engineer is assigned to multiple GPs, which makes inter and intra-GP travel an important component of their work. Similarly, as field experiences suggest, GRSs are also supposed to visit the local sub-district/block office regularly for MIS-related tasks as well as in-person meetings regarding their performance and administrative briefings. The GP-level bureaucrat (Sachiv) is also assigned to several GPs, and

[21] As discussed later, due to phased implementation of online payments across different states, publicly available data for other states for 2014-15 is not completely reliable.

as mentioned above, their presence and signature/approval is important for payment processing. In summary, better roads can ensure faster completion of these steps by ensuring easier commutes for on-field officials.

In Stage 2, on the other hand, uploaded data is processed by centralized servers and payments are credited into bank accounts. For a long period since the scheme's inception (including my study period), data regarding time taken to complete this stage is not available publicly. At the same time, there seems to be much less scope for roads to impact timeliness of payments in this stage, apart from making it easier for aggrieved beneficiaries to visit the district headquarters (where most online transaction-related issues are resolved) for grievance redressal. Hence, I focus on Stage 1 only.

2.3.2 About PMGSY (Treatment)

In December 2000, the then Indian Prime Minister inaugurated the Pradhan Mantri Gram Sadak Yojana (PMGSY), a national rural roads scheme aimed at providing "good all-weather road connectivity to unconnected" rural areas by means of paved roads. Officially, the aim of the scheme is to connect unconnected habitations (smaller components of a village) that are located at least 500 meters (0.3 mile) away from an all-weather road or another connected habitation. As part of this scheme, these unconnected habitations are to be connected "to nearby Habitations already connected by an All-weather road or to another existing All-weather road so that services (educational, health, marketing facilities etc.), which are not available in the unconnected Habitation, become available to the residents."

In order to execute this scheme, three arbitrary population thresholds were decided. As per guidelines, first priority is to be accorded to habitations (or hamlets) with over 1,000 people, followed by those with 500-999 people and 250-499 people. However, the program's guidelines also state that other qualitative factors may be considered while deciding an unconnected habitation's priority status. For instance, smaller villages can be prioritized if they are too close to a village prioritized on the basis of a population cut-off point. Similarly, factors such as proximity to local market, health facilities or local administrative headquarters, and suggestions of local legislators are also considered before finalizing projects. Nevertheless, as discussed in Section 2.4.2, one can identify causal effects of road construction even if the population-based rules are partially followed.

2.4 Identification Strategy

2.4.1 Sample

After discussions with the National Rural Roads Development Agency, Asher and Novosad (2020) found that owing to different reasons, PMGSY roads were constructed as per population thresholds across six states: Chhattisgarh, Gujarat, Madhya Pradesh, Maharashtra, Odisha and Rajasthan.[22] At the same time, digitization of the payment process under MGNREGA happened in a phased manner across the country. After telephonic consultations with relevant government officials at the National Informatics Center (NIC) at India's Ministry of Rural Development, and officials in the MGNREGA departments of Madhya Pradesh and Gujarat, I found that of the six states mentioned in Asher and Novosad (2020), by FY 2014-15, four had digitized their payment systems, thereby providing reliable data related to transactions.

The four states-Madhya Pradesh, Odisha, Gujarat and Rajasthan-are shown in Figure 13, and constitute the sample area for this paper. As per the Population Census of 2011, these states make up for around 20% of the national population. Whereas over 70% of the population in Madhya Pradesh (72.4%), Odisha (83.3%) and Rajasthan (75.1%) is rural, Gujarat is one

[22] As per Asher and Novosad (2020), in other states such as those in southern India, owing to high development of infrastructure, the scope of additional road placement was very limited. In other states with weaker state capacity, such as Bihar, guidelines were not followed.

of India's more urbanized states with 57.4% rural population. During FY 2014-15, beneficiary payments across these states accounted for roughly 20% of all beneficiary payments across India. As shown in Figure 12, during FY 2014-15, only 26% of all transactions across these states were processed within 15 days, whereas 31% were processed within the 16-30 day interval. Another 23% of all transactions were processed within 31-60 days while the remaining (roughly) 20% were equally distributed between the remaining two intervals: 61-90 days and more than 90 days.

Further, in Figure 14 and Figure 15, I show the distribution of ratio of villages connected by PMGSY roads across GPs in these states (as of March 31, 2014), and the ratio of villages connected by any concrete road by the year 2009.[23] By March 31, 2014, over 50% of GPs had no PMGSY roads, whereas in around 20% of GPs, all villages were connected. On average, approximately one in three villages in each GP was connected by a road, although this number rises to 70% when one considers a broader definition of road connectivity, i.e. any kind of concrete road. As Asher and Novosad (2020) mention, during their consultations with officials at the National Rural Roads Development Agency, they found out that while Madhya Pradesh adhered to the 500 and 1,000 thresholds, the remaining three states adhered to only one of the cut-off points, i.e. 500. Hence, while for Madhya Pradesh, I consider both the relevant thresholds (500 and 1,000), I only use the 500 threshold for the remaining three states.

2.4.2 Identification Strategy

With cross-sectional data for a given financial year, the ideal OLS specification at the habitation level (the level at which road placement is decided) is shown in Equation (4).

[23]These graphs represents 35.6 thousand of the 57.8 thousand GPs across the four states (62%). These GPs comprise the unrestricted sample, which I use to create the final sample for my IV estimation strategy. I drop the remaining 38% GPs due to absence of string matches across the different datasets that I use to compile the final dataset.

$$AveragePaymentTime_h = \eta + \psi PMGSY_h + \phi X_h + m_h \tag{4}$$

The outcome variable, $AveragePaymentTime_h$, represents the average time taken to complete Stage 1 of the payment process in habitation h. In a simple OLS specification, this is explained by a binary variable $PMGSY_h$ that equals 1 if the habitation is connected by a PMGSY road, along with habitation-level controls (X_h). In this case, ψ will be the treatment effect and a negative sign on ψ will indicate better last-mile delivery due to road connectivity. However, this strategy faces several challenges. Omitted variable bias is one of them. For instance, if roads are built in areas with higher levels of civic awareness, the latter may also influence timeliness of beneficiary payments. Hence, if roads do reduce payment processing time, the negative impact will be over estimated. Similarly, politicians may lobby for roads in areas with traditionally marginalized sections such as lower caste households. Despite road construction, the latter may face the apathy of the local bureaucracy. In such a scenario, the negative impact of road construction on payment processing time may be under estimated. I, therefore, turn to an instrumental variable strategy for estimation of treatment effects. In addition, publicly available data on payment delays under MGNREGA is available at the Gram Panchayat (GP) level, which is much coarser, for a GP may contain one or more villages, which, in turn, may contain one or more habitations. As a result, I use a modified version of the fuzzy RD setup as my benchmark empirical specification (Bhalotra and Clots-Figueras 2014).

As discussed above, under PMGSY, population thresholds are used to determine a habitation's eligibility to receive a road. Within a narrow bandwidth around these cut-off points, a habitation may be deemed as "just eligible" or "just ineligible". Within a limited geography, habitations with populations just above (treatment) and just below (control) the cut-off points should be similar, which makes the placement of a PMGSY road in a just eligible habitation an exogenous process, especially because the eligibility criteria refer to population data from 2001 and the outcome variable comes from the period FY 2014-15. However, Asher and Novosad (2020) note that habitation population numbers seem to be manipulated around the cut-off points in PMGSY's internal data records, which is also visible around the 250, 500 and 1,000 marks in Figure 16. Hence, like Asher and Novosad (2020), I use village-level population data from the Population Census of 2001 (and not the individual habitation population from PMGSY records) to determine eligibility. As Figure 17 shows, unlike in Figure 16, there is no visible bunching for village-level population around these cut-off points. This ensures that villages that received roads despite ineligibility (and vice versa) due to manipulation are treated as non-compliers.

Usage of population data at the village level (instead of the habitation level, the intended unit of treatment) is not problematic for at least two reasons. Of all the villages in the states that I consider in my sample, more than 70% villages are made up of a single habitation, which effectively means that the habitation is the same as the village. At the same time, under the rules of the scheme, populations of neighboring habitations (within a radius of 500 meters) in a village can be clubbed in order to determine the area's overall eligibility. Therefore, treatment at the village level (i.e. PMGSY connection) is defined as the provision of PMGSY connectivity to at least one habitation in the village (before FY 2014-15).[24] In case more than one habitaton is connected, following Asher and Novosad (2020), I use the earliest PMGSY road as the date of treatment.

As mentioned above, within a narrow bandwidth around the population cut-off points, treatment status of villages just eligible and just ineligible is exogenous. I exploit this exogeneity to predict the total proportion of villages connected by PMGSY roads in a GP. To the extent that I

[24]In Appendix B.6, I show that my main results do not change if I restrict my sample to those GPs where all villages have only one habitation.

use such close cases to predict the overall proportion of PMGSY-connected villages, this empirical design is somewhat similar to that used by Bhalotra and Clots-Figueras (2014). However, while Bhalotra and Clots-Figueras (2014) control for a third-order polynomial of the running variable to estimate treatment effects, following the recommendations of Imbens and Lemieux (2008), and Gelman and Imbens (2019), I use the running variable's (village population from 2001) first order polynomial. Equation (5) and Equation (6) characterize my empirical strategy.

First Stage:

$$PMGSY\_Coverage_{gnbD,2014} = \delta + \kappa Z_{gnbD,2001} + \pi J_{gnbD,2001} + \theta_D + r_b + \nu_{gnbD} \quad (5)$$

Second Stage:

$$Y_{\tau,gnbD,2014-15} = \alpha_\tau + \beta_\tau PMGSY\_\widehat{Coverage}_{gnbD,2014} + \gamma_\tau J_{gnbD,2001} + \lambda_{\tau,D} + h_{\tau,b} + \epsilon_{\tau,gnbD} \quad (6)$$

In the first stage (Equation (5)), I predict the fraction of villages in a GP g (with a total of n villages in district D) connected by a PMGSY road (as of the beginning of FY 2014-15) using the number of villages just eligible under the PMGSY scheme (as determined on the basis of population in 2001), i.e. $Z_{gnbD,2001}$.[25] Owing to aggregation from the habitation/village level (where the discontinuity is relevant) up to the GP level in the final specification, I am not able to use a suitable data-driven technique to find the optimal bandwidth. I, therefore, begin with a narrow bandwidth of +/- 50 people around the cut-off point and then conduct robustness checks with different bandwidths. As a result, in the main specification, just eligible villages are those with population between 500 and 550. Additionally, in the case of Madhya Pradesh,

[25] The subscript b is representative of GP g's composition in terms of total number of villages within the benchmark bandwidths (i.e., both just eligible and just ineligible villages on either side of the cut-off points), which I explain in Appendix B.2.

villages with population between 1,000 and 1,050 are also considered as just eligible. Similarly, just ineligible villages are defined as those with populations between 450 and 500 (for all states in sample), and 950 and 1,000 (for Madhya Pradesh only).

Next, I use the predicted values for $PMGSY\_Coverage_{gnbD,2014}$ from the first stage and estimate Equation (6). The format of the outcome variable ($Y_{\tau gnbD,2014-15}$) is such that I have the number of transactions and amount for which Stage-1 (field work plus data upload from the field) was completed within the following time periods (τ) during FY 2014-15: a) 15 days, b) 16-30 days, c) 31-60 days, d) 61-90 days and e) more than 90 days. I, therefore, estimate Equation (6) for each of the aforementioned values of τ.

In both the specifications, apart from the total number of villages in the GP, I control for a rich set of GP-level baseline covariates from the year 2001 in order to control for any possible selection on observables ($J_{gnbD,2001}$, explained in Section 2.4.3 and Appendix B.1). I also control for district fixed effects (θ_D and $\lambda_{\tau,D}$ in Equation (5) and Equation (6) respectively) in order to capture heterogeneity in governance levels across different areas.[26] In addition, in order to control for the effects (if any) of the total number of villages within the bandwidth (either side), as explained in Appendix B.2, I control for fixed effects related to the overall number of villages within the bandwidth (r_b and $h_{\tau,b}$ in Equation (5) and Equation (6) respectively). For instance, these fixed effects will distinguish between GPs with 2 villages within the overall 450-550 bandwidth and those with 3 villages within the overall 450-550 bandwidth. In the case of Madhya Pradesh, these will also control for any potential effects related to the two different cut-off points being used.

[26]In Section 2.7, instead of district fixed effects, I control for sub-district fixed effects, and find that results remain robust. This is relevant, for officials at both the district and sub-district/block levels are responsible for the implementation of PMGSY and MGNREGA at the local level.

2.4.3 Data

Table 8 lists all the data sources used in the analysis.[27] Asher and Novosad (2020) scraped PMGSY roads data at the level of habitation/hamlet from the scheme's official website (until January 2015). Because the roads dataset does not contain the name of the GP, I obtained the GP-village mapping data from the Government of India's Local Government Directory.[28] [29] I scraped data on MGNREGA payment processing for FY 2014-15 from the MGNREGA website. As shown in Figure 12, Stage 1 for only 26% of transactions across these states was completed within the stipulated 15 days. Within the sample, there is notable heterogeneity (Figure 18). In Gujarat, for instance, for 45% of transactions, Stage 1 was completed within 15 days. In Madhya Pradesh, on the other hand, this ratio is much lower, i.e. around 16%.

I merged the MGNREGA and PMGSY data with datasets containing socio-economic variables obtained from the SHRUG (Socioeconomic High-resolution Rural-Urban Geographic Dataset for India) Database (Asher and Novosad 2019), which is effectively a compilation of different official datasets (population census, economic census, election results, etc.) related to India. In the benchmark specification, I control for the following baseline covariates from 2001: access to road pre-2001, access to power, access to bank, access to communication, access to medical facilities, access to primary school, distance to nearest town, night light luminosity (square root), proportion of scheduled caste/scheduled tribe population, proportion of illiterate population and ratio of irrigated land to the total farmed land. However, SHRUG provides this data at the village-level, whereas my unit of analysis is a GP, which may contain more than one village. Therefore, for all covariates, I calculated a weighted average of village-level

[27] Definitions of all variables can be found in Appendix B.1.

[28] Given my identification strategy, in my final sample, I use only those GPs for which the following two conditions are satisfied: a) MGNREGA take up for FY 2014-15 was non-zero, b) Data for PMGSY and other controls is available for each of the villages in the GP.

[29] Due to the absence of unique numerical identifiers at the GP level across the GP-Village mapping, PMGSY and MGNREGA datasets, I used fuzzy string matching. In order to ensure high levels of efficiency, I create mini datasets for each block and merged them with similar block-level mini datasets in order to ensure that the fuzzy merge score comes from the GP name only.

values. For all covariates apart from the ratio of irrigated land to the total farmed land, a village's baseline population (2001) was used as its weight.[30] For the ratio of irrigated to total farmed land, a village's farmed land area was used as its weight. Finally, for heterogeneity analysis, I obtained data on distance between villages and sub-district headquarters from the 2011 Population Census.

2.5 Main Results

2.5.1 OLS Results

I begin with a naive OLS regression analysis for all GPs for which data is available across the four states in sample. I estimate Equation (7), where the variables and notations are the same as explained in Section 2.4.2. The coefficient of interest is μ_τ.

$$Y_{\tau gD,2014-15} = \omega_\tau + \mu_\tau PMGSY\_Coverage_{gD,2014} + \rho_\tau J_{gD,2001} + \zeta_{\tau,D} + s_{\tau,gD} \quad (7)$$

The results, as reported in Table 9 and Table 10, suggest that as the ratio of villages with PMGSY coverage rises in a GP, the average time to process Stage 1 does not seem to change in any economically significant manner. As explained in Section 2.4.2, there are reasons to believe that these results may be biased, and that the direction of the bias is indeterminate.

2.5.2 IV Strategy: First Stage and Placebo Checks

Strength and Relevance

Before showing estimates of causal effects, I test the strength and relevance of my instrument by estimating Equation (5), i.e. the first stage. For the same, I regress the ratio of PMGSY-

[30]For night light luminosity, I calculate the square root of the weighted average.

connected villages in a GP on the number of villages in that GP that are just eligible for PMGSY road construction. I define just eligible villages as those with population within the range of 500-550 people. In addition, for Madhya Pradesh, I also consider villages with population in the range of 1000-1050 within this category. While doing so, I control for the total number of villages in the GP and baseline (2001) GP-level covariates as well as fixed effects for district and composition of the total number of villages within the bandwidth (as explained in Appendix B.2).

As shown in Table 11, robust to inclusion of controls and fixed effects, the coefficient on the instrument, the number of just eligible villages, is statistically significant and is in the range of 3.0% to 5.8%. Given that the control mean (i.e. ratio of villages with PMGSY in GPs with no such just eligible villages) equals 29.8%, this implies that an additional just eligible road increases the coverage ratio by 10-19.5%. The F-statistic for the instrument also remains comfortably above the commonly used benchmark of 10 as well as the Stock-Yogo 10% critical value of 16.38.[31]

Placebo Checks and Validity

In this section, I confirm the validity of the empirical strategy by running some placebo checks. If the strategy is valid, then the instrument should not impact any pre-determined outcomes, i.e. those before 2001. In Table 40 (Appendix B.3), I report estimates for the coefficient on the instrument, using each of the 2001 baseline covariates as outcome variables (while controlling for the others). As shown, the instrument is not correlated with any of the pre-determined outcomes, which could potentially impact payment delays under MGNREGA (the main outcome).[32]

[31] Lee et al. (2022) mention that the commonly used practice of comparing the first stage F-statistic against the rule-of-thumb threshold of 10 may lead to the construction of misleading confidence intervals. In Section 2.5.3, I address this concern.

[32] In the entire paper, SC/ST referes to Scheduled Castes and Scheduled Tribes, i.e., socio-economically vulnerable groups as identified by the Government of India.

In addition, to my knowledge, no other national-level program targets villages using these cut-off points at the habitation/village level. Spears (2012) suggests that under the Total Sanitation Program (TSC) (a national-level program), there are different incentive structures for GPs above and below the 1000 threshold (which I use only in the case of Madhya Pradesh) for achieving sanitation targets. Ideally, this should not have been an issue, for sanitation levels should in no way affect bureaucratic efficacy. However, beginning 2011, the Government of India allowed convergance of TSC and MGNREGA, i.e. sanitation works could be completed under the umbrella of MGNREGA. In that case, given the discontinuity at 1,000, there could be a scenario where GPs with population above 1,000 and below 1,000 people may be exerting different levels of pressure with regard to timely inspections so that they become eligible for incentives under TSC. However, TSC guidelines are implemented at the GP level, whereas PMGSY guidelines are implemented at the habitation/village level. Hence, any overlap in the implementation of TSC and PMGSY guidelines should occur in GPs with only one village. In my final sample, only 4% of GPs contain a single village. Furthermore, in Table 41 (Appendix B.3), I run another set of placebo checks using the two stage least squares (2SLS) strategy. I find no causal effects of PMGSY roads on baseline covariates.

2.5.3 IV Strategy: 2SLS Results

In Table 12 and Table 13, I report results of the benchmark two stage least squares empirical strategy using the first stage specification from column (4) of Table 11, i.e. including baseline GP-level controls and fixed effects. In accordance with Andrews, Stock and Sun (2018), apart from the Kleibergen-Paap F-statistic and the Cragg-Donald F-statistic, I also report the Effective F-statistic. Across the three statistics, the values are comfortably above the benchmark value of ten, as well as the Stock-Yogo 10% critical value of 16.38.

Results from Column (2) in Table 12 suggest that on average, in a GP that contains only just eligible villages, 100% PMGSY road coverage leads to an increase of 12.8 percentage points in the ratio of transactions processed within the 16-30 day interval. This result is significant at the 10% level. At the same time, results from Column (3) suggest that 100% road coverage leads to a reduction of 12.4 percentage points in the ratio of transactions processed within the 31-60 day interval. This result is also significant at the 10% level. In addition, while point estimates from Columns (1) and (5) suggest a similar trend-roads lead to fewer transactions being processed beyond the 60-day mark-they are not statistically significant at conventional levels.

In Table 12, I show the impact of road connectivity on the proportion of transactions processed within different time periods. Similarly, in Table 13, I show these effects on the proportion of total amount processed within the same time periods. Both sets of results are consistent with each other. Point estimates in Table 13 suggest that the proportion of amount processed beyond the 30-day mark declines in treated villages and that these amounts are now processed within the 0-30 day time window. However, as Columns (2) and (3) suggest, effects are significant only for the 16-30 and 31-60 day time periods. As per these results, in treated GPs, the percentage of amount processed within the 31-60 day interval declines by 14.4 percentage points and that processed within the 16-30 day interval rises by 15.3 percentage points. Cumulatively, this implies that while in control villages, 54% of the amount is processed within the 30-day window, treating the effect on the 15-day margin in Column (1) as null, in treated villages, this proportion rises to 69.3%. Across GPs in my sample, on average, a total of 1185 transactions worth USD 17,684 were processed during the financial year 2014-15. Statistically significant point estimates shown in Table 12 and Table 13, therefore, suggest that on average, 100% PMGSY road coverage among just eligible villages would have led to quicker processing of transactions worth USD 2,193-2,705 during the financial year.[33]

[33] 1 USD=INR 61 (approximate average figure for 2014-15)

Lee et al. (2022) suggest that construction of confidence intervals following a satisfactory comparison of the first stage F-statistic against the rule-of-thumb threshold of 10 may lead to incorrect conclusions. As an alternative, they provide a method (the tF procedure) to adjust the t-statistic based on the first stage F-statistic. In Table 12 and Table 13, I show 95% confidence intervals calculated using the tF procedure. For estimates that are conventionally significant at the 5% and 10% levels, the tF-adjusted 95% confidence intervals either do not span zero or do so marginally.

2.6 Robustness Checks

In this section, I conduct robustness checks. I begin with falsification tests using arbitrary thresholds in the first stage. Using the actual thresholds (500 and 1,000), I then vary the size of the bandwidth. I also change the endogenous regressor to the proportion of a GP's population (instead of villages) connected by PMGSY roads. Finally, I estimate the benchmark specification using a different and more broadly defined endogenous regressor (all paved roads) from a different data source (Population Census 2011).

2.6.1 Different Cut-Off Points

In Table 14, I test the validity of the first stage estimates obtained around the 500 and 1,000 population thresholds. I use some placebo cut-off points around these two thresholds and run the specification used in column (4) of Table 11, i.e. first stage with GP-level baseline covariates and fixed effects. Estimates in Table 14 suggest that there is no such jump at any of these placebo thresholds.

2.6.2 Different Bandwidths

In my benchmark specification, I use an arbitrary bandwidth of +/-50 around the population thresholds. I test for the robustness of these results using different bandwidths. As shown in Appendix B.5, for all the outcome variables, the estimates are statistically indistinguishable across alternate bandwidths (+/-30, 40, 60 and 70). Specifically, for the 16-30 and 31-60 day margins, as the size of the bandwidth increases, point estimates shrink in magnitude. A plausible explanation for stronger effects within narrower bandwidths could be their localized nature. I compare compliers on both sides, i.e. areas where roads are provided even if they are just eligible and those where roads are not provided even if they are just ineligible. As the bandwidth shrinks, compliance on either side implies higher adherence to administrative rules in both control and treatment areas. It is possible that treatment effects of rural roads are witnessed more strongly in such villages.

2.6.3 Variation in Endogenous Regressor

In the benchmark specification, the endogenous regressor is the ratio of villages in a GP connected by a PMGSY road. I re-estimate the benchmark specification using the ratio of a GP's population connected by a PMGSY road as the endogenous regressor. This is relevant, for in addition to the ratio of connected villages in a GP, potential improvements in bureaucratic efficacy my also be a function of the proportion of people in the GP who can access roads. The results for all outcome variables, as shown in Figure 19 and Figure 20, are statistically indistinguishable from the benchmark estimates.

2.6.4 Different Data Source

Finally, following Aggarwal (2018), I expand the scope of the endogenous regressor. Using Census data from 2011, I use an indicator variable that equals 1 for *any* paved road in a village (PMGSY or no PMGSY). Apart from the endogenous regressor, everything else remains the same as in the benchmark specification. The specification with this new endogenous regressor uses approximately 13% fewer observations (approximately 825 observations) when compared to the benchmark case owing to concerns regarding data quality. While the Census was released in 2011, the reference date of data collection is 2009. When I compare the two variables, i.e. PMGSY coverage and paved road coverage by the end of 2009, for approximately 825 GPs, I find that PMGSY coverage is higher than paved road (PMGSY and non-PMGSY) coverage. I drop these problematic observations for this new endogenous regressor. For the purpose of comparison with the benchmark case, I look at estimates with and without these "bad" data points using the original endogenous regressor. As Figure 21 and Figure 22 suggest, the results are statistically indistinguishable for all outcome variables across the three specifications.[34]

2.7 Potential Mechanisms

Headline results suggest that construction of rural roads led to quicker beneficiary payments under MGNREGA. The original hypothesis was that construction of rural roads implies higher state capacity, which can be leveraged by on-field functionaries to quicken the payment process. On-field bureaucrats involved in Stage 1 are usually stationed in and around their assigned GPs and are frequently on the move. Hence, if rural roads lead to quicker payments due to easier on-field movement for these bureaucrats, one should expect treatment effects to be spread out evenly across all kinds of treated GPs.

[34]Results for >90 days omitted to avoid overcrowding in the graphs, although they remain robust.

However, as shown in Figure 23 and Figure 24, sub-sample analysis suggests that treatment effects vary with a treated GP's distance to its sub-district headquarters.[35] After dividing the sample into tertiles based on GPs' distance to their sub-district headquarters, I find that headline treatment effects are driven by GPs farther away from the headquarters. In fact, rural roads lead to slower payment processing in areas close to the latter.

Simple OLS regression of quality of PMGSY roads on the GP's distance to sub-district headquarters suggests that differential quality of roads does not explain this result (Appendix B.4). Another possible explanation may arise from Asher and Novosad (2020), who suggest that rural roads provide better access to neighboring labor markets, facilitating movement of rural labor out of agriculture. This effect can potentially dampen the demand for MGNREGA, thereby reducing officials' workload, which may lead to faster payments in the area. I, therefore, estimate the effect of road construction on total MGNREGA transactions for the entire sample, as well as for the three tertiles separately. For this purpose, I use a modified version of the benchmark specification, replacing district fixed effects with sub-district fixed effects. In Appendix B.7, I show that although imprecisely estimated, roads lead to an increase in MGNREGA transactions across all tertiles. The non-negative estimates are not entirely surprising. Given the nature of rural society in India, working age men are the likeliest candidates in a household to move out of villages for non-farm work (Asher and Novosad 2020). Nevertheless, given that each household is entitled to 100 days of work every year, women and older men can take up MGNREGA work. At the same time, if one were to take the positive point estimates seriously, a possible explanation is that the expectation of more efficient administration of the scheme may have encouraged more workers to demand MGNREGA work.

[35] Results in Figure 23 and Figure 24 are obtained by using a variation of the benchmark specification, where instead of district fixed effects, I control for sub-district fixed effects.

Hence, while I don't have definite causal evidence to explain this spatial heterogeneity in treatment effects, I offer a justification based on existing literature and some descriptive evidence. Asher et al. (2018) find that areas far away from local government headquarters witness lower provision of public goods due to higher costs of monitoring by higher officials. Unfortunately, there is no public data available on the frequency of random audits conducted by senior officials in the context of MGNREGA. However, consistent with Asher et al. (2018), across GPs in the control group in my sample, I observe that proportion of transactions processed within the 30-day interval decreases and the proportion of those processed beyond the 30-day interval increases in the distance to sub-district headquarters (Figure 25 and Figure 26). It is, therefore, likely that at least for GPs in tertiles 2 and 3, improved road connectivity enhances state capacity at two levels: operational for on-field functionaries and monitoring for officials at the sub-district level. For GPs in tertile 1, on the other hand, improvement in monitoring capacity may be less important.

It is relevant to note that in general, on-field bureaucrats are not incentivized by performance-linked incentives, but may be influenced by the fear of punishment if they are seen to be performing poorly by higher officials at the district/sub-district level. The latter, on the other hand, need to ensure smooth implementation of schemes like MGNREGA to climb up the administrative hierarchy. Hence, whereas bureaucrats involved with monitoring have the incentive to ensure higher efficiency across all areas under their jurisdiction, on-field bureaucrats may be motivated to work more efficiently in areas where higher officials are expected to focus more. Improved efficacy due to rural roads in tertiles 2 and 3 may, therefore, be a result of increased focus of higher officials who are now able to access those areas with greater ease. As a result, on-field bureaucrats may leverage better roads to focus in these areas. Admittedly, results from tertile-1 are puzzling, and while there may be some speculative explanations, it is better to explore them within the realm of a separate research question.[36]

2.8 Conclusion

Countries across the world deploy social safety nets and administer welfare schemes aimed at provision of public goods. However, across resource-constrained environments, last mile delivery of these schemes remains a challenge. Performance of the local bureaucracy is an important determinant of timely delivery of these services. Existing literature explores the role of incentives, diversity, and oversight and monitoring mechanisms as factors influencing bureaucratic performance. Nevertheless, across developing countries, factors related to state capacity such as inadequate salaries and lack of infrastructure can also impede their efficacy.

In this paper, I identify the impact of improved state capacity, as measured by newly built rural roads, on the performance of on-field bureaucrats in the context of a national public works scheme. During the period of my study, on-field officials associated with the latter had to commute frequently to ensure coordination in order to process payments for beneficiaries. Using village-level data on two national-level rural welfare schemes from four Indian states, I assess the causal effect of better road infrastructure on time taken to complete on-field administrative steps. I overcome the challenge of endogeneity of road placement by using an IV strategy inspired by the fuzzy RD design (Bhalotra and Clots-Figueras 2014). Specifically, I exploit the fact that under the said national rural road construction program, arbitrary population-based thresholds were implemented to prioritize road construction across villages. I find that, on average, construction of rural roads leads to a reduction in the average time taken to complete on-field steps. Depending on the outcome variable, the treatment effect is in the range of 47-60% of the control mean. On average, these effects imply quicker processing of annual payments

[36] For instance, owing to their proximity to sub-district headquarters, beneficiaries in tertile-1 may access nearby labor markets with greater ease (Asher and Novosad 2020). While demand for MGNREGA does not seem to reduce in these areas, it is possible that beneficiaries are not too concerned about delayed payments.

worth USD 2,193-2,705 in an area with total annual beneficiary payments worth USD 17,684. These effects are robust to several changes in the benchmark empirical specification and the usage of different data sources for the endogenous regressor.

Further, I find that treatment effects vary by distance to local government (sub-district) headquarters and are driven by areas farther away from the latter. This rules out the possibility of a general improvement in the performance of on-field bureaucrats due to enhanced state capacity. Evidence suggests that this spatial heterogeneity is not caused by lower take-up of the public works scheme due to out-migration caused by road construction. Neither is the quality of roads built in these areas any different from the rest of the sub-district. Hence, although suggestive, I offer an explanation based on enhanced monitoring capacity as a potential mechanism.

Existing literature (Asher et al. 2018) and suggestive evidence from MGNREGA show that high cost of transportation may impede adequate monitoring of government programs in areas far away from local government headquarters. Therefore, introduction of rural roads in such remote areas may lead to more frequent audits by officials at the sub-district level, who are motivated by the possibility of promotion and have an incentive to perform well across all areas within their jurisdiction. On-field bureaucrats, on the other hand, seldom receive any performance-linked incentives, but fear punishment if such senior officials deem their performance as unsatisfactory. Given these facts, due to additional fear of negative incentives, on-field officials in remote villages may be leveraging newly constructed roads more effectively. Generally speaking, results from this paper suggest that in the absence of appropriate incentives, enhanced state capacity may not result in improved bureaucratic performance by itself. This conclusion may have important implications for policymakers involved in augmenting bureaucratic capacity through capital investments.

2.9 Chapter 2: Tables

DATA	LEVEL	DATA SOURCE
PMGSY Roads	Habitation	Asher and Novosad (2020)
MGNREGA Payment Delays	Gram Panchayat	MGNREGA Website
Baseline Controls	Village	SHRUG Database
GP Village Mapping	Gram Panchayat	Local Government Directory
GP-Sub-District Distance	Village	Population Census 2011
All Paved Roads (2011)	Village	Population Census 2011

Table 8: Data Sources

Table 9: Naive OLS Regression Analysis (Outcome: Ratio of Transactions Processed during FY 2014-15)

	(1) Processed Within 15 Days	(2) Processed Within 16-30 Days	(3) Processed Within 31-60 Days	(4) Processed Within 61-90 Days	(5) Processed After 90 Days
Ratio of PMGSY-Connected Villages	-0.006*	-0.005	0.004	0.002	0.004
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Control Mean (No PMGSY Roads)	.34	.27	.21	.09	.09
N	28811	28811	28811	28811	28811

Robust Standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

All specifications control for number of villages in GP and district fixed effects. All specifications also control for baseline GP-level variables (2001), which include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 10: Naive OLS Regression Analysis (Outcome: Ratio of Amount Processed during FY 2014-15)

	(1) Processed Within 15 Days	(2) Processed Within 16-30 Days	(3) Processed Within 31-60 Days	(4) Processed Within 61-90 Days	(5) Processed After 90 Days
Ratio of PMGSY-Connected Villages	-0.005	-0.005	0.004	0.002	0.003
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Control Mean (No PMGSY Roads)	.34	.27	.21	.09	.09
N	28797	28797	28797	28797	28797

Robust Standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

All specifications control for number of villages in GP and district fixed effects. All specifications also control for baseline GP-level variables (2001), which include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 11: First Stage: Ratio of Connected Villages in GP Instrumented by Number of Just Eligible Villages Within Bandwidth (+/-50)

	(1)	(2)	(3)	(4)
	No Controls	GP-Level Controls	Fixed Effects	Controls +Fixed Effects
No. of Just Eligible Villages (BW=50)	0.030***	0.045***	0.055***	0.058***
	(0.007)	(0.006)	(0.007)	(0.007)
F-Stat	20.2	51.32	66.05	78.07
N	6542	6445	6542	6445

Robust standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

Sample includes Rajasthan, Gujarat, Odisha and Madhya Pradesh. Outcome variable is the overall ratio of villages in the GP connected by roads under PMGSY by March 31, 2014. Controls include number of villages. Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights. Fixed effects imply fixed effects for districts and composition of number of villages within the bandwidth (on either side).

Table 12: 2SLS Estimates with Number of Just Eligible Villages Within Bandwidth as IV (Bandwidth=50)

	(1) Ratio of Transactions Within 15 days	(2) Ratio of Transactions 16-30 days	(3) Ratio of Transactions 31-60 days	(4) Ratio of Transactions 61-90 days	(5) Ratio of Transactions >90 days
PMGSY Village Coverage	0.022	0.128*	-0.124*	0.008	-0.035
	(0.077)	(0.070)	(0.071)	(0.048)	(0.065)
Control Mean	.27	.27	.24	.1	.12
Kleibergen-Paap F-Stat	78.07	78.07	78.07	78.07	78.07
Cragg-Donald F-Stat	75.40	75.40	75.40	75.40	75.40
Effective F-Stat	78.07	78.07	78.07	78.07	78.07
tF-Adjusted 95% CI	(-0.13, 0.17)	(-0.01, 0.27)	(-0.27, 0.01)	(-0.09, 0.10)	(-0.16, 0.09)
N	6445	6445	6445	6445	6445

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Sample includes Rajasthan, Gujarat, Odisha and Madhya Pradesh. Outcome variable is the percentage of transactions processed within given time periods. Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking and educational facilities, night lights, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 13: 2SLS Estimates with Number of Just Eligible Villages Within Bandwidth as IV (Bandwidth=50)

	(1) Ratio of Amount Processed Within 15 days	(2) Ratio of Amount Processed 16-30 days	(3) Ratio of Amount Processed 31-60 days	(4) Ratio of Amount Processed 61-90 days	(5) Ratio of Amount Processed >90 days
PMGSY Village Coverage	0.024	0.153**	-0.144**	0.001	-0.036
	(0.078)	(0.071)	(0.071)	(0.048)	(0.065)
Control Mean	.27	.27	.24	.1	.12
Kleibergen-Paap F-Stat	78.32	78.32	78.32	78.32	78.32
Cragg-Donald F-Stat	75.64	75.64	75.64	75.64	75.64
Effective F-Stat	78.32	78.32	78.32	78.32	78.32
tF-Adjusted 95% CI	(-0.13, 0.18)	(0.01, 0.29)	(-0.29, -0.00)	(-0.10, 0.10)	(-0.17, 0.09)
N	6444	6444	6444	6444	6444

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Sample includes Rajasthan, Gujarat, Odisha and Madhya Pradesh. Outcome variable is the percentage of amount processed within given time periods. Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking and educational facilities, night lights, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 14: First Stage: Robustness Checks with Different Cut-Off Values (BW=50)

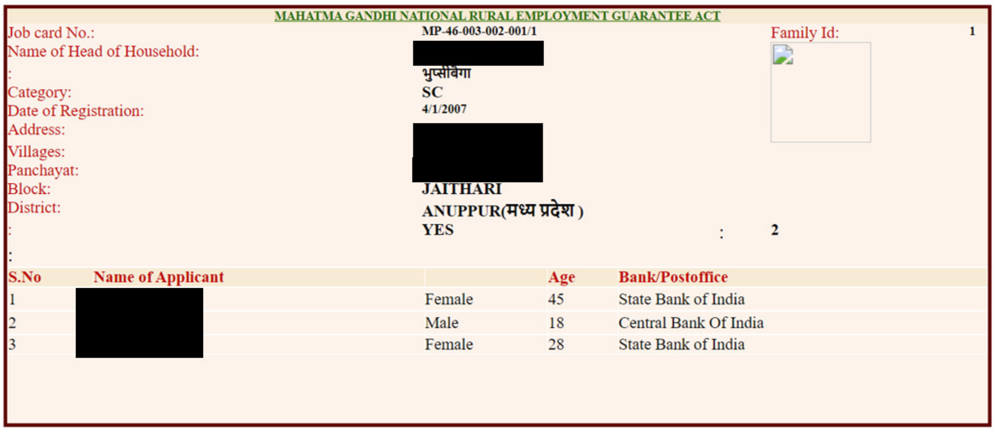
	(1)	(2)	(3)	(4)
No. of Just Eligible Villages	-0.000	-0.003	-0.000	-0.006
	(0.006)	(0.006)	(0.007)	(0.007)
F-Stat	0	.32	0	.76
Thresholds	400, 900	450, 950	550, 1050	600, 1100
N	7395	7177	6970	6711

Robust Standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Outcome variable is the ratio of villages covered under PMGSY by March 31, 2014. Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

2.10 Chapter 2: Figures



MAHATMA GANDHI NATIONAL RURAL EMPLOYMENT GUARANTEE ACT

Job card No.:	MP-46-003-002-001/1	Family Id:	1
Name of Head of Household:	[redacted]		
:	भुप्सीबिगा		
Category:	SC		
Date of Registration:	4/1/2007		
Address:	[redacted]		
Villages:			
Panchayat:			
Block:	JAITHARI		
District:	ANUPPUR(मध्य प्रदेश)		
:	YES	:	2
:			

S.No	Name of Applicant		Age	Bank/Postoffice
1	[redacted]	Female	45	State Bank of India
2		Male	18	Central Bank Of India
3		Female	28	State Bank of India

Figure 10: Sample MGNREGA Job Card

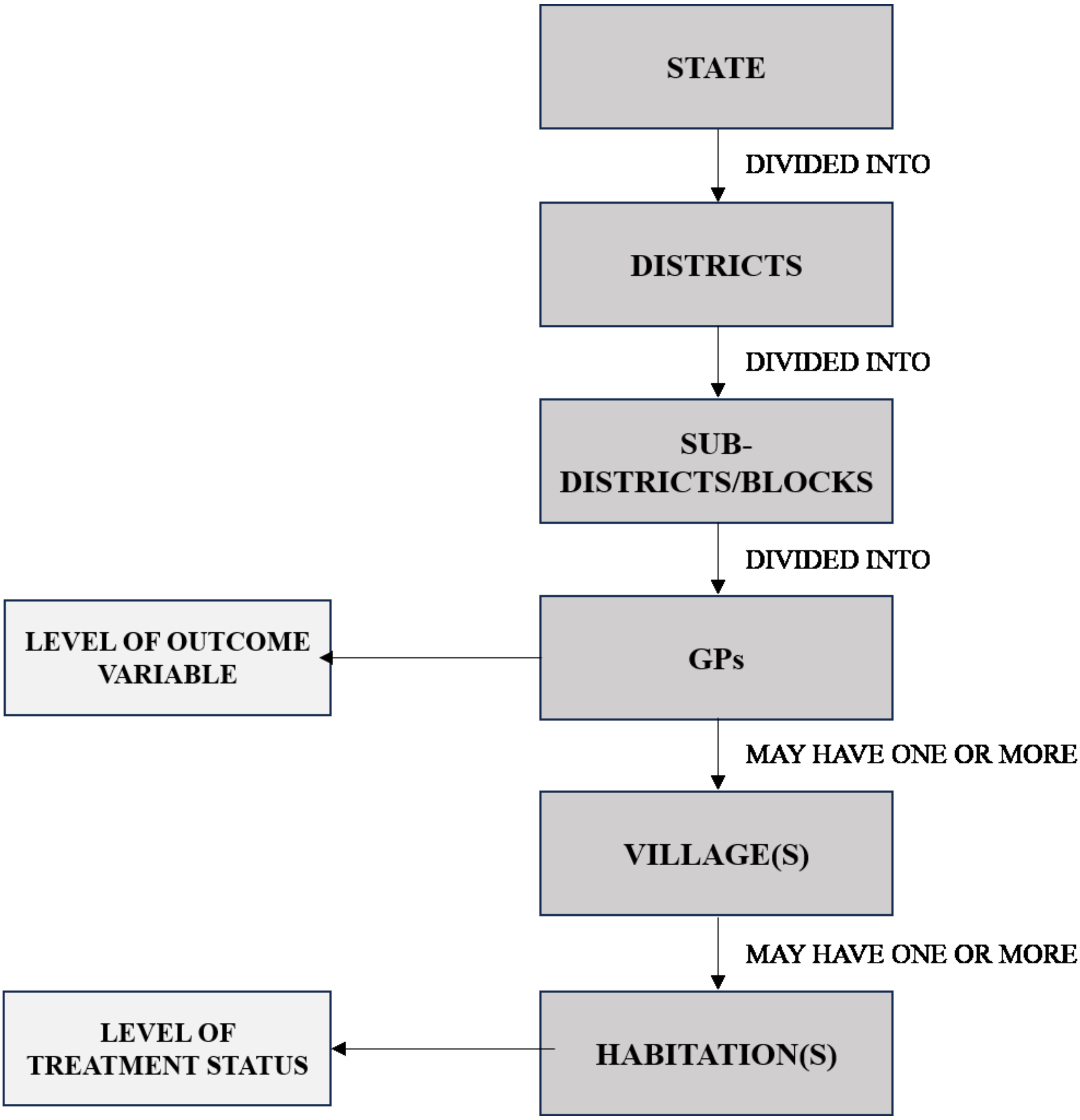


Figure 11: Administrative Hierarchy in India

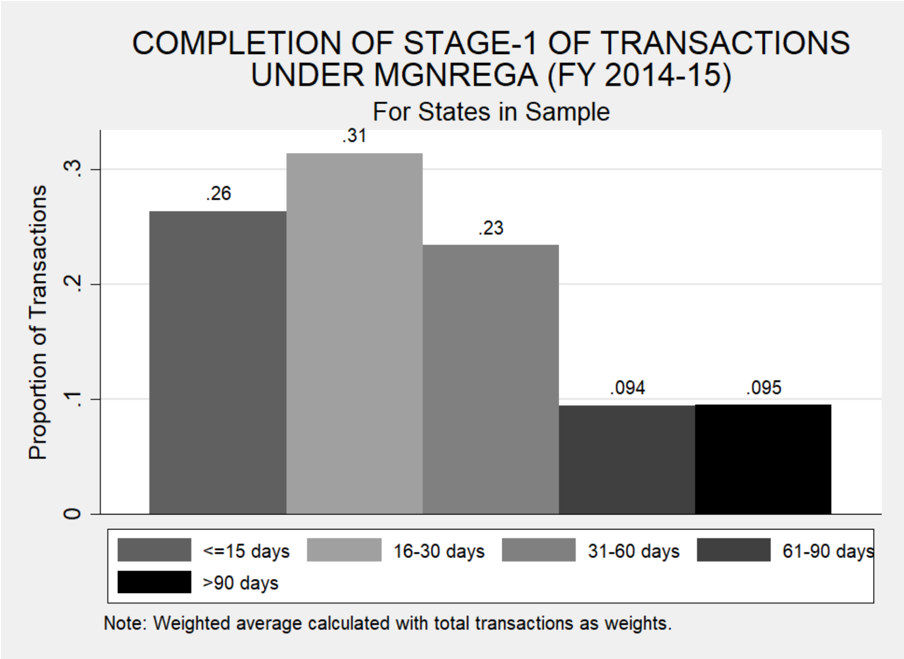


Figure 12: Time taken to complete Stage-1 during FY 2014-15

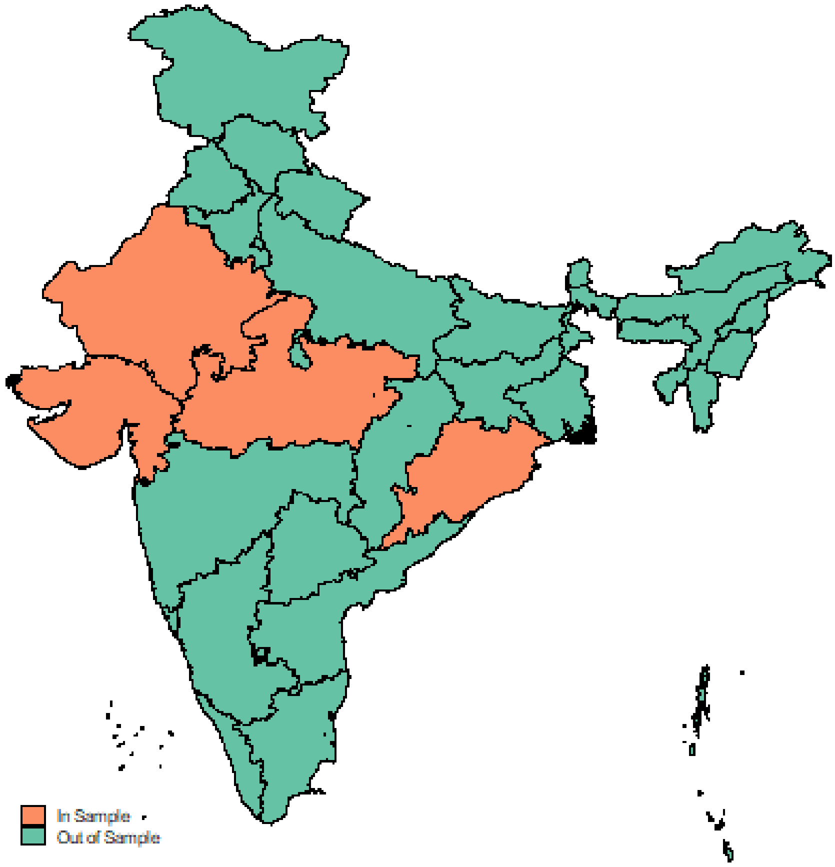


Figure 13: Sample States: Map of India

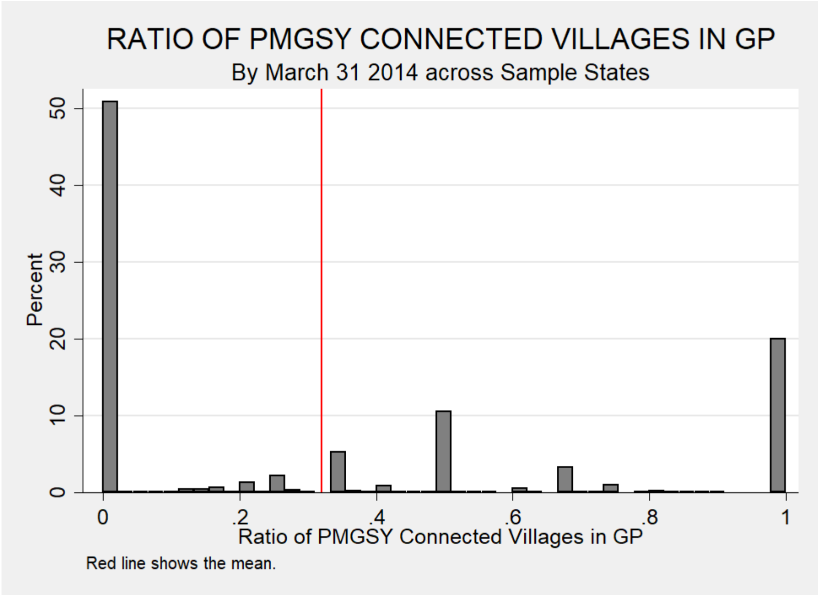


Figure 14: Histogram: Ratio of Villages connected by PMGSY road in GP

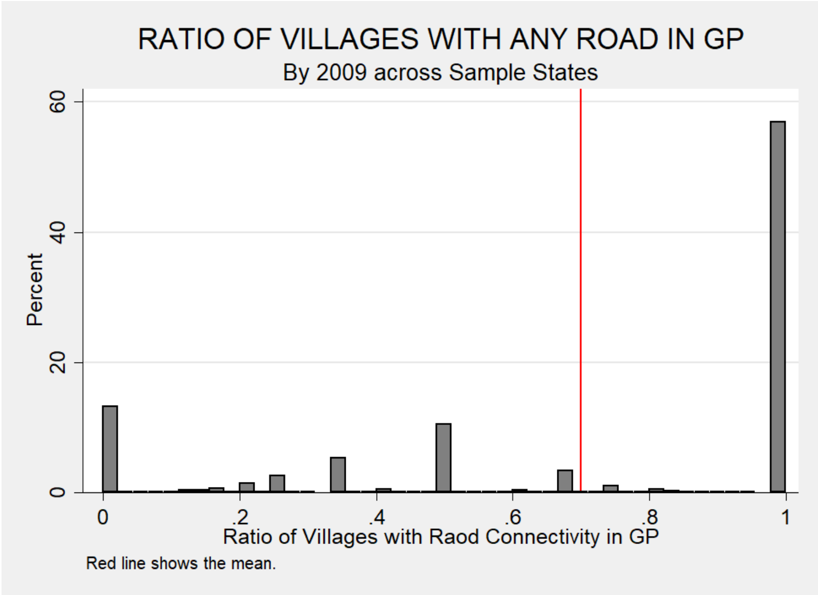


Figure 15: Histogram: Ratio of Villages connected by any road in GP

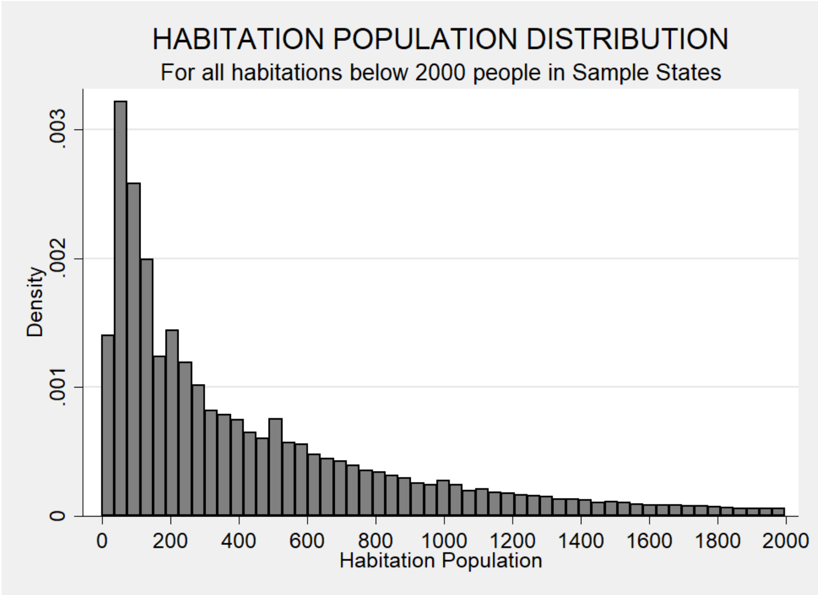


Figure 16: Checking for Manipulation: Distribution of Habitation Population

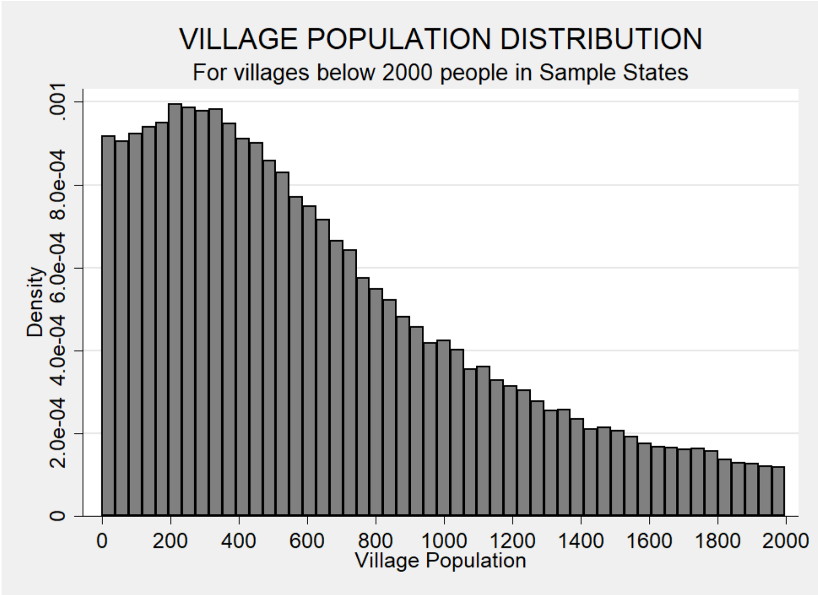


Figure 17: Checking for Manipulation: Distribution of Village Population

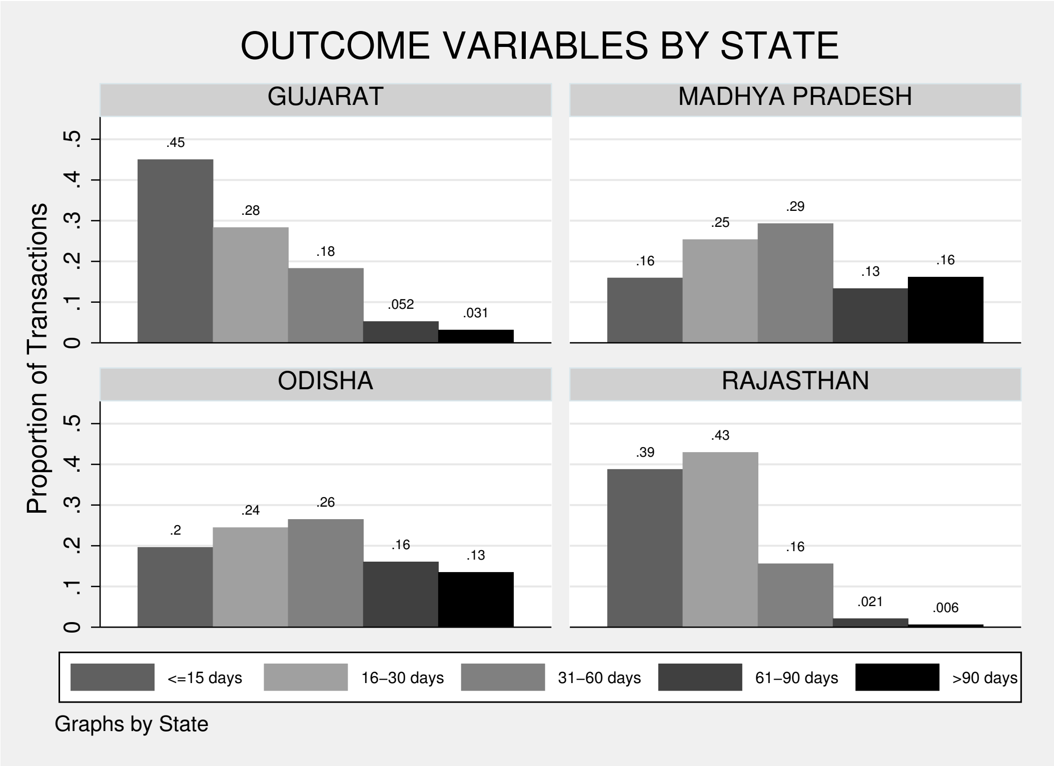


Figure 18: MGNREGA Payment Processing Time (Stage-1) by State

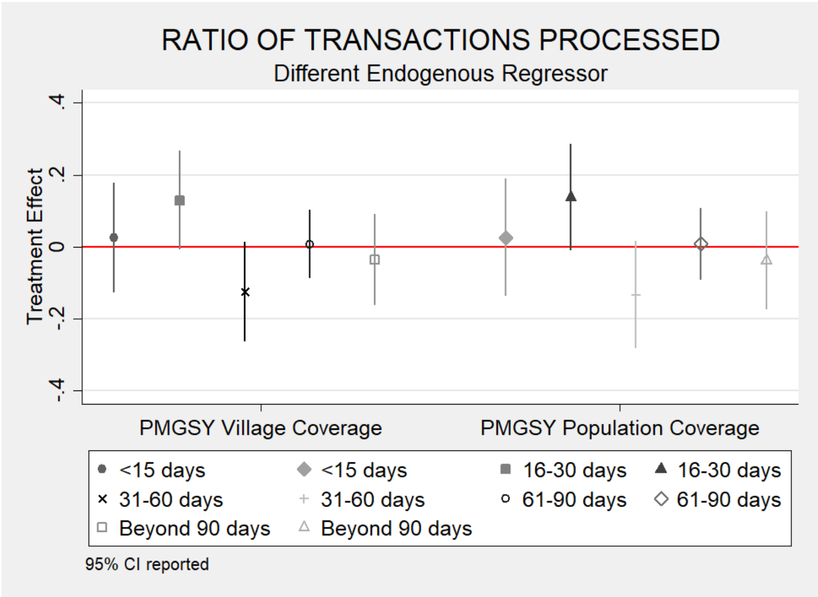


Figure 19: Robustness Check (transactions): PMGSY Population Coverage as Endogenous Regressor

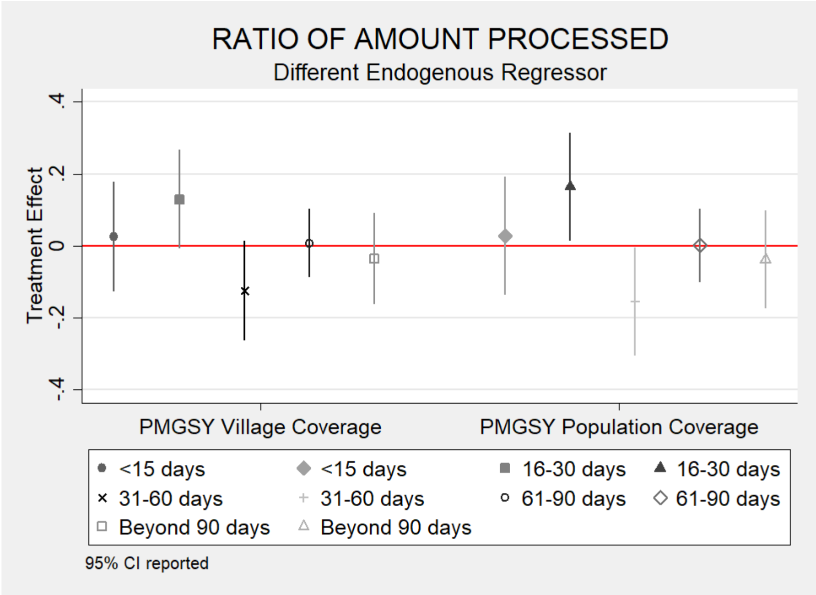


Figure 20: Robustness Check (amount): PMGSY Population Coverage as Endogenous Regressor

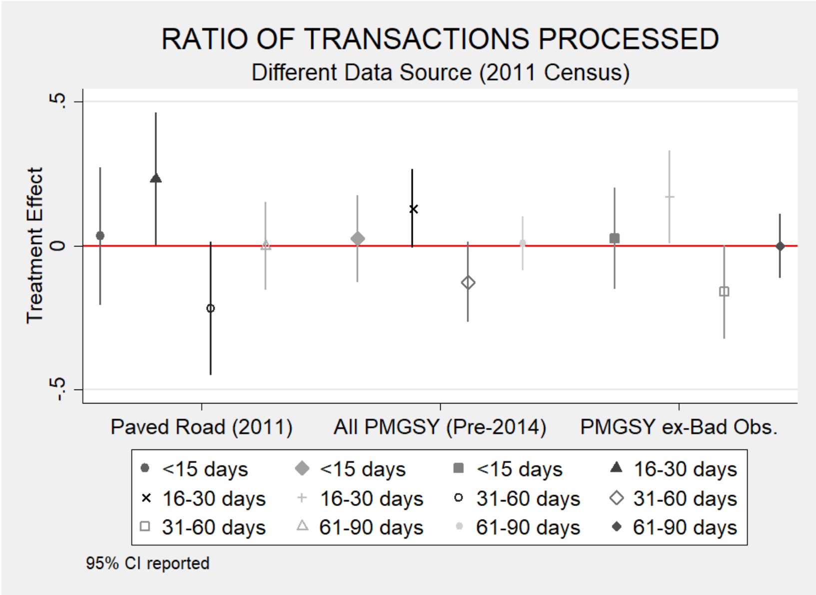


Figure 21: Robustness Check (transactions): Any Paved Road Coverage as Endogenous Regressor

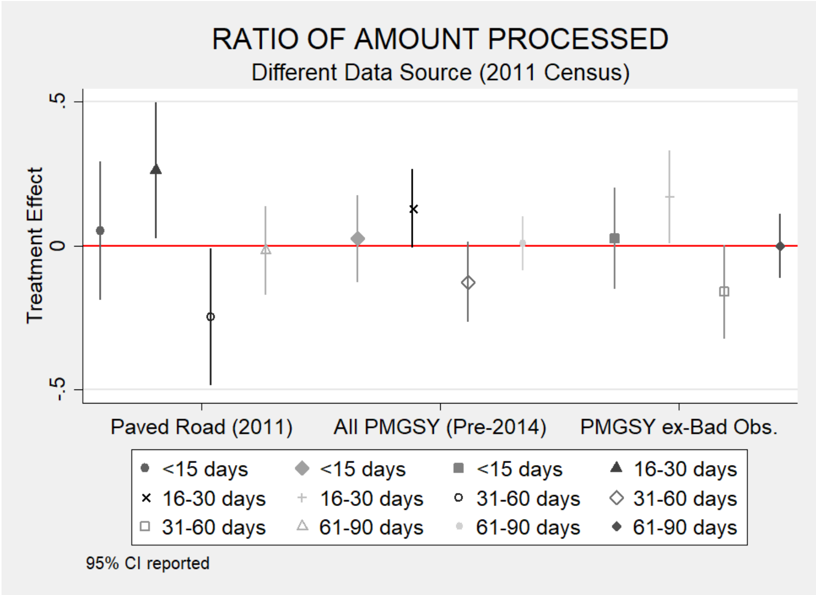
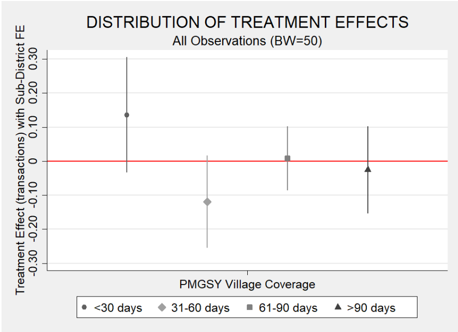
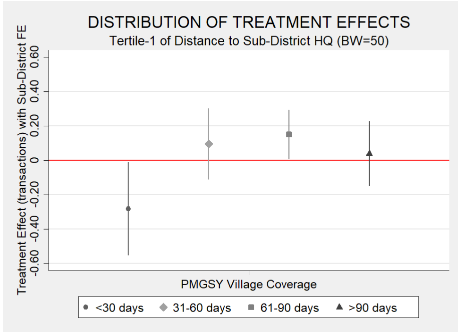


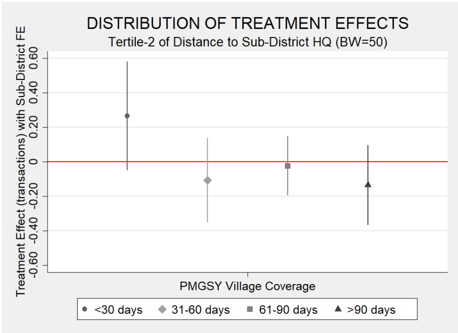
Figure 22: Robustness Check (amount): Any Paved Road Coverage as Endogenous Regressor



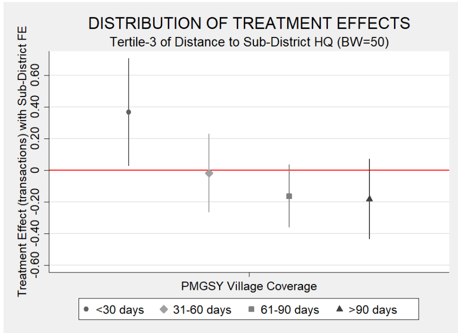
(a) All Observations



(b) Tertile-1

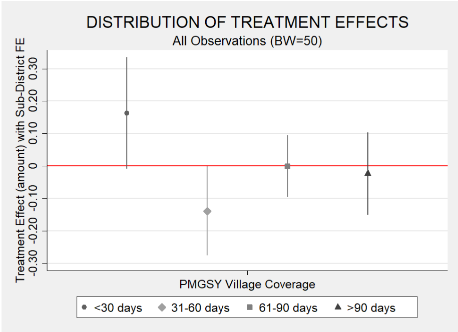


(c) Tertile-2

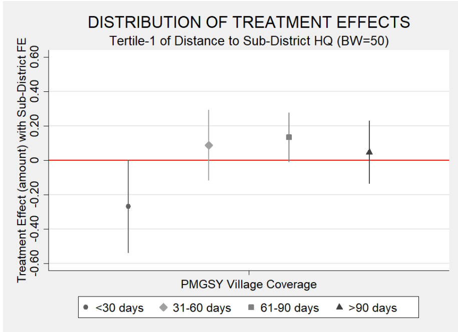


(d) Tertile-3

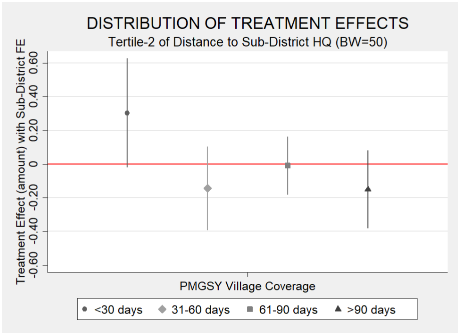
Figure 23: Heterogeneity Analysis of Treatment Effects (transactions): By Distance to Sub-District Headquarters



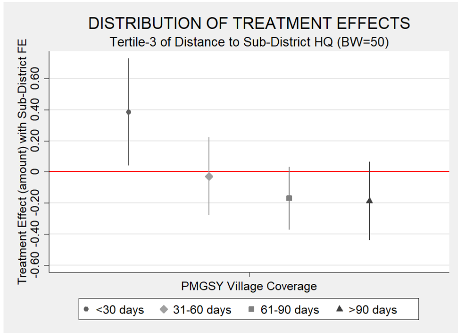
(a) All Observations



(b) Tertile-1



(c) Tertile-2



(d) Tertile-3

Figure 24: Heterogeneity Analysis of Treatment Effects (amount): By Distance to Sub-District Headquarters

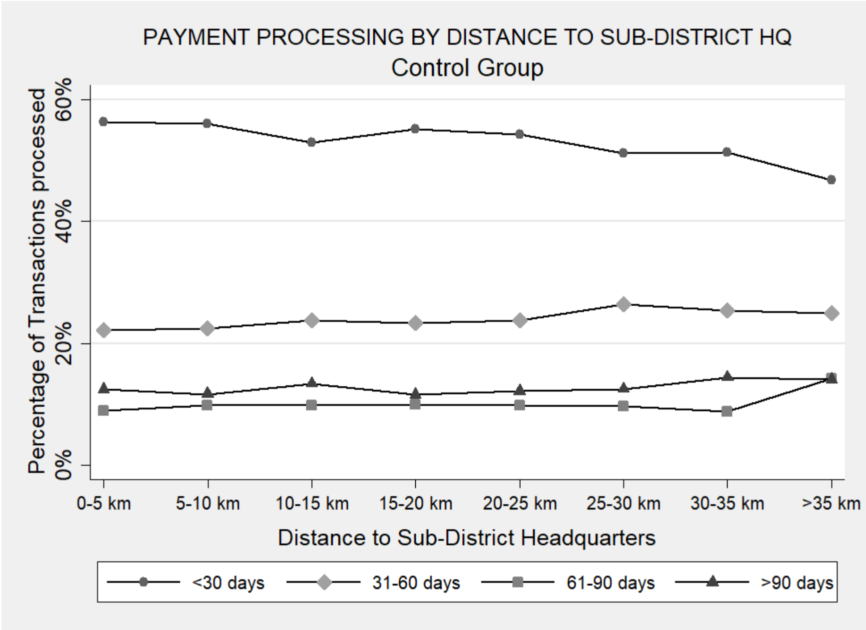


Figure 25: Time taken to process MGNREGA transactions by distance to Sub-District HQ for Control Group

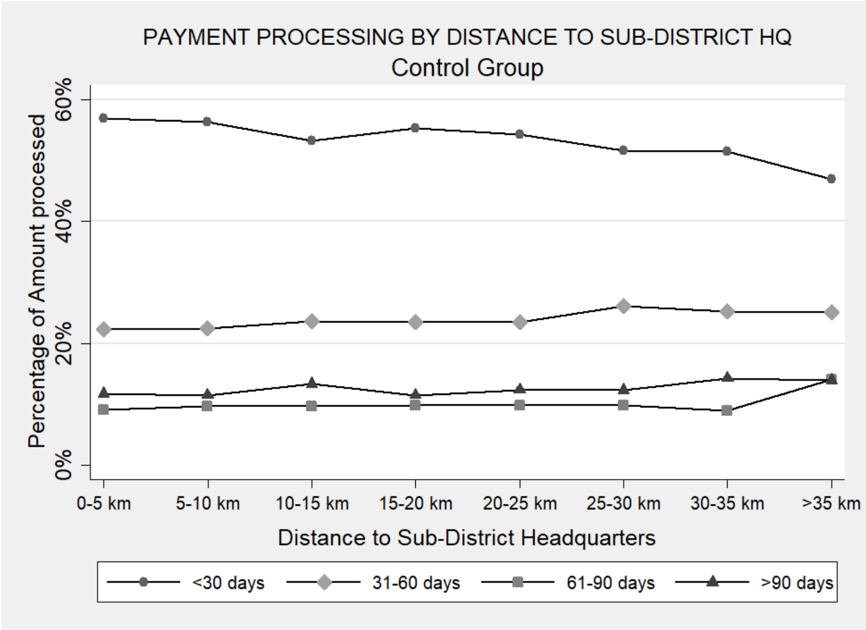


Figure 26: Time taken to process MGNREGA amount by distance to Sub-District HQ for Control Group

3 CHAPTER 3: Party Size and Candidate Wealth: Evidence from India

Why are some elected legislators wealthier than others? In this paper, I examine whether the size a of a political party determines the economic background of its candidates. Unlike parties with a national footprint, smaller regional parties in a multi-party system may have limited avenues for resource mobilization. I hypothesize that such parties overcome this disadvantage by fielding wealthier candidates. Using data from state elections across India during 2005-2016, I estimate whether challengers fielded by regional parties in narrowly lost seats are wealthier than those fielded by national parties. RD estimates suggest that challengers fielded by regional parties may, on average, be 44-103% wealthier than those fielded by national parties. These results have important implications for understanding electoral finance in low- and middle-income countries like India.

3.1 Introduction

What explains spatial differences in the affluence of political representatives? This question is important in light of evidence that a political representative's socio-economic profile impacts outcomes related to governance (Prakash et al., 2019). In this paper, I explore one possible explanation related to the size of political parties. In a multi-party system, political parties of different sizes exist. Whereas some hold political influence across different parts of the country, others may be limited to specific political geographies. I argue that when compared to those in the former category, parties in the latter category have limited avenues to raise funds. I hypothesize that as a result of this difference, regional parties rely more on wealthier candidates. Using state-level election data during the period 2005-2016, I test whether regional parties that barely lose elections to national parties during election cycle t field wealthier challengers during election cycle $t + 1$, when compared to national party challengers.

As per the Election Commission of India (2023), there are over 2,500 political parties across

India. Of these, six are recognized as "national parties", whereas over 70 are considered as state-level parties, and over 2,500 are unrecognized parties. Among the 6 national parties, only two-the Indian National Congress (INC) and the Bharatiya Janata Party (BJP)- have substantial presence across a large part of India. Since India's independence in 1947, the Indian Prime Minister has belonged to either of these parties for over 90% of the duration. The rest of the parties are mostly limited to a specific state, whereas many have presence within specific regions of a specific state.

Treating the INC and the BJP as national parties, and the rest as regional parties, I test whether the latter field wealthier candidates than the former. There are at least two reasons to believe that this may be the case. First, unlike most regional parties, the INC and the BJP have been in power at at federal level. It is, therefore, possible that they are better connected with potential donors owing to their past, as well as the possibility of being in power in the future. At the same time, unlike regional parties, these two parties head state governments of different states, which may provide them with additional avenues to mobilize resources. Regional parties, on the other hand, have fewer avenues to raise resources in a centralized manner, and may, therefore, rely on wealthier candidates for their campaigns.

Suggestive evidence points in this direction. Vaishnav (2011) points out, "Because parties are in need of finances to fund activities such as campaigning, voter mobilization and vote buying, they must strategically select candidates who will not be a drain on finite party coffers." Verniers and Jaffrelot (2020) suggest that in the context of national elections, during 2004-2019, while average individual wealth levels for all elected candidates jumped to six times the original level, the same for elected candidates from regional parties rose to thirteen times the baseline value. In this paper, I try to find causal evidence regarding the impact of the politico-geographical reach of a political party on its choice of candidates along the dimension of candidate wealth.

A naive regression of a candidate's wealth on the nature of their political party (national or regional) may face challenges related to endogeneity. I overcome this challenge by looking at seats where national and regional parties are involved in close contests. In such a seat, the victory of a national party candidate against a regional party challenger can be treated as exogenous. Similarly, the victory of a regional party candidate against a national party challenger can also be considered as exogenous. In my sample, I focus on seats that are decided by a narrow margin during election cycle e_t. Following a close election regression discontinuity setup, I compare the wealth of regional and national party challengers during election cycle e_{t+1}.

In the benchmark specification, I use candidates' self-declared assets as well as net worth (self-declared assets minus self-declared liabilities) as the primary outcome variables. I normalize these variables by dividing them by the number of electors in the constituency. Depending on the outcome variable, estimates suggest that candidates fielded by regional parties in seats lost by narrow margins are 44-103% wealthier than candidates fielded by national parties in similar seats. Results are robust to a battery of empirical checks.

The nature of the results, however, is localized. I focus on those seats that witness contests between regional and nationl parties in both election cycles e_t as well as e_{t+1}. Next, in the context of this paper, a treated seat is one where a regional party challenges an incumbent national party during election cycle e_{t+1} after being narrowly beaten during election cycle e_t. Control seats, on the other hand, are those where a national party challenges an incumbent regional party during election cycle e_{t+1} after being narrowly beaten during election cycle e_t. Hence, the nature of treatment is bundled, i.e. a regional party that acts as challenger and a national party that is incumbent. I partially address the last concern by controlling for characteristics of the candidates during e_{t+1}. The results remain robust to the inclusion of these covariates.

The rest of the paper is organized as follows. In Section 3.2, I discuss the existing literature and this paper's contribution to the same. In Section 3.3, I lay down the context for this study. In Section 3.4, I explain the empirical strategy, while in Section 3.5, I provide details regarding the data and its sources. In Section 3.6, I establish the validity of the main empirical strategy. In Section 3.7, I show the main results, and in Section 3.8, I show results from robustness checks. Finally, in Section 3.9, I discuss the said results, and provide concluding remarks.

3.2 Literature Review

The paper contributes to existing literature related to challenger and incumbent spending. The importance of money in elections was noted by Abramowitz (1988). Specifically, that paper focuses on the ability of challengers to raise funds for their campaigns. In the context of the United States of America, Abramowitz (1991) noted the declining ability of House challengers to raise campaign funds. At the same time, there is evidence that challenger spending impacts the challenger's fate significantly. While Jacobson (1990) suggests that challenger spending is more effective than incumbent spending, Green and Krasno (1988) suggest that the role of campaign financing is a complement to the candidate's quality.

In the Indian context, Gowda and Sridharan (2012) cite the National Election Audit (1999) conducted by the Center for the Study of Developing Societies, which suggests that "to be competitive, candidates need to spend at least a certain minimum amount", although above certain thresholds, the correlation between spending and vote shares is lost. However, given data constraints related to election spending in India, the paper will focus on challengers' assets (and net worth) instead of electoral spending. In that regard, Vaishnav (2011) points out that political parties are attracted towards criminal politicians because they possess the ability to self-finance, and they do not act as a drain on their party's resources. Chauchard et al. (2019)

suggest that while voters see candidate wealth with suspicion related to corruption and political violence, there are mechanisms that justify victories of "wealth accumulators" despite these negative reactions. However, to my best knowledge, at least in the Indian context, this is among the first few papers that examines the causal impact of a political party's politico-geographic reach on its candidate selection along the dimension of wealth.

3.3 Context

The Indian legislative system has three tiers. Voters elect representatives to each of these tiers, i.e. Members of Parliament to the Parliament of India (national), Members of Legislative Assembly to their respective state assemblies, and representatives to their areas' local legislative bodies. Each of these elections follows the "first past the post" system. In any given election, a candidate who secures the highest number of votes in a constituency is elected to the legislature from that constituency. In this paper, I will focus on elections to state assemblies. I make this choice (over the the other two types of elections) for multiple reasons.

When compared to national elections, the contest between regional and national parties in state elections is much stiffer. For instance, as of March 2024, almost a dozen Indian states are governed by Chief Ministers from regional parties. At the same time, for almost 90% of India's 76-year old post-independence history, a member belonging to either of the two national parties-Indian National Congress (INC) and Bharatiya Janata Party (BJP)-has been the Prime Minister.[37] In addition, elections to the national parliament are held once every 5 years, whereas there are 31 states (and union territories) where elections are held once every 5 years. These state elections take place in a staggered manner, such that there is at least one state assem-

[37]Technically, as per the Election Commission of India, as of 2024, there are 6 national parties, i.e. Bahujan Samaj Party, Bharatiya Janata Party, National People's Party, Communist Party of India (Marxist), Indian National Congress and Aam Aadmi Party. For the purpose of this project, I am considering only the INC and BJP because, unlike the other national parties, they are present across several states, have formed local governments across many of these states and have been the focal point of almost all federal governments in India's history so far.

bly ever year. Hence, these elections provide a much larger number of observations to detect effects. Finally, when compared to local body elections, data availability for state elections is much more robust.

As explained in the subsequent section, my empirical strategy hinges on the idea that constituencies with stiff competition between a regional and national party at time t will be assessed for outcomes at time $t+1$. It is to be noted that in 2008, electoral boundaries were redrawn across states. This complicates the empirical strategy. For instance, assembly elections were held to the state of Uttar Pradesh in 2002, 2007, 2012 and 2017. While constituencies in the 2002 election are comparable with those in 2007, the same is not true of constituencies in 2007 and 2012. Similarly, in the state of Kerala, elections were held in 2001, 2006, 2011 and 2016. As in the case of Uttar Pradesh, constituencies in 2001 are the same as those in 2006, and those in 2011 are the same as those in 2016. However, I cannot use 2006 as t and 2011 as $t+1$, because the configuration of constituencies changed during the intervening period. In addition, candidate affidavits, which contain information about their finances, criminality and education became compulsory only beginning 2003-04. That data is available only beginning 2004-05 and has been collated until 2017. Finally, there are states where the contest is restricted to the two national parties.

Considering the said factors in mind, as shown in Appendix C.1.1, constituencies from 18 states/Union Territories are considered for this study. This amounts to more than 50% of the total states/Union Territories where elections to state assembly elections are held every 5 years. Of these, I use a single set of "t" and "t+1" for 10 states, and two sets of "t" and "t+1"for the remaining 8 states.

3.4 Empirical Strategy

A naive OLS regression aimed at the identification of the impact of the size of a political party on the wealth of its nominated candidates is shown below (Equation (8)). As per Equation (8), the wealth of candidate i nominated by political party p in constituency c in year t is regressed on whether the party has a national footprint (or not) in year t, candidate-level controls ($X'_{i,p,c,t}$), party-level controls ($Z'_{p,t}$), constituency-level controls ($J'_{c,t}$), and time fixed effects (δ_t). In this case, β is the coefficient of interest. A positive β will imply that all else equal, candidates belonging to regional parties are, on average, wealthier than those belonging to national parties.

$$CandidateWealth_{i,p,c,t} = \alpha + \beta RegionalParty_{p,t} + X'_{i,p,c,t}\lambda + Z'_{p,t}\eta + J'_{c,t}\kappa + \delta_t + \epsilon_{i,p,t} \quad (8)$$

However, this specification may suffer from omitted variable bias. For instance, candidate wealth may be influenced by unobservables such as a constituency's political culture. There may also be concerns regarding reverse causality. The primary identification strategy in this paper will, therefore, be a sharp regression discontinuity design (Lee, 2008; Imbens and Lemieux, 2008).

Focusing on close contests between national and regional party candidates, I will estimate the impact of the election of a national party candidate in a constituency during an election cycle e_t on the wealth of the challenger candidate fielded by the competing regional party during the subsequent election cycle e_{t+1}. The counterfactual to these treated constituencies will be those where during e_t, a regional party candidate narrowly wins, and is challenged by a national party during e_{t+1}. The methodology is based on existing literature that uses close election RD designs (among others, Lee et al., 2004 and Prakash et al., 2019).

The underlying assumption behind this setup is that at the end of election cycle e_t, constituencies where a national party defeats a regional party by a narrow margin (let's call these type "A" constituencies) are, on average, similar to constituencies where a regional party defeats a national party by a narrow margin (call these type "B" constituencies) along all dimensions except that incumbents in type "A" constituencies belong to national parties, and those in type "B" belong to regional parties. During election cycle e_{t+1}, a candidate from a regional party will challenge the national party incumbent in type "A" constituencies, and a candidate from a national party will challenge the regional party incumbent in type "B" constituencies. A comparison of the wealth of these challenger candidates belonging to regional (treated) and national (control) political parties will then give us the causal effect of a regional party's challenge against an incumbent national party on the wealth of the former's candidate. The empirical strategy can be formalized as follows (Equation (9)).

$$CandidateWealth_{i,p,c,s,t+1} = \alpha + \beta RegionalPartyChallenger_{c,t+1} + f(Margin_{c,s,t}) + J'_{c,s,t}\kappa + \theta_s + \delta_t + \epsilon_{i,p,c,s,t+1} \tag{9}$$

As mentioned above, I will focus on close contests between national party and regional party candidates. The running variable in this case will be the difference between the national party candidate's vote share and the regional party's vote share during election cycle e_t. Therefore, the cut-off point will be zero. The final sample will contain constituencies that were won with a very narrow margin on either side of the cut-off, i.e. constituencies where national parties just won against regional parties (treated) and those where national parties just lost against regional parties (control) during election cycle e_t. In the benchmark specification, I will use a bandwidth of 5 percentage points. In subsequent sections, I test the robustness of the main results by varying the bandwidth.

The outcome variables ($CandidateWealth_{i,p,c,s,t+1}$) will represent the wealth of the challenger candidate i belonging to party p in constituency c in state s during election cycle e_{t+1}. For the main results, I will use two different outcomes: a challenger's total assets per elector in the constituency ($\frac{Assets}{Number of Electors}$) and their net wealth per elector ($\frac{Assets-Liabilities}{Number of Electors}$) in the constituency. I use the normalized values of assets and net worth instead of aggregate values because of the vast differences in constituency sizes across the country. For instance, across all state assembly elections held in 2016, as per the SHRUG database, the range of number of electors in a constituency is 10,747 to 60,2472. The same for elections held in 2017 is 12,764 to 8,65,650. Similar variations exist within states as well.

The variable $RegionalPartyChallenger_{c,t+1}$ is a binary variable (treatment) that equals 1 if during election cycle e_{t+1} in constituency c, a regional party candidate challenges a national party incumbent. In other words, $RegionalPartyChallenger_{c,t+1}$ equals one if during e_t, in constituency c, a national party narrowly won the election against a regional party candidate. Following Cattaneo et al., 2019, the expression $f(Margin_{c,s,t})$ represents a local linear polynomial in distance to the cut-off point, i.e. margin of victory for the national party during election e_t. $J_{c,s,t}$ is a set of constituency-level controls. θ_s and δ_t represent state and time fixed effects respectively. Finally, $\epsilon_{i,p,c,s,t+1}$ is the error term. The coefficient of interest is β. A positive β implies that regional party challengers are wealthier than national party challengers. Given that the winning candidate from time t may not necessarily be fielded again at time *t+1*, this empirical strategy focuses on incumbency at the party level, and not at the individual level.

3.5 Data

- **Outcomes:** As per a 2003 Supreme Court of India order, all candidates appearing in elections are required to disclose information regarding their financial, educational, and criminal background by means of a sworn affidavit submitted to India's Election Commission. This publicly available data has been compiled by the Association for Democratic

Reforms, an India-based think-tank, which was further collated by Prakash et al. (2019). I obtain this data from the Socioeconomic High-resolution Rural-Urban Geographic Dataset for India (SHRUG; see Asher et al., 2021 for details). The dataset contains each candidate's self-declared assets and liabilities. I use this data to create two different outcome variables- value of candidate's assets per elector in their constituency, and candidate's net worth (assets minus liabilities) per elector.

- **Explanatory Variables:** For all election cycles in the sample, I obtain data on each candidate's vote share (running variable) from Jensenius and Verniers (2017). This data is also a part of the SHRUG. The dataset also contains other information regarding the constituency: turnout, number of electors, reservation status (whether it is a general constituency or reserved for historically disadvantaged groups), number of candidates and political parties in contention, etc. I use some of these variables from election cycle e_t as controls (see Appendix C.1.2). Finally, I obtain constituency-level data on geographical and economic covariates from SHRUG: elevation, ruggedness and employment levels (1998).[38]

3.6 Validity of the RD Design

A valid sharp RD setup should satisfy two conditions. First, within the bandwidth around the cut-off point, both treated and control constituencies should be balanced on all explanatory variables except treatment status. Second, we should not see any bunching of observations around the cut-off point. I shall test each of these assumptions.

In order to confirm the validity of the RD design, I first run the McCrary density test to check if there is any manipulation around the cut-off, i.e. where the running variable equals zero. The t-statistic for this test turns out to be -1.44, which makes the coefficient statistically

[38] The authors of SHRUG obtain raw data on employment from the Economic Census of 1998, and map it to assembly constituencies.

insignificant at conventional levels (see Figure 27).

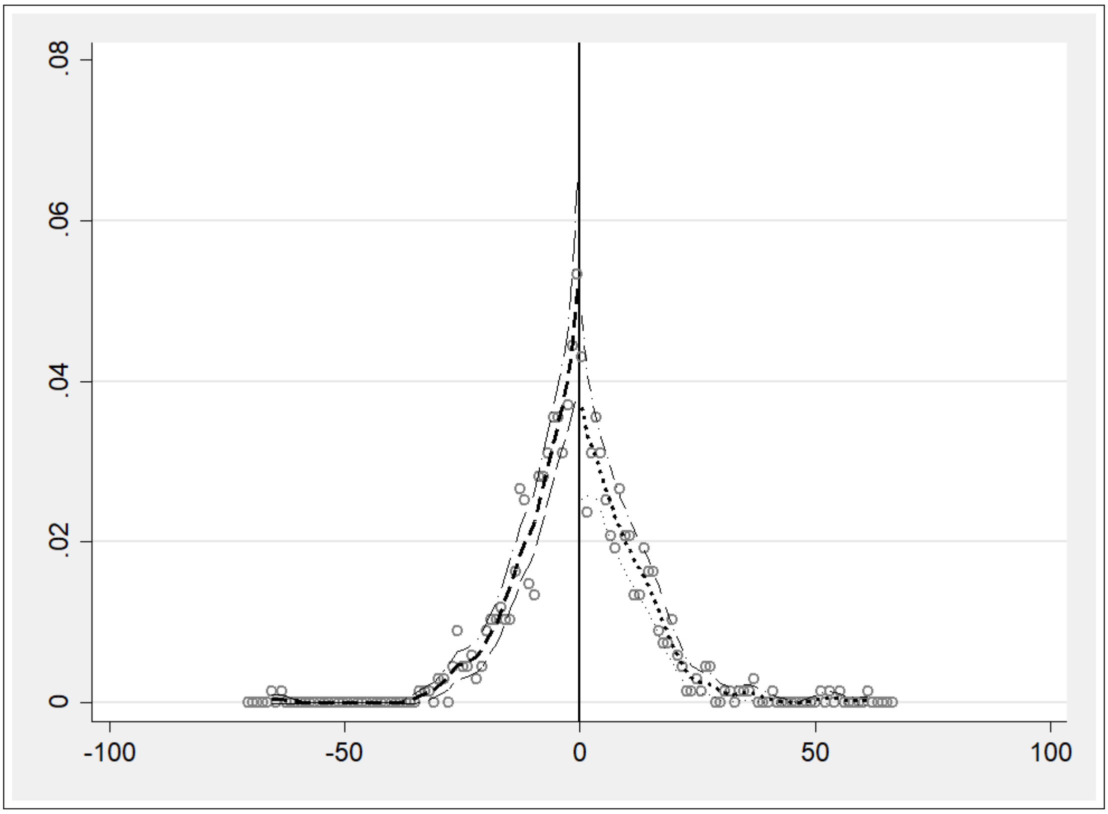


Figure 27: McCrary Test (Bandwidth=5 percentage points)

In Table 15, I show results from balance checks within the bandwidth. To arrive at these estimates, I use the empirical strategy outlined in Equation (9), but without covariates. Of the 15 covariates, I find that the estimate on the treatment variable is statistically significant for two variables: at the 5% level for whether the winner at the baseline was a member of socially disadvantaged groups (scheduled caste/SC or scheduled tribe/ST), and at the 10% level for whether the runner-up at baseline was a member of socially disadvantaged groups (scheduled caste/SC or scheduled tribe/ST). These results are not very worrying, for even if the coefficients are null, one expects one to two estimates to be statistically significant by chance. Furthermore, I include all these covariates in the benchmark specification.

Table 15: Balance Checks (Baseline Covariates)

	(1) RD Estimate
National Party=INC?	-0.246
	(0.132)
Winner SC/ST (t)?	-0.388**
	(0.035)
Runner-up SC/ST (t)?	-0.325*
	(0.067)
Winner female (t)?	0.037
	(0.894)
Runner-up female (t)?	-0.251
	(0.184)
Turnout percentage (t)	-0.047
	(0.735)
Effective number of parties (t)	0.010
	(0.969)
No. of candidates (t)	-0.095
	(0.615)
SC Reserved constituency?	-0.235
	(0.354)
Elevation	-0.105
	(0.463)
Ruggedness	-0.205
	(0.155)
Percentage employed (1998)	0.024
	(0.832)
Winner's age (t)	0.185
	(0.499)
Runner-up's age (t)	-0.178
	(0.601)
Number of Electors (t)	-0.080
	(0.367)
N	243

p-values in parentheses

*p<0.10, **p<0.05, ***p<0.01

RD estimate using local linear polynomial. All outcome variables are standardized. The specification controls for state and year fixed effects. Robust standard errors in parentheses. Number of observations for age-related variables is 142.

3.7 Main Results

I report the main results in Table 16. While all estimates are positive, not all are statistically significant at the 5% level. Given that the number of observations in the specifications without any controls (columns 1 and 4), and those with all controls except age-related variables (columns 2 and 5) is 243, statistical insignificance may be a result of lack of power. On the other hand though, when I include variables related to candidates' age at baseline, 101 observations drop out due to lack of data. Estimates obtained from this sample are statistically significant (columns 3 and 6). The coefficients are consistently high across all specifications. As per these estimates, the impact of a regional party narrowly losing an election to a national party during e_t on its candidate's wealth during e_{t+1} is anywhere in the range of 44-103% of the control mean.[39] The control mean for challenger's net worth per elector, i.e. challenger's net worth per elector in tightly contested constituencies where national party candidate challenges an incumbent regional party during e_{t+1} is INR 148. The same for the national party challenger's total assets per elector is INR 165 per elector.

Table 16: Main Results

	Log of Challenger's Total Assets per Elector			Log of Challenger's Total Net Worth per Elector		
	(1)	(2)	(3)	(4)	(5)	(6)
Regional Party Challenger=1	0.535*	0.368	0.709**	0.577**	0.445*	0.643**
	(0.289)	(0.265)	(0.304)	(0.282)	(0.243)	(0.314)
Bandwidth (pp.)	5	5	5	5	5	5
Baseline Covariates	None	All ex age	All	None	All ex age	All
N	243	243	142	241	241	142

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. All specifications control for state fixed effects. Robust standard errors in parentheses

[39] This range comes from estimates with minimum and maximum magnitudes, i.e. 0.368 and 0.709.

3.8 Robustness Checks

Choice of Bandwidth

In the main specification, I use an arbitrarily chosen narrow bandwidth (margin of victory), i.e. 5 percentage points. In Figure 28, I show that results are robust to the choice of different bandwidths, both narrower and wider.

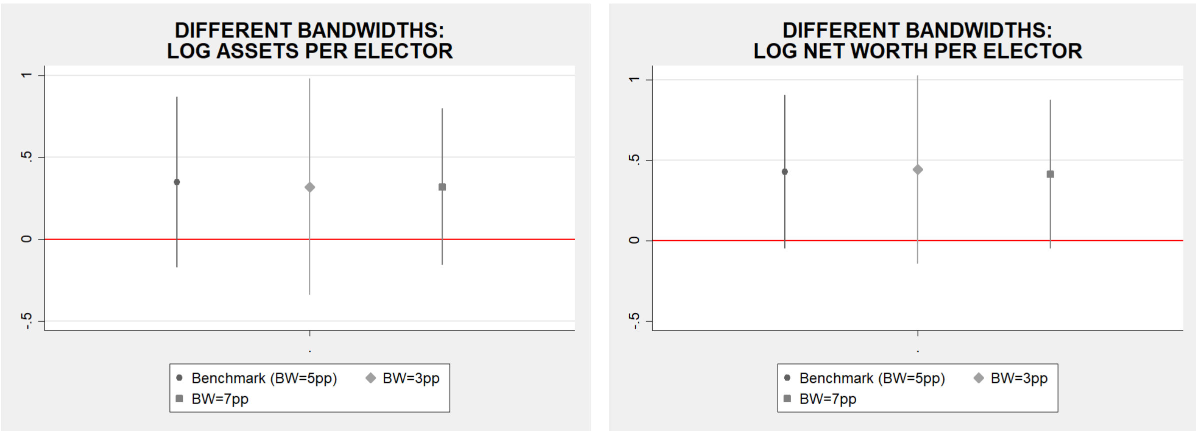


Figure 28: Robustness Checks: Different Bandwidths

Different Kernel Functions

In Figure 29, I show that results are robust to the usage of different types of kernels. In the benchmark specification, I use a triangular kernel function to assign weights to observations. In Figure 29, I show that results do not change if I use rectangular or Epanechnikov kernel functions.

Inclusion of e_{t+1} Controls

It is plausible that the treatment, i.e. election of a national party winner against a regional party during e_{t+1} may impact other characteristics of candidates during e_{t+1}. These include factors such as candidates' age, education, criminal history, etc. This may also include the incumbent

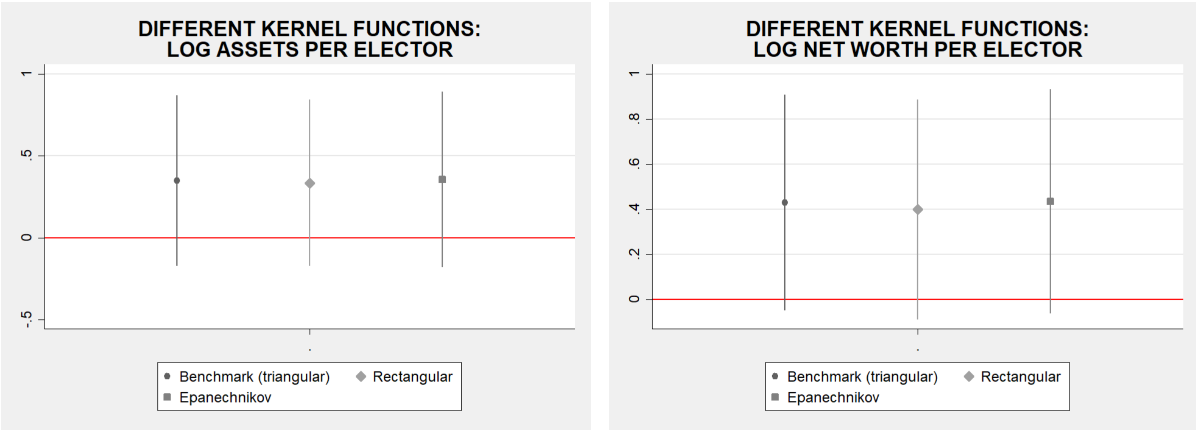


Figure 29: Robustness Checks: Different Kernel Functions

party's candidate's net worth. If these factors are correlated with challenger's wealth during e_{t+1}, then the coefficient on the treatment variable may be picking up the impact of treatment on these variables, and not challenger net worth. I address this concern by controlling for these factors in the benchmark specification. As shown in Figure 30, after the inclusion of these controls, the estimates remain stable, and the confidence intervals shrink.

Figure 30: Robustness Checks: Inclusion of Additional Controls (e_{t+1})

3.9 Discussion and Conclusion

Electoral success depends on several important factors, including electoral funding. In the Indian context, the distinction between regional and national parties is important. While national parties such as the Indian National Congress and the Bharatiya Janata Party are spread across several states and are usually at the center of all federal governments, regional parties are restricted to only a handful of states (and in many cases one state). As a result, it is reasonable to assume that the former have access to larger pools of financial resources than the latter.

As a result, I hypothesize that regional parties are more dependent on wealthy individuals, who can finance their own campaigns as well as contribute to the party's resource pool. Focusing on state-level assembly elections, I estimate whether regional party candidates who challenge incumbent national parties are wealthier than national party candidates who challenge incumbent regional parties. Using a sharp RD design, I focus on seats where regional parties narrowly win/lose against national parties during election cycle e_t. Seats where regional parties narrowly lose against national parties (treated) are those where they field challengers during election cycle e_{t+1}. whereas in seats where national parties narrowly lose against regional parties (control) are those where the former field challengers during e_{t+1}. The treatment effect is the difference between wealth of the regional party challenger and the national party challenger during election cycle e_{t+1}.

My estimates suggest that on average, regional parties field wealthier challengers than national parties during election cycle e_{t+1}. The magnitudes of these estimates suggest that these effects are in the range of 44-103% of the control mean. While results are robust to standard checks, I also add candidate-level controls from e_{t+1} to ensure that effects are not driven by candidate-level characteristics. The estimates are stable even after the inclusion of these controls.

Given that these estimates are localized-both in terms of focusing on constituencies with close contests as well as focusing on constituencies with contests between regional and national parties-they should be interpreted with caution. There may also be reasons to believe that these estimates are a lower bound of the true treatment effect. For instance, across several states (such as Bihar, Punjab and Maharashtra), there are instances where a regional party, in alliance with a national party, is competing against another national party or another alliance of a regional party and national party. For example, for a very long time in the state of Maharashtra, the Indian National Congress (a national party) and the Nationalist Congress Party (NCP, a regional party) have allied against the Bharatiya Janata Party (a national party) and the Shiv Sena (a regional party). In such cases, in a regional-national party contest, the regional party may not be too dependent on wealthy candidates because their alliance partner (a national party) may also be contributing to the former's resource pool. Similarly, candidates may not have all their wealth/assets listed in their own name. Instead, a substantial share of their true wealth may be amassed in the name of their family members. I cannot estimate these proportions separately, which makes it a limitation of the analysis. Overall, however, these results open up avenues for future research in the relatively underexplored area of electoral finance in developing countries like India.

A Appendix: Chapter 1

A.1 Land Ownership by Social Groups (1900s)

Figure 31: Land Ownership by Social Groups

A.2 Empirical Strategy: An Illustration

Villages in the green buffer zone (reformed) are compared with those within the orange buffer zone (unreformed). The border is divided into 25 segments of equal length.

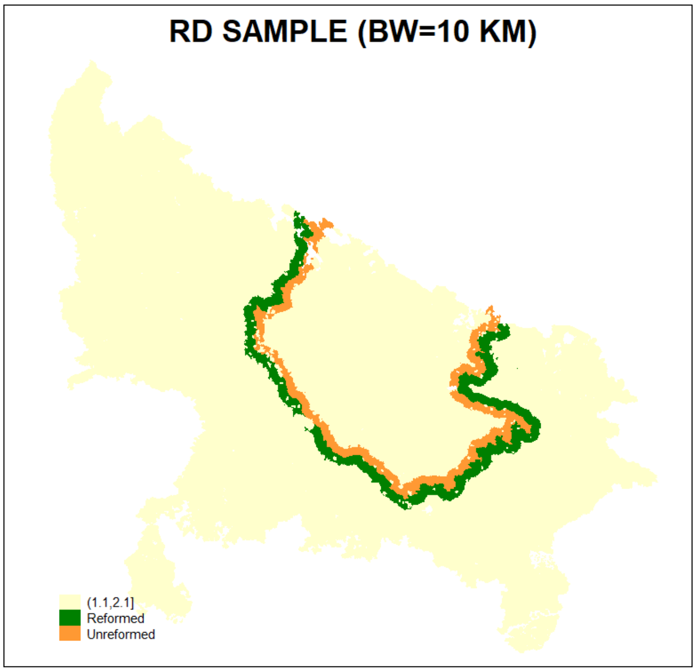


Figure 32: Identification Strategy

A.3 Construction of Index of Asset Ownership

Population Census of India (2011) gives data on ratio of HH within village with at least one of following assets.

- Car
- Mobile Phone
- Fixed Line Phone
- Two Wheeler (Scooter, Motorcycle)
- Bicycle
- Radio
- Television

In this paper, the Asset Index equals weighted average of these ratios, where the weight equals average regional price of asset as per National Sample Survey 2011-12. **Final Weights (INR):** Car (339,944), Mobile (1,906), Fixed Line (1,646), Two Wheeler (49,300), Bicycle (2,674), Radio (1,053), TV (4,358)

A.4 Sample from Ain-i-Akbari

	Bighas Biswas.	Revenue D.	Suyúrghál D.	Cavalry.	Infantry.	Elephants.	Castes.
Báwan	56,156	1,161,235	26,488	20	1,000	...	Ditto.
Basrah	60,063	......	.....	...	300	...	Various.
Bharwárah, has a brick fort	8,971-18	43,543	......	50	2,500	...	Ahnín [1]
Basárá	21,740	276,066	......	...	200	...	Bachhal.
Pilá	981-14	48,202	......	...	200	...	Ahnín.[1]
Chhatyápúr	64,706	1,765,641	41,094	50	700	...	Rajput Gaur.
Khairábád with suburban District, 2 Mahals, has a brick fort ...	159,072	2,161,234	174,191	50	2,000	...	Bráhman.
Sánḍi, has a brick fort...	211,804	3,055,339	195,106	20	2,000	...	Sombansi.
Sarah	68,832	2,091,983	8,666	60	500	...	Chauhán.
Sadrpúr	120,698	881,175	15,581	20	500	...	Janwár, Báchhal.
Gopámau, has a brick fort	107,368-5	5,620,466	562,037	100	3,000	...	Rájput Kuár.
Kherí, do. do. ...	260,168	3,250,522	50,522	60	1,500	...	Bisen, Rájpút, Janwár.
Khairigarh, one of the most important fortresses in Hindustán. There are 6 forts of brick and mortar, at a short distance from it.	43,052-7	1,829,328	......	300	1,500	...	Bais, Bisen, Bachhal, Kahnah.
Kharkhelá	15,815-16	478,727	......	20	500	...	Asín.[1]
Khánkhat Mau ...	3,058-11	235,656	......	...	400	...	Various.
Láharpúr	208,288	3,029,479	209,079	50	1,000	...	Bráhman.
Machharhaṭṭah ...	71,069	2,112,176	2,430	30	2,000	...	Rájpút, Bachhal.
Nímkhár, has a brick fort	58,775-18	3,566,055	66,055	100	1,500	...	Ahír.
Hargaráoṇ	66,952	200,000	26,385	20	500	...	Bráhman.

Sarkár of Lucknow.

Containing 55 *Mahals*, 3,307,426 *Bíghas*, 2 *Biswas*. Revenue 80,716,160 *Dáms* in money. *Suyúrghál*, 4,572,526 *Dáms*. Castes various. Cavalry 2,680. Elephants 36. Infantry 83,450.

	Bighas. Biswas.	Revenue D.	Suyúrghál D.	Cavalry.	Infantry.	Elephants	Castes.
Abethi (Amethí), has a brick fort	117,381	3,076,480	300,217	300	2,000	20	Ansárí,
Unám, has a brick fort.	61,045	2,012,372	253,747	50	4,000	...	Sayyid.
Isauli, has a brick fort on the Gúmti. ...	1,670,093[2]	4,208,046	240,846	50	2,000	...	Rájpút, Bachgoti.

Figure 33: Sample from Ain-i-Akbari)

A.5 The Process of Georeferencing

(a) Modern Day (b) Colonial Map (c) Pre-Colonial Map

Figure 34: Georeferencing Across Time

A.6 Variable Definitions and Sources (Secondary)

Variables	Definition	Source	Resolution
Asset Index	The Population Census lists percentage of households owning at least one of the following assets (separately): car, bicycle, two-wheeler, radio, TV, mobile phone, fixed line phone and computer. The Socio-Economic Caste Census has data on percentage of households owning at least one of the following assets (separately): car, two-wheeler, mobile phone, fixed line phone and refrigerator. I compiled an asset index using this data as explained in Appendix A.3.	Population Census (2011), Socio-Economic Caste Census (2011)	Village
Nonfarm Employment	An establishment is defined as "an enterprise or part of an enterprise that is situated in a single location in which one or predominantly one kind of economic activity is carried out". In the benchmark empirical strategy, I used the number of people employed in non-agricultural economic establishments per 100 residents of the village. It is possible that not all workers employed in an establishment belong to the village. Hence, I remove those observations where total workers employed in a village is higher than the village's population.	Economic Census (2013)	Village
Years of schooling	Years of schooling are recorded in the Socio-Economic Caste Census (2011) at the individual level. For the purpose of analysis, I calculated the average years of schooling per individual within a village.	Socio-Economic Caste Census (2011)	Village

Variables	Definition	Source	Resolution
Soil Variables (2015)	These include the following variables: Sand Mass, Coarse Fragments, Silt, Clay, Bulk Density, Water Capacity, Organic Carbon, Nitrogen, C/N Ratio, Electrical Conductivity, Soil Reaction, Cation Exchange, Base Saturation, Exchangeable Sodium, Carbonate Equivalent, Clay Size, Gypsum, Effective Cation Exchange and Aluminium. These are available at a resolution of 1 km^2. For each village, I calculated the average value using data from all polygons of 1 km^2 that can be mapped into it.	International Soil Reference and Information Centre (2015)	1 km^2
Distance to Nearest River	Using coordinates of rivers from this shape file, I calculated distance between villages and the nearest river (using GIS).	National River Linking Project	Geographical Coordinates of Rivers
Village-Level Maps		Meiyappan et al. (2018)	Geographical Coordinates of Villages from 2001 Census
Pre-Treatment Covariates	Using an archival source, I compiled data on tax demands per unit land and important caste and religious groups from the pre-treatment period, i.e. the 16^{th} century.	Ain-i-Akbari (16th Century)	Sub-District
Colonial Tax Data	Using colonial gazettes, I compiled data on tax demands per unit land from the first decade of the 20^{th} century.	Colonial Gazettes (1900s)	Sub-District
Historical Agricultural Income		Donaldson (2018)	District
Historical Inter-District Migration	In order to obtain data on inter-district migration, I digitized tables on population (residents) by district of birth from archival documents.	Population Census (1881)	District

A.7 Survey Representativeness

In this Figure 35 and Figure 36, I show representativeness of the randomly selected villages across three different kinds of variables: soil, bio-climatic and socio-economic outcomes.

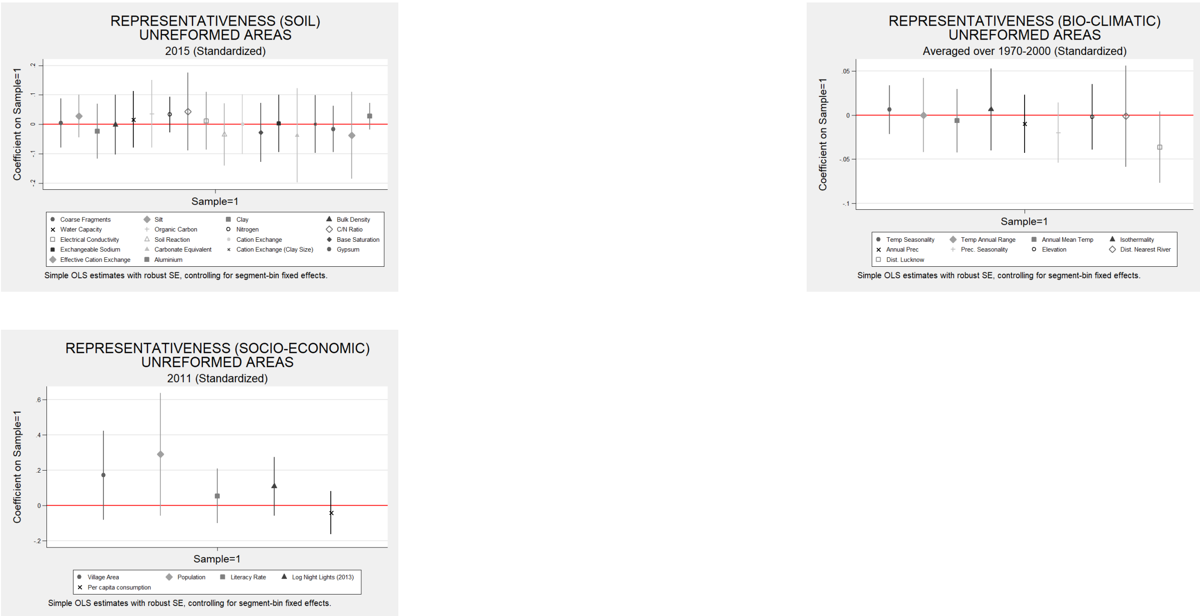


Figure 35: Sample Representativeness (Unreformed Areas)

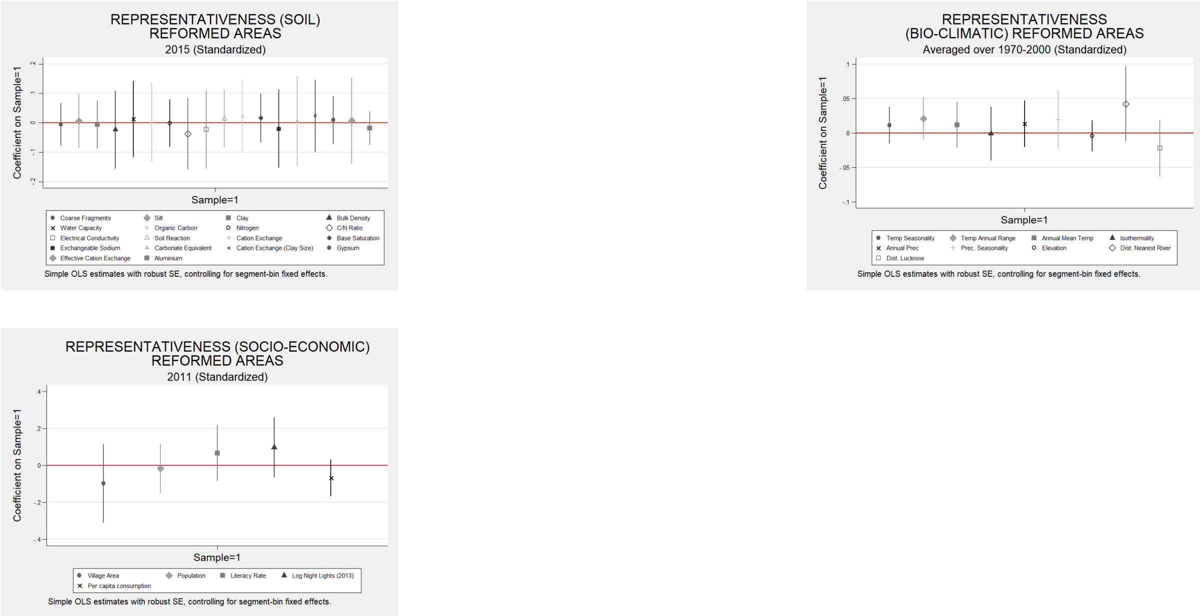


Figure 36: Sample Representativeness (Reformed Areas)

A.8 Balance Checks: Geographical Variables

Table 17: Balance Checks (Soil Characteristics, 2015)

	(1) RD Estimate
Sand Mass	-0.0014
	(0.829)
Coarse Fragments	0.0160
	(0.365)
Silt	0.0065
	(0.538)
Clay	-0.0062
	(0.416)
Bulk Density	0.0031
	(0.776)
Water Capacity	0.0020
	(0.906)
Organic Carbon	0.0074
	(0.466)
Nitrogen	0.0095
	(0.284)
C/N Ratio	0.0193
	(0.143)
Electrical Conductivity	0.0143
	(0.509)
Soil Reaction	-0.0091
	(0.163)
Cation Exchange	-0.0057
	(0.552)
Base Saturation	-0.0093
	(0.122)
Exchangeable Sodium	0.0085
	(0.681)
Carbonate Equivalent	-0.0130
	(0.391)
Clay Size	-0.0045
	(0.842)
Gypsum	-0.0099
	(0.667)
Effective Cation Exchange	-0.0099
	(0.372)

Table 18: Balance Checks (Geography and Bioclimatic Variables, 1970-2000)

	(1) RD Estimate
Log Dist. Nearest River	0.0551
	(0.238)
Temp Seasonality	0.0032
	(0.641)
Temp Annual Range	0.0064
	(0.611)
Prec Wettest Q	0.0167
	(0.416)
Prec Wettest M	0.0015
	(0.951)
Prec Warmest Q	-0.0347
	(0.107)
Prec Seasonality	0.0364**
	(0.040)
Prec Driest Q	-0.0256*
	(0.070)
Prec Driest M	0.0205
	(0.556)
Prec Coldest Q	-0.0081
	(0.316)
Min Temp Coldest M	0.0152
	(0.171)
Mean Temp Wettest Q	0.0289
	(0.210)
Mean Temp Warmest Q	0.0147
	(0.181)
Mean Temp Driest Q	-0.0056
	(0.918)
Mean Temp Coldest Q	0.0115
	(0.240)
Mean Diurnal Range	-0.0061
	(0.797)
Max Temp Warmest M	0.0162
	(0.271)
Isothermality	-0.0220
	(0.380)
Elevation	0.0136
	(0.200)
Annual Mean Temp	0.0162

A.9 Balance on Pre-Treatment Variables

Table 19: Balance on Pre-Treatment Covariates (Sub-District Level, 16th century)

	(1) Log Infantry Per Unit Land	(2) Log Cavalry Per Unit Land	(3) Log Land Revenue Demanded Per Unit Land	(4) Log Charitable Grants Per Unit Land
Reformed Area=1	0.541	0.363	0.201	0.238
	(0.347)	(0.327)	(0.123)	(0.365)
Control Mean (level)	.032	.001	43.882	1.493
N	55	48	59	47

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Control mean is in absolute terms (not log). Each sub-district is assigned a weight that equals the sum of RD weights assigned to its villages in the main specification. The specification controls for agro-climatiz zone fixed effects. Robust standard errors reported.

A.10 Cross-Border Migration within Bandwidth (1881)

This analysis is done at the level of the colonial districts. The earliest Census was conducted in 1881. In this table, I show the average of the proportion of people who moved from reformed to unreformed districts and vice versa within the bandwidth. I then test whether the difference between these two statistics is statistically significant or not. As the p-value shows, it is not significant.

	Reformed to Unreformed	Unreformed to Reformed	Difference	p-value
Proportion Migrated Out	.03312	.03029	.0028	.76159
Observations	8	8		

A.11 Main Treatment Effects without Controls

In the following table, I show estimates of treatment effects without any controls. However, as discussed in Section 3.4, I do control for agro-climatic zone and segment fixed effects.

Table 20: Village Level Socio-Economic Indicators without Controls (2011)

	(1) Average Asset Ownership Index	(2) Nonfarm Workers Per 100 People	(3) Log Avg Years of Schooling
Reformed Area=1	0.223*	1.285***	0.054
	(0.123)	(0.211)	(0.040)
Control Mean	4.94	2.63	4.28
Bandwidth	10 km	10 km	10 km
N	12579	12504	11871

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

A.12 Robustness

A.12.1 Different Bandwidths

In the following tables, I show that estimates of treatment effects for the three outcome variables are robust to the usage of different bandwidths (5-15 kilometers). In addition, in Figure 37, Figure 38 and Figure 39, I show that these results remain robust even if I extend the bandwidth up to 25 kilometers.

Table 21: Robustness Across Bandwidths: Index of Durable Assets

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Reformed Area=1	0.307***	0.397***	0.380***	0.356***	0.334***	0.320***	0.295**	0.289**	0.282**	0.271**	0.264**
	(0.116)	(0.114)	(0.117)	(0.118)	(0.118)	(0.118)	(0.115)	(0.116)	(0.117)	(0.118)	(0.118)
Control Mean	4.94	4.96	4.96	4.96	4.96	4.95	4.96	4.96	4.96	4.95	4.94
Bandwidth (km)	10	5	6	7	8	9	11	12	13	14	15
N	10412	5917	6873	7776	8668	9534	11276	12136	12970	13751	14514

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

Table 22: Robustness Across Bandwidths: Number of Nonfarm Workers per 100 people

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Reformed Area=1	1.333***	1.235***	1.295***	1.318***	1.321***	1.321***	1.333***	1.306***	1.277***	1.247***	1.216***
	(0.238)	(0.253)	(0.251)	(0.247)	(0.243)	(0.241)	(0.234)	(0.231)	(0.231)	(0.232)	(0.233)
Control Mean	2.63	2.45	2.51	2.53	2.57	2.6	2.64	2.62	2.63	2.62	2.62
Bandwidth (km)	10	5	6	7	8	9	11	12	13	14	15
N	10364	5885	6839	7740	8627	9489	11228	12086	12918	13694	14460

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

Table 23: Robustness Across Bandwidths: Log Average Years of Schooling

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Reformed Area=1	0.063*	0.061*	0.063*	0.063*	0.063*	0.063*	0.063*	0.063*	0.062*	0.061*	0.061*
	(0.035)	(0.036)	(0.036)	(0.036)	(0.036)	(0.035)	(0.034)	(0.034)	(0.033)	(0.033)	(0.033)
Control Mean	4.28	4.27	4.27	4.27	4.28	4.28	4.28	4.28	4.28	4.28	4.28
Bandwidth (km)	10	5	6	7	8	9	11	12	13	14	15
N	9862	5630	6524	7373	8218	9034	10682	11502	12293	13029	13765

Robust standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

Each bar represents a point estimate of the treatment effect and its 95% confidence interval. The first bar in each graph is the estimate using a bandwidth of 10 kilometers. Each subsequent bar represents estimates from a bandwidth that is 1 kilometer wider than the previous bar's. The last bar represents a bandwdith of 25 kilometers.

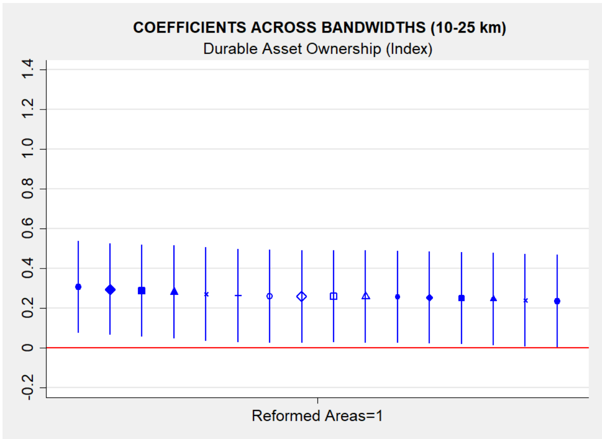


Figure 37: Different Bandwidths (Index of Durable Assets, Standardized)

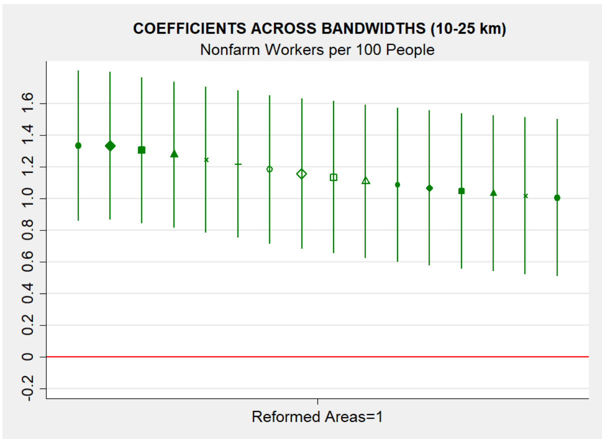


Figure 38: Different Bandwidths (Nonfarm Employment per 100 People, Standardized)

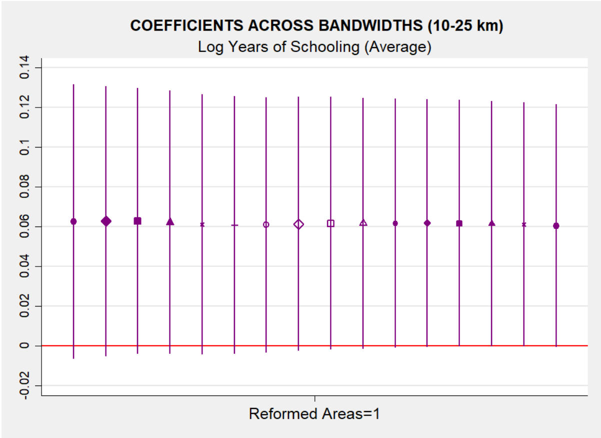


Figure 39: Different Bandwidths (Log Average Years of Schooling, Standardized)

A.12.2 Donut Hole RD Estimates

Of the 12 donut-hole estimates (columns (2) to (4) of Table 24, Table 25 and Table 26), three are statistically insignificant at conventional levels, although they are statistically indistinguishable from the benchmark estimates. The p-values of the estimates are listed below.

1. **Assets (2 km donut-hole):** 0.16

2. **Assets (3 km donut-hole):** 0.31

3. **Years of Schooling (2 km donut-hole):** 0.22

4. **Nonfarm Employment per 100 pople (4 km donut-hole):** 0.14

Table 24: Donut Hole RD Estimates: Index of Durable Assets

	(1)	(2)	(3)	(4)
Reformed Area=1	0.307***	0.221	0.191	0.529***
	(0.116)	(0.155)	(0.187)	(0.188)
Control Mean	4.94	4.98	4.99	4.93
Donut Hole (km)	0	2	3	4
Bandwidth (km)	10	10	10	10
N	10412	9238	9028	8815

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects.

Table 25: Donut Hole RD Estimates: Nonfarm Workers per 100 people

	(1)	(2)	(3)	(4)
Reformed Area=1	1.333***	0.941***	0.670**	0.429
	(0.238)	(0.246)	(0.268)	(0.287)
Control Mean	2.63	2.73	2.77	2.74
Donut Hole (km)	0	2	3	4
Bandwidth (km)	10	10	10	10
N	10364	9211	9005	8788

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects.

Table 26: Donut Hole RD Estimates: Log Average Years of Schooling

	(1)	(2)	(3)	(4)
Reformed Area=1	0.063*	0.035	0.048*	0.066**
	(0.035)	(0.028)	(0.027)	(0.030)
Control Mean	4.28	4.3	4.29	4.28
Donut Hole (km)	0	2	3	4
Bandwidth (km)	10	10	10	10
N	9862	8693	8513	8318

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects.

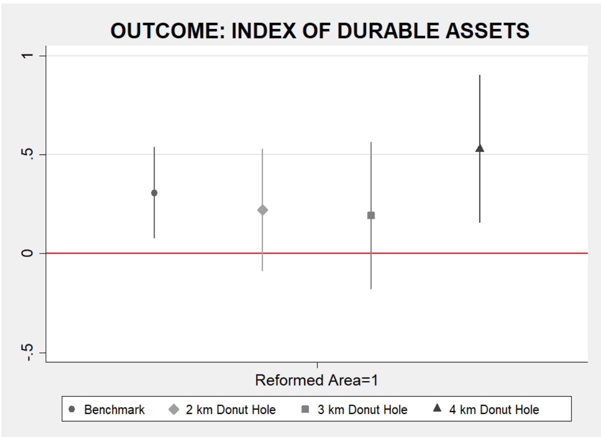


Figure 40: Donut Hole RD (Index of Durable Assets)

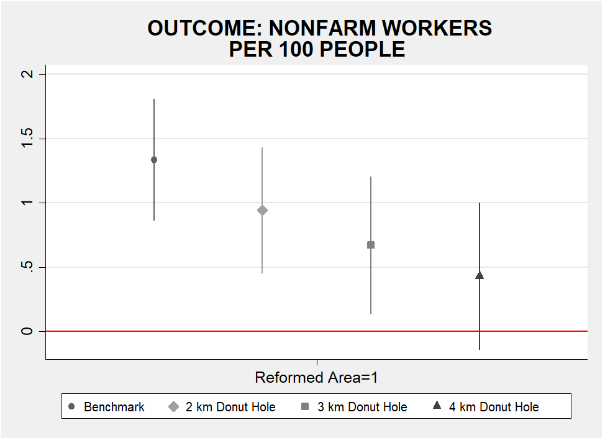


Figure 41: Donut Hole RD (Nonfarm Employment per 100 People)

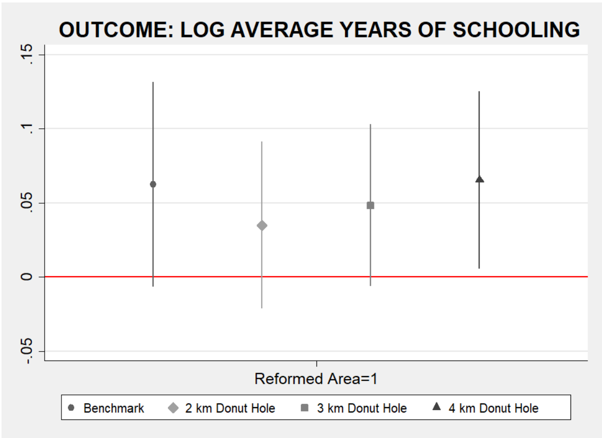


Figure 42: Donut Hole RD (Log Average Years of Schooling)

A.12.3 Inclusion of Pre-Treatment Covariates

In Table 27, I show estimates of the treatment effect with and without the inclusion of pre-treatment covariates in the benchmark specification. These covariates are compiled from *Ain-i-Akbari*, a book written by Abul Fazl, a member of Mughal emperor Akbar's court, in the 16^{th} century. This book has been cited frequently across colonial documents as reference material for the pre-colonial era. These covariates include land revenue demand, cavalry and infantry per unit area, and categorical variables for important social groups (upper caste, intermediate caste, lowest caste and Muslims) at the sub-district level. Owing to absence of data at the village level, I assign values compiled for a historical sub-district to all modern day villages matched to it. The sample size (number of villages) in these specifications is roughly 20% smaller than that in the main results, for I was not able to match those villages to areas mentioned in historical documents.

Table 27: Robustness to Inclusion of Pre-Treatment Covariates

	(1) Average Asset Ownership Index	(2) Average Asset Ownership Index	(3) Nonfarm Workers Per 100 People	(4) Nonfarm Workers Per 100 People	(5) Log Avg Years of Schooling	(6) Log Avg Years of Schooling
Reformed Area=1	0.338***	0.368***	1.600***	1.653***	0.086**	0.081*
	(0.114)	(0.115)	(0.279)	(0.274)	(0.039)	(0.045)
Control Mean	5.08	5.08	2.67	2.67	4.32	4.32
Pre-Treatment Covariates?	No	Yes	No	Yes	No	Yes
Bandwidth (km)	10 km	10 km	10 km	10 km	10 km	10 km
N	8431	8431	8392	8392	8063	8063

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

A.12.4 Removal of Southern Boundary

Table 28: Robustness to Removal of Southern Boundary: All Indicators

	(1) Average Asset Ownership Index	(2) Average Asset Ownership Index	(3) Nonfarm Workers Per 100 People	(4) Nonfarm Workers Per 100 People	(5) Log Avg Years of Schooling	(6) Log Avg Years of Schooling
Reformed Area=1	0.307***	0.307**	1.333***	1.154***	0.063*	0.033
	(0.116)	(0.127)	(0.238)	(0.267)	(0.035)	(0.031)
Control Mean	4.94	5.29	2.63	2.65	4.28	4.28
Sample	Benchmark Sample	Ex-Southern Boundary	Benchmark Sample	Ex-Southern Boundary	Benchmark Sample	Ex-Southern Boundary
Bandwidth (km)	10	10	10	10	10	10
N	10412	7872	10364	7836	9862	7341

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects.

A.12.5 Removal of Permanently Settled Areas

Table 29: Robustness to Removal of Permanently Settled Areas: All Indicators

	(1) Index of Durable Assets	(2) Index of Durable Assets	(3) Nonfarm Workers per 100 People	(4) Nonfarm Workers per 100 People	(5) Log Average Years of Schooling	(6) Log Average Years of Schooling
Reformed Area=1	0.307***	0.334**	1.333***	1.700***	0.063*	0.069
	(0.116)	(0.147)	(0.238)	(0.247)	(0.035)	(0.043)
Control Mean	4.94	4.65	2.63	2.55	4.28	4.12
Sample	Benchmark	Ex-Permanently Settled	Benchmark	Ex-Permanently Settled	Benchmark	Ex-Permanently Settled
Bandwidth (km)	10	10	10	10	10	10
N	10412	8513	10364	8467	9862	8072

Robust standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

A.12.6 Removal of Rohilkhand (Western Border)

Table 30: Robustness to Removal of Rohilkhand: All Indicators

	(1) Index of Durable Assets	(2) Index of Durable Assets	(3) Nonfarm Workers per 100 People	(4) Nonfarm Workers per 100 People	(5) Log Average Years of Schooling	(6) Log Average Years of Schooling
Reformed Area=1	0.307***	0.275**	1.333***	1.132***	0.063*	0.073*
	(0.116)	(0.118)	(0.238)	(0.250)	(0.035)	(0.037)
Control Mean	4.94	5.06	2.63	2.69	4.28	4.36
Sample	Benchmark	Ex-Rohilkhand	Benchmark	Ex-Rohilkhand	Benchmark	Ex-Rohilkhand
Bandwidth (km)	10	10	10	10	10	10
N	10412	9003	10364	8971	9862	8518

Robust standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

A.12.7 Usage of Different Kernels

Table 31: Robustness to Choice of Kernels: Index of Durable Assets

	(1)	(2)	(3)
Reformed Area=1	0.307***	0.242*	0.289**
	(0.116)	(0.130)	(0.122)
Control Mean	4.94	4.94	4.94
Kernel	Triangular	Rectangular	Epanechnikov
Bandwidth (km)	10	10	10
N	10412	10412	10412

Robust standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

Table 32: Robustness to Choice of Kernels: Nonfarm Workers per 100 People

	(1)	(2)	(3)
Reformed Area=1	1.333***	1.347***	1.343***
	(0.238)	(0.243)	(0.241)
Control Mean	2.63	2.63	2.63
Kernel	Triangular	Rectangular	Epanechnikov
Bandwidth (km)	10	10	10
N	10364	10364	10364

Robust standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

Table 33: Robustness to Choice of Kernels: Log Average Years of Schooling

	(1)	(2)	(3)
Reformed Area=1	0.063*	0.064*	0.062*
	(0.035)	(0.034)	(0.035)
Control Mean	4.28	4.28	4.28
Kernel	Triangular	Rectangular	Epanechnikov
Bandwidth (km)	10	10	10
N	9862	9862	9862

Robust standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

A.12.8 Different Levels of Clustering

Table 34: Robustness to Different Levels of Clustering: Index of Durable Assets

	(1)	(2)	(3)	(4)	(5)	(6)
Reformed Area=1	0.307***	0.307***	0.307***	0.307**	0.307**	0.307**
	(0.116)	(0.109)	(0.100)	(0.137)	(0.137)	(0.133)
Control Mean	4.94	4.94	4.94	4.94	4.94	4.94
Level of Clustering	Segment-District	No Clustering	Segment	District	Sub-District	Development Block
Bandwidth (km)	10	10	10	10	10	10
N	10412	10412	10412	10412	10412	10412

Robust standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects.

Table 35: Robustness to Different Levels of Clustering: Nonfarm Workers per 100 People

	(1)	(2)	(3)	(4)	(5)	(6)
Reformed Area=1	1.333***	1.333***	1.333***	1.333***	1.333***	1.333***
	(0.238)	(0.157)	(0.310)	(0.294)	(0.274)	(0.262)
Control Mean	2.63	2.63	2.63	2.63	2.63	2.63
Level of Clustering	Segment-District	No Clustering	Segment	District	Sub-District	Development Block
Bandwidth (km)	10	10	10	10	10	10
N	10364	10364	10364	10364	10364	10364

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects.

Table 36: Robustness to Different Levels of Clustering: Log Average Years of Schooling

	(1)	(2)	(3)	(4)	(5)	(6)
Reformed Area=1	0.063*	0.063***	0.063	0.063	0.063*	0.063*
	(0.035)	(0.015)	(0.040)	(0.043)	(0.036)	(0.034)
Control Mean	4.28	4.28	4.28	4.28	4.28	4.28
Level of Clustering	Segment-District	No Clustering	Segment	District	Sub-District	Development Block
Bandwidth (km)	10	10	10	10	10	10
N	9862	9862	9862	9862	9862	9862

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects.

A.12.9 Usage of Conley S.E.

Table 37: Robustness to Usage of Conley Standard Errors

	(1) Index of Durable Assets	(2) Index of Durable Assets	(3) Nonfarm Workers per 100 People	(4) Nonfarm Workers per 100 People	(5) Log Average Years of Schooling	(6) Log Average Years of Schooling
Reformed Area=1	0.242*	0.242**	1.347***	1.407***	0.064*	0.066*
	(0.130)	(0.122)	(0.243)	(0.316)	(0.034)	(0.039)
Control Mean	4.94	4.94	2.63	2.63	4.28	4.28
Type of S.E.	Clustered S.E.	Conley S.E.	Clustered S.E.	Conley S.E.	Clustered S.E.	Conley S.E.
Bandwidth (km)	10	10	10	10	10	10
N	10412	10412	10364	10364	9862	9862

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using rectangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects.

A.13 Statements related to Caste-Based Stereotyping (Scale of 1-5)

1. On the whole, having political leaders from an upper caste is better for the voters.

2. A university education is more important for a child from an upper caste family than other children in the village.

3. When jobs (government or private) are scarce, employers should give priority to people of upper caste families.

4. If Dalit (lowest caste) families earn more money than rest of the village, it's almost certain to cause problems.

5. One's social status is determined by will of god and last life's deeds, and it should not be challenged.

6. One's social status is determined by will of god and last life's deeds, but it can be changed.

7. Families should ideally pursue professions that the varna/caste system dictates.

A.14 Treatment Effects on Colonial Governance

A.14.1 Colonial Tax Demands

	Mean (Control)	Mean (Treatment)	Difference	p-value
Log of Tax Demand (Cultivable Area)	.7908924	.7537783	.0371141	.6593173
Observations	54			

	Mean (Control)	Mean (Treatment)	Difference	p-value
Log of Tax Demand (Overall Area)	.2678363	.2857629	-.0179265	.8325507
Observations	54			

A.14.2 Presence of School

Table 38: Outcome: Presence of Colonial School (Early 20th Century)

	(1) Probability of Colonial School Presence
Reformed Area=1	0.011
	(0.008)
Control Mean	.05
Bandwidth	10 km
N	11034

Robust standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

A.15 Model (Acemoglu and Robinson, 2006)

Consider the following stakeholders in the economy: a group of citizens, with mass normalized to 1, an incumbent ruler, and an infinite stream of potential new rulers. All agents live infinitely and maximize the net discounted income, using a discount factor β. While citizens live infinitely, rulers, once replaced by a new ruler, receive no utility. In this economy, there is only one good, which is produced using the following production function.

$$y_t = A_t \tag{10}$$

Here, A_t is the state of technology at time t, such that it also captures the institutional context that is critical for production. For instance, apart from technological innovation in the literal sense, A_t also captures factors such as rule of law and enforcement of property rights. When a new change (to the institution of caste, in the case of this paper) is introduced, A rises to αA, such that $\alpha > 1$. The cost of bringing about this change is normalized to zero. In addition, replacement of the elite with a new group is associated with a cost, so is the process of the introduction of change. In particular, the cost of replacing the incumbent is zA, whereas the cost of bringing about change is $z'A$. The following equations explains the relationship.

$$A_t = A_{t-1}((1-p_t)(1+(\alpha-1)x_t)) + p_t(1+(\alpha-1)\hat{x_t} - x_t z' - (1-x_t)z)) \tag{11}$$

In the above equation, x_t equals 1 if the change happens at time t under the existing elite, and zero otherwise. Similarly, $\hat{x_t}$ equals 1 if the change happens at time t under a new elite, and zero otherwise. p_t equals 1 if the elite is replaced and zero otherwise. The model further assumes that z and z' are random variables. They are assumed to be drawn from different distributions. While z is drawn from the distribution F^N, z' is drawn from F^I. The model assumes that F^I is first-order stochastically dominated by F^N. This assumption captures the idea that introduction of the change erodes the initial elite's authority.

As a simplification, the model assumes that F^I is a uniform distribution over the interval $[\mu - 1/2, \mu + 1/2]$. Similarly, F^N is a uniform distribution over the interval $[\gamma\mu - 1/2, \gamma\mu + 1/2]$, where $\gamma \geq 1$. μ represents the elite's political monopoly (or absence of political competition). A lower μ represents low incumbent advantage and vice versa. If $\mu = 0$, it implies that there is no advantage to the incumbent. γ represents the said socio-economic change's potential to erode elite's political monopoly. A higher γ, therefore, gives lower incentive to the elite to support the said change.

Finally, rulers also impose a tax T on citizens. The model assumes that citizens have access to a non-taxable informal technology that produces $(1 - \tau)A$, thus making any T higher than τ non-optimal. In the following bullet points, the timing of events is spelled out.

1. Period starts with A_t
2. Incumbent makes the decision regarding x_t
3. Stochastic costs of replacement- z_t and z'_t are revealed
4. Citizens decide on p_t
5. If $p_t = 1$, a new elite comes to power and takes the decision regarding $\hat{x}_t$
6. The ruler in power decides T_t

Using Bellman equations, the authors then characterize the social planner's solution as well as a decentralized equilibrium of the game, which lead to the following propositions.

- The social planner always innovates
- γ and innovation are negatively correlated. The effect of μ is ambiguous: when μ is

sufficiently small or large (political competition very high or very low), the elites will always innovate. For intermediate values of μ, economic change may be blocked.

Further, after incorporating recurring political rents to the elite R, human capital of the non-elite h (such that $y_t = hA_t$) and possibility of external threats, the following additional propositions are laid out.

- Political elites are more likely to block economic change when political rents R are high and human capital h of the non-elite is low.
- Political elites are less likely to block development when there is a severe external threat and when the perpetrator is more developed.

A.16 Literacy Levels (1881)

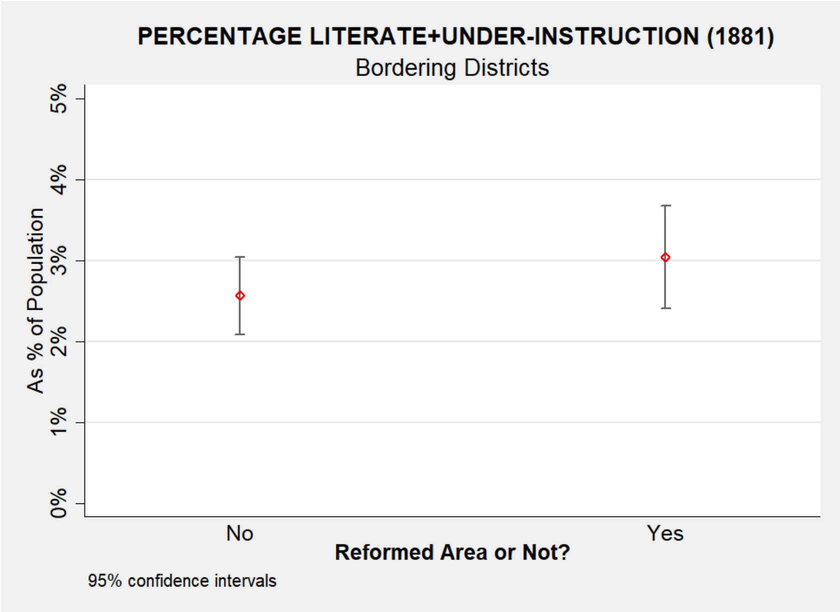


Figure 43: Literacy Rates in Bordering Districts (1881)

A.17 The Ultimatum Game

This game is usually played between two players: a Proposer and a Respondent. The surveyor gives the Proposer an amount to split. The Proposer is supposed to propose a share out of that amount to Respondent. If accepted by the Respondent, both get the Proposer and the Respondent get payouts. If the Respondent declines, neither gets anything. The structure of the game incentivizes both parties to cooperate. Rational agents should accept any non-zero offer. In the survey, following Kubota et al (2013), I read out 20 random offers from fictitious proposers to the respondent. The Respondent was told that the Proposer has been asked to split INR 40 (worth a McDonald's burger in India) between the two. Across the 20 offers, the amount and the Proposer's caste (gauged from the last name) varied.

- **Ex.:** "Of INR 40, Raj **Maurya** wants to keep 25 for himself and give you 15"

Respondents were told that if the offer was accepted, both Proposer and Respondent may be entitled to a payout, otherwise both get nothing. Respondents were asked to respond within 4-6 seconds to avoid cognitive override (Hoffman et al., 1994; Cram et al., 2018). In the following matrix, the offers made to the respondent have been summarized.

		OFFER		
		INR 5	**INR 10**	**INR 15**
CASTE OF PROPOSER	**Landed Elite Caste**	2	2	3
	Intermediate	2	2	1
	Lowest	2	2	2
	Placebo	1	1	0

Figure 44: Offers in the Ultimatum Game

A.18 Treatment Effects on Number of Workers per Economic Establishment

Table 39: Treatment Effects on Number of Workers per Economic Establishment (2013)

	(1)	(2)	(3)
	All Establishments	SC/ST Owned	OBC Owned
Reformed Area=1	-0.277***	-0.121***	-0.244**
	(0.105)	(0.043)	(0.104)
Control Mean	2.15	1.47	1.79
Bandwidth	10 km	10 km	10 km
N	9526	5103	7964

Robust standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Local linear regression coefficient estimated using triangular kernels. Winsorized outcomes (1 per cent on both sides, if not in log terms) used. Specifications control for geographic covariates, segment and agro-climatic zone fixed effects. Standard errors clustered at the district-segment level.

B Appendix: Chapter 2

B.1 Variable Definitions

- **Outcome Variables:**
  - **Proportion of Transactions processed during the interval t:** Number of beneficiary transactions processed within interval t (example: 16-30 days) divided by total number of beneficiary transactions processed in a GP during FY 2014-15.
  - **Proportion of Amount processed during the interval t:** Amount of beneficiary payments processed within interval t (example: 16-30 days) divided by total amount of beneficiary payments processed in a GP during FY 2014-15.
- **Endogenous Regressors:**
  - **Ratio of GP's villages connected by PMGSY roads (before FY 2014-15):** Number of villages in a GP connected by a PMGSY road until March 31, 2014 divided by total number of villages in the GP.
  - **Ratio of GP's population connected by PMGSY roads (before FY 2014-15):** Number of people in a GP connected by a PMGSY road until March 31, 2014 divided by total number of people in the GP. Village-level population figures are obtained from the Population Census of 2011.
  - **Ratio of GP's population connected by any roads (before 2009):** Number of villages in a GP connected by any paved road as reported in the Population Census of 2011 (data collected in 2009) divided by total number of villages in the GP.
- **Instrument:**
  - **Number of just eligible villages (benchmark bandwidth=50):** Total number of villages in a GP with population 500-550 (for all four states in sample) as per the Population Census of 2001. In the case of Madhya Pradesh, I also add villages with population 1000-1050.

- **Covariates (2001):**
  - **Ratio of GP's villages connected by road before 2001:** Weighted average of the number of villages with road connectivity before the year 2001. A village's population as per the Population Census of 2001 is used as its weight.
  - **Ratio of GP's villages with access to power:** Weighted average of the number of villages with access to power as per the Population Census of 2001. A village's population as per the Population Census of 2001 is used as its weight.
  - **Ratio of GP's villages with access to bank:** Weighted average of the number of villages with access to a bank as per the Population Census of 2001. A village's population as per the Population Census of 2001 is used as its weight.
  - **Ratio of GP's villages with access to communication facilities:** Weighted average of the number of villages with access to communication facilities as per the Population Census of 2001. A village's population as per the Population Census of 2001 is used as its weight.
  - **Ratio of GP's villages with access to medical facilities:** Weighted average of the number of villages with access to medical facilities as per the Population Census of 2001. A village's population as per the Population Census of 2001 is used as its weight.
  - **Distance between GP and nearest town:** Weighted average of the distance of each village in a GP to the nearest town. A village's population as per the Population Census of 2001 is used as its weight.
  - **Ratio of GP's villages with access to maternity and child welfare centers:** Weighted average of the number of villages with a maternity and child welfare center

as per the Population Census of 2001. A village's population as per the Population Census of 2001 is used as its weight.

  - **Ratio of GP's villages with primary school:** Weighted average of the number of villages with a primary school as per the Population Census of 2001. A village's population as per the Population Census of 2001 is used as its weight.

  - **Share of Scheduled Caste/Scheduled Tribe population in GP's total population:** Number of SC/ST people in a GP divided by the GP's total population, as per the Population Census of 2001.

  - **Share of illiterate population in GP's total population:** Number of illiterate people in a GP divided by the GP's total population, as per the Population Census of 2001.

  - **Ratio of GP's irrigated to total farm land:** Total irrigated land in a GP divided by the total farmed land in the GP, as per the Population Census of 2001.

  - **Night Lights:** Square root of the weighted average of villages' total night lights from 2001, where a village's 2001 population is its weight.

- **Distance between GP and Sub-District Headquarters:** Weighted average of the distance of each village in a GP to its sub-district headquarters. A village's population as per the Population Census of 2001 is used as its weight.

B.2 About the Instrument and GP Composition Fixed Effects

As mentioned in the main text, I use the number of just eligible villages (as per 2001 Census) as instrument for the endogenous regressor, i.e. ratio of villages in GP covered by PMGSY scheme. As per the national government's guidelines, three thresholds based on village/habitation population were outlined for prioritization of road construction, i.e. 1,000 people, 500 people and 250 people (priority in decreasing order). However, these thresholds were not adhered to uniformly across the country. The thresholds adhered to in the sample of my states are 500 and 1,000. Among the four states, the following thresholds were adhered to (Asher and Novosad 2020).

- Gujarat (only 500)
- Rajasthan (only 500)
- Odisha (only 500)
- Madhya Pradesh (500 and 1000)

For the benchmark case, I use a bandwidth of +/- 50 people in order to determine whether villages are just eligible or just ineligible. For the first three states, the situation is straightforward, i.e., the number of villages in a GP with population 500-550 people is used to predict the overall proportion of villages in GP connected by PMGSY. However, for Madhya Pradesh, the number of villages in a GP with population 500-550 as well as 1000-1050 is used to predict the overall proportion of villages in GP connected by PMGSY.

Note that the final causal effect is determined from the difference observed between compliers on either side of the cut-off, i.e. GPs where just eligible villages get PMGSY vs GPs where just ineligible villages do not get PMGSY. Therefore, in order to take care of any heterogeneity in governance that might come about across GPs with different levels of the 500-threshold and 1000-threshold villages, I use GP bandwidth composition fixed effects. For instance, all

GPs are put into categories that suggest the *total number of villages* (just eligible and just ineligible) within bandwidth around the 500 point (450-550) and the 1000 point (950-1050). For states apart from Madhya Pradesh, all GPs fall in one of the following catgeories: one 500-bandwidth village and none 1000-bandwidth village, two 500-bandwidth villages and none 1000-bandwidth village, three 500-bandwidth villages and none 1000-bandwidth village and so on. For Madhya Pradesh, the categories are more complex, i.e., one 500-bandwidth village and none 1000-bandwidth village, one 500-bandwidth village and one 1000-bandwidth village, two 500-bandwidth villages and one 1000-bandwidth village, two 500-bandwidth villages and two 1000-bandwidth villages and so on. In my specifications, therefore, I control for fixed effects for these categories.

B.3 Placebo Checks

Table 40: Placebo Checks on Baseline Covariates (First Stage, Bandwidth=50)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Proportion of Road pre-2001	Average Power Access	Average Bank Access	Average Communication Access	Average Medical Facility Access	Average Distance to Nearest Town	Average MCW Center Access	Average Primary School Access	Proportion of SC/ST	Illiteracy Rate	Proportion of Irrigated Area	Square Root of Average Night Lights
No. of Villages Just Eligible	0.005	0.004	-0.001	0.007	0.008	0.250	-0.000	0.000	-0.004	0.002	0.003	-0.025
	(0.391)	(0.362)	(0.644)	(0.284)	(0.256)	(0.440)	(0.875)	(0.992)	(0.279)	(0.665)	(0.501)	(0.502)
N	6445	6445	6445	6445	6445	6445	6445	6445	6445	6445	6445	6445
Control Mean	.21	.92	.04	.29	.31	23.58	.03	.93	.46	.55	.31	2.73

Robust standard errors in parentheses

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Simple OLS regression used with fixed effects for districts and composition of villages within bandwidth in GP. For each outcome variable, controls include number of villages, area of GP and all the other outcome variables. Robust Standard errors used.

Table 41: Placebo Checks (2SLS Estimates, Bandwidth=50)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Proportion of Road pre-2001	Average Power Access	Average Bank Access	Average Communication Access	Average Medical Facility Access	Average Distance to Nearest Town	Average MCW Center Access	Average Primary School Access	Proportion of SC/ST	Illiteracy Rate	Proportion of Irrigated Area	Square Root of Average Night Lights
PMGSY Village Coverage	0.083	0.062	-0.023	0.118	0.133	4.296	-0.007	0.000	-0.075	0.030	0.053	-0.433
	(0.396)	(0.360)	(0.640)	(0.285)	(0.251)	(0.435)	(0.874)	(0.992)	(0.278)	(0.661)	(0.497)	(0.498)
N	6445	6445	6445	6445	6445	6445	6445	6445	6445	6445	6445	6445
Effective F-Statistic	73.38	78.08	78.05	76.29	78.61	78.37	78.10	77.64	77.8	78.08	77.88	78.04
Control Mean	.3	.94	.04	.34	.31	21.56	.02	.94	.43	.54	.33	3.03

Robust standard errors in parentheses

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Two stage least squares estimation results wherein ratio of villages with PMGSY in GP is instrumented by number of just eligible villages. Specifications include fixed effects for districts and composition of villages within bandwidth in GP. For each outcome variable, controls include number of villages and all the other outcome variables. Robust Standard errors used.

B.4 Road Quality and Distance from Sub-District HQ

In this section, I regress a variable that equals one if a road built under PMGSY (in a GP) is made of bitumen or concrete (the preferred materials under PMGSY) on the distance between the connected GP and district headquarters, and sub-district headquarters.

Table 42: Road Quality by Distance to Sub-District HQ

	(1)
Distance from Sub-Distance HQ (km)	0.00045
	(0.00043)
N	3733

Robust Standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

B.5 Robustness Check (Different Bandwidths)

Table 43: Robustness Checks for Ratio of Transactions within 15 days

	(1) Bandwidth=50 (Benchmark)	(2) Bandwidth=30	(3) Bandwidth=40	(4) Bandwidth=60	(5) Bandwidth=70
PMGSY Village Coverage	0.025	0.105	0.094	0.034	0.015
	(0.078)	(0.100)	(0.086)	(0.069)	(0.064)
Control Mean	.27	.27	.27	.27	.27
Effective F-Stat	77.99	46.22	64.09	98.01	115.34
N	6445	4221	5392	7366	8309

Robust Standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 44: Robustness Checks for Ratio of Transactions within 16-30 days

	(1) Bandwidth=50 (Benchmark)	(2) Bandwidth=30	(3) Bandwidth=40	(4) Bandwidth=60	(5) Bandwidth=70
PMGSY Village Coverage	0.129*	0.167*	0.151*	0.078	0.068
	(0.070)	(0.089)	(0.077)	(0.063)	(0.058)
Control Mean	.27	.26	.27	.27	.27
Effective F-Stat	77.99	46.22	64.09	98.01	115.34
N	6445	4221	5392	7366	8309

Robust Standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 45: Robustness Checks for Ratio of Transactions within 31-60 days

	(1) Bandwidth=50 (Benchmark)	(2) Bandwidth=30	(3) Bandwidth=40	(4) Bandwidth=60	(5) Bandwidth=70
PMGSY Village Coverage	-0.125*	-0.127	-0.122	-0.110*	-0.057
	(0.071)	(0.091)	(0.078)	(0.063)	(0.058)
Control Mean	.24	.24	.23	.24	.24
Effective F-Stat	77.99	46.22	64.09	98.01	115.34
N	6445	4221	5392	7366	8309

Robust Standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 46: Robustness Checks for Ratio of Transactions within 61-90 days

	(1) Bandwidth=50 (Benchmark)	(2) Bandwidth=30	(3) Bandwidth=40	(4) Bandwidth=60	(5) Bandwidth=70
PMGSY Village Coverage	0.007	-0.003	-0.009	-0.006	-0.005
	(0.048)	(0.062)	(0.053)	(0.044)	(0.041)
Control Mean	.1	.1	.1	.1	.1
Effective F-Stat	77.99	46.22	64.09	98.01	115.34
N	6445	4221	5392	7366	8309

Robust Standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 47: Robustness Checks for Ratio of Transactions beyond 90 days

	(1) Bandwidth=50 (Benchmark)	(2) Bandwidth=30	(3) Bandwidth=40	(4) Bandwidth=60	(5) Bandwidth=70
PMGSY Village Coverage	-0.036	-0.142	-0.114	0.003	-0.021
	(0.065)	(0.087)	(0.072)	(0.058)	(0.054)
Control Mean	.12	.12	.13	.12	.12
Effective F-Stat	77.99	46.22	64.09	98.01	115.34
N	6445	4221	5392	7366	8309

Robust Standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 48: Robustness Checks for Ratio of Amount Processed within 15 days

	(1) Bandwidth=50 (Benchmark)	(2) Bandwidth=30	(3) Bandwidth=40	(4) Bandwidth=60	(5) Bandwidth=70
PMGSY Village Coverage	0.026	0.104	0.092	0.040	0.022
	(0.078)	(0.100)	(0.086)	(0.070)	(0.064)
Control Mean	.27	.27	.27	.27	.27
Effective F-Stat	78.24	46.49	64.33	98.03	115.13
N	6444	4220	5391	7365	8307

Robust Standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 49: Robustness Checks for Ratio of Amount Processed within 16-30 days

	(1) Bandwidth=50 (Benchmark)	(2) Bandwidth=30	(3) Bandwidth=40	(4) Bandwidth=60	(5) Bandwidth=70
PMGSY Village Coverage	0.154**	0.183**	0.174**	0.097	0.085
	(0.071)	(0.090)	(0.078)	(0.063)	(0.059)
Control Mean	.27	.27	.27	.27	.27
Effective F-Stat	78.24	46.49	64.33	98.03	115.13
N	6444	4220	5391	7365	8307

Robust Standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 50: Robustness Checks for Ratio of Amount Processed within 31-60 days

	(1) Bandwidth=50 (Benchmark)	(2) Bandwidth=30	(3) Bandwidth=40	(4) Bandwidth=60	(5) Bandwidth=70
PMGSY Village Coverage	-0.146**	-0.140	-0.140*	-0.125*	-0.080
	(0.071)	(0.091)	(0.079)	(0.064)	(0.058)
Control Mean	.24	.24	.23	.24	.24
Effective F-Stat	78.24	46.49	64.33	98.03	115.13
N	6444	4220	5391	7365	8307

Robust Standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 51: Robustness Checks for Ratio of Amount Processed within 61-90 days

	(1) Bandwidth=50 (Benchmark)	(2) Bandwidth=30	(3) Bandwidth=40	(4) Bandwidth=60	(5) Bandwidth=70
PMGSY Village Coverage	0.001	-0.017	-0.019	-0.011	-0.005
	(0.048)	(0.062)	(0.053)	(0.045)	(0.041)
Control Mean	.1	.1	.1	.1	.1
Effective F-Stat	78.24	46.49	64.33	98.03	115.13
N	6444	4220	5391	7365	8307

Robust Standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 52: Robustness Checks for Ratio of Amount Processed beyond 90 days

	(1) Bandwidth=50 (Benchmark)	(2) Bandwidth=30	(3) Bandwidth=40	(4) Bandwidth=60	(5) Bandwidth=70
PMGSY Village Coverage	-0.036	-0.132	-0.108	-0.002	-0.023
	(0.065)	(0.085)	(0.072)	(0.058)	(0.054)
Control Mean	.12	.12	.12	.12	.12
Effective F-Stat	78.24	46.49	64.33	98.03	115.13
N	6444	4220	5391	7365	8307

Robust Standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking, medical and educational facilities, night lights, access to power, road access pre-2001 and distance to nearest town with village population being used as weights.

B.6 Single Habitation Villages Only

Table 53: 2SLS Estimates with Number of Just Eligible Villages Within Bandwidth as IV (Bandwidth=50)

	(1) Ratio of Transactions Within 15 days	(2) Ratio of Transactions 16-30 days	(3) Ratio of Transactions 31-60 days	(4) Ratio of Transactions 61-90 days	(5) Ratio of Transactions ¿90 days
PMGSY Village Coverage	0.043	0.157**	-0.135*	-0.052	-0.012
	(0.083)	(0.076)	(0.079)	(0.056)	(0.077)
Control Mean	.24	.26	.25	.11	.14
Kleibergen-Paap F-Stat	55.96	55.96	55.96	55.96	55.96
Cragg-Donald F-Stat	54.46	54.46	54.46	54.46	54.46
Effective F-Stat	55.96	55.96	55.96	55.96	55.96
N	4064	4064	4064	4064	4064

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Sample includes GPs with only single habitation villages in the states of Rajasthan, Gujarat, Odisha and Madhya Pradesh. Outcome variable is the percentage of transactions processed within given time periods. Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking and educational facilities, night lights, road access pre-2001 and distance to nearest town with village population being used as weights.

Table 54: 2SLS Estimates with Number of Just Eligible Villages Within Bandwidth as IV (Bandwidth=50)

	(1) Ratio of Amount Within 15 days	(2) Ratio of Amount 16-30 days	(3) Ratio of Amount 31-60 days	(4) Ratio of Amount 61-90 days	(5) Ratio of Amount ¿90 days
PMGSY Village Coverage	0.055	0.164**	-0.149*	-0.059	-0.009
	(0.083)	(0.077)	(0.079)	(0.056)	(0.077)
Control Mean	.24	.26	.25	.11	.14
Kleibergen-Paap F-Stat	56.17	56.17	56.17	56.17	56.17
Cragg-Donald F-Stat	54.66	54.66	54.66	54.66	54.66
Effective F-Stat	56.17	56.17	56.17	56.17	56.17
N	4063	4063	4063	4063	4063

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Sample includes GPs with only single habitation villages in the states of Rajasthan, Gujarat, Odisha and Madhya Pradesh. Outcome variable is the percentage of amount processed within given time periods. Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking and educational facilities, night lights, road access pre-2001 and distance to nearest town with village population being used as weights.

B.7 Impact of Roads on Volume of MGNREGA Transactions and Amount)

Table 55: Treatment Effect on MGNREGA Transactions (Bandwidth=50)

	Log Number of Transactions				Log Amount			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Tertile-1	Tertile-2	Tertile-3	All	Tertile-1	Tertile-2	Tertile-3
PMGSY Village Coverage	0.301	0.432	0.049	0.336	0.307	0.254	-0.138	0.517
	(0.363)	(0.513)	(0.615)	(0.662)	(0.377)	(0.508)	(0.634)	(0.692)
Control Mean	988.26	916.62	1040.7	1017.79	8.88	8.43	9.27	9.01
Effective F-Stat	61.53	18.34	14.26	12.9	61.65	18.54	14.26	12.9
N	6445	2110	2313	1961	6444	2109	2313	1961

Robust Standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

Control mean for amount of money processed in Columns (5)-(8) is in INR lakhs (00,000). Controls include number of villages, and fixed effects for districts and composition of number of villages within the bandwidth (on either side). Baseline GP-level controls (2001) include population literacy levels, proportion of irrigated land, proportion of SC/ST population and weighted averages of village-level indicators such as banking and educational facilities, night lights, road access pre-2001 and distance to nearest town with village population being used as weights.

C Appendix: Chapter 3

C.1 States in Sample and Variable Definitions

C.1.1 States in Sample

- * For Delhi, I consider 2013 and 2015 because of a peculiar scenario where owing to an unclear mandate, Delhi, after the assembly election in late 2013, witnessed President's rule for almost a year and a fresh round of polls was conducted in early 2015.

- ** Because of Jammu and Kashmir's erstwhile special status as a state, elections used to be held every 6 years there.

- *** The state of Telangana was created in 2014. However, pooling the districts that now comprise Telangana, I used it as a state even in 2009.

	Year "t"	Year "t+1"
Andhra Pradesh	2009	2014
Assam	2001 2011	2006 2016
Bihar	2010	2015
Delhi*	2013	2015
Goa	2002 2012	2007 2017
Haryana	2000 2009	2005 2014
Jammu & Kashmir**	2008	2014
Jharkhand	2009	2014
Karnataka	2008	2013
Kerala	2001 2011	2006 2016
Maharashtra	2009	2014
Odisha	2009	2014
Puducherry	2001	2006
Punjab	2002 2012	2007 2017
Tamil Nadu	2001 2011	2006 2016
Telangana***	2009	2014
Uttar Pradesh	2002 2012	2007 2017
West Bengal	2001 2011	2006 2016

Table 56: Description of Sample

C.1.2 Control Variables

- **National Party=INC?:** Whether the national party in question is the Indian National Congress (INC) or not. If it is the BJP (the second of the 2 national parties), this variable equals zero.

- **Winner SC/ST (t)?:** Whether the winning candidate at baseline (t) belonged to socially disadvantaged groups (Scheduled caste or scheduled tribe).

- **Runner-up SC/ST (t)?:** Whether the runner-up at baseline (t) belonged to socially disadvantaged groups (Scheduled caste or scheduled tribe).

- **Winner female (t)?:** Whether the winning candidate at baseline (t) was a female or not.

- **Runner-up female (t)?:** Whether the runner-up candidate at baseline (t) was a female or not.

- **Turnout percentage (t):** Percentage of total voters who came out to vote during election cycle e_t.

- **Effective number of parties (t):** Effective number of parties derived from vote shares of each candidate in a constituency.

- **Number of candidates (t):** Number of candidates contesting the election during e_t.

- **SC Reserved constituency?:** Whether the constituency is reserved for representatives belonging to scheduled castes or not.

- **Elevation:** Average elevation of the constituency. Elevation data is from SRTM (Shuttle Radar Topography Mission) at 1-arc second resolution. Its Digital Object Identifier (DOI) number is /10.5066/F7PR7TFT.

- **Ruggedness:** Average ruggedness of the constituency. The ruggedness measure TRI is constructed from the elevation data using the method from Riley et al (1999), and used by Nunn and Puga (2012) "Ruggedness: The Blessing of Bad Geography in Africa". The TRI expresses the elevation difference between adjacent pixels. It is the square root of the mean squared difference in elevation values between a pixel and the eight cells immediately surrounding it.

- **Percentage employed (1998):** Percentage of electors employed in economic establishments (except government administration, national defense and agriculture) in a constituency.

- **Winner's age (t):** Age of the winner during election cycle e_t.

- **Runner-up's age(t):** Age of the runner-up during election cycle e_t.

- **Number of electors (t):** Total number of electors in a constituency (those who voted plus those who did not) during election cycle e_t.

References

[1] Abadie, A., Athey, S., Imbens, G.W., Wooldridge, J., 2017. When Should You Adjust Standard Errors for Clustering? Technical Report. National Bureau of Economic Research.

[2] Abramowitz, A., 1988. Explaining Senate Election Outcomes. American Political Science Review , 385–403.

[3] Abramowitz, A., 1991. Incumbency, Campaign Spending, and the Decline of Competition in the U.S. House. Journal of Politics , 34–57.

[4] Acemoglu, D., Reed, T., Robinson, J.A., 2014. Chiefs: Economic Development and Elite Control of Civil Society in Sierra Leone. Journal of Political Economy 122, 319–68.

[5] Acemoglu, D., Robinson, J.A., 2006. Economic Backwardness in Political Perspective. American Political Science Review 100, 115–131.

[6] Adukia, A., Asher, S., Novosad, P., 2020. Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction. American Economic Journal: Applied Economics , 884–907.

[7] Aggarwal, S., 2018. Do rural roads create pathways out of poverty? Evidence from India. Journal of Development Economics , 375–395.

[8] Ahmad, E., 1952. Rural Settlement Types in the Uttar Pradesh (United Provinces of Agra and Oudh). Annals of the Association of American Geographers 42, 223–246.

[9] Algan, Y., Cahuc, P., 2010. Inherited Trust and Growth. American Economic Review 100, 2060–2092.

[10] Anderson, M.L., 2008. Multiple Inference and Gender Differences in the Effects of Early

Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. Journal of the American statistical Association 103.

[11] Anderson, S., 2011. Caste as an Impediment to Trade. American Economic Journal: Applied Economics 3, 239–63.

[12] Anderson, S., Francois, P., Kotwal, A., 2015. Clientelism in Indian villages. American Economic Review 105, 1780–1816.

[13] Andrews, I., Stock, J., Sun, L., 2018. Weak Instruments and What to do About Them, in: Summer Institute 2018 Methods Lectures, NBER, National Bureau of Economic Research.

[14] Asher, S., Lunt, T., Matsuura, R., Novosad, P., 2021. Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the shrug open data platform. The World Bank Economic Review 35, 845–871.

[15] Asher, S., Nagpal, K., Novosad, P., 2018. The Cost of Distance: Geography and Governance in Rural India. World Bank Working Paper .

[16] Asher, S., Novosad, P., 2019. Socioeconomic High-resolution Rural-Urban Geographic Dataset for India (SHRUG). Harvard Dataverse doi:`https://doi.org/10.7910/DVN/DPESAK`.

[17] Asher, S., Novosad, P., 2020. Rural Roads and Local Economic Development. American Economic Review , 797–823.

[18] Ashraf, N., Bandiera, O., Jack, B., 2014. No margin, no mission? A field experiment on incentives for public service delivery. Journal of Public Economics , 1–17.

[19] Baden-Powell, B.H., 1892. The Land-Systems of British India. volume 2. Clarendon Press.

[20] Baden-Powell, B.H., 1894. A Short Account of the Land Revenue and Its Administration in British India.

[21] Banerjee, A., Duflo, E., Imbert, C., Mathew, S., Pande, R., 2020. E-governance, Accountability, and Leakage in Public Programs: Experimental Evidence from a Financial Management Reform in India. American Economic Journal: Applied Economics , 34–57.

[22] Banerjee, A., Iyer, L., 2005. History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India. American Economic Review 95, 1190–1213.

[23] Banerjee, A., Iyer, L., Somanathan, R., 2005. History, Social Divisions, and Public Goods in Rural India. Journal of the European Economic Association 3, 639–647.

[24] Bardhan, P., Mookherjee, D., 2011. Subsidized Farm Input Programs and Agricultural Performance: A Farm-Level Analysis of West Bengal's Green Revolution, 1982-1995. American Economic Journal: Applied Economics 3, 186–214.

[25] Besley, T., Burchardi, K., Ghatak, M., 2012. Incentives and the De Soto Effect. The Quarterly Journal of Economics 127, 237–282.

[26] Besly, T., Burgess, R., 2002. The Political Economy of Government Responsiveness: Theory and Evidence from India. The Quarterly Journal of Economics 117, 1415–1451.

[27] Bhalotra, S., Clots-Figueras, I., 2014. Health and the Political Agency of Women. American Economic Journal: Economic Policy , 615–635.

[28] Björkman, M., Svensson, J., 2009. Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda. The Quarterly Journal of Economics 124, 735–769.

[29] Bozcaga, T., 2020. The Social Bureaucrat: How Social Proximity among Bureaucrats Affects Local Governance. Middle East Initiative .

[30] Bryceson, D.F., Bradbury, A., Bradbury, T., 2008. Roads to Poverty Reduction? Exploring Rural Roads' Impact on Mobility in Africa and Asia. Development Policy Review 11.

[31] Calonico, S., Cattaneo, M.D., Farrell, M.H., 2020. Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. The Econometrics Journal 23, 192–210.

[32] Cattaneo, M.D., Idrobo, N., Titiunik, R., 2019. A Practical Introduction to Regression Discontinuity Designs: Foundations. Cambridge University Press.

[33] Chanana, P., 2013. Caste Ties, Allodial Rights and Colonial Administration in Oudh during Summary Settelement and after 1858: The 'Taluqdar' Association and the British Policies upto 1870's. Proceedings of the Indian History Congress 74, 458–469.

[34] Chauchard, S., Klašnja, M., Harish, S., 2019. Getting Rich Too Fast? Voters' Reactions to Politicians' Wealth Accumulation. The Journal of Politics 81.

[35] Conley, T.G., 2010. Spatial Econometrics. Palgrave Macmillan UK, London. pp. 303–313.

[36] Davis, J., 2004. Corruption in public service delivery: Experience from South Asia's water and sanitation sector. World Development 32, 53–71.

[37] Deliège, R., 1999. The Untouchables of India. Berg Publishers.

[38] Dell, M., 2010. The Persistent Effects of Peru's Mining Mita. Econometrica 78, 1863–1903.

[39] Dell, M., Lane, N., Querubin, P., 2018. The Historical State, Local Collective Action, and Economic Development in Vietnam. Econometrica 86, 2083–2121.

[40] Dell, M., Olken, B.A., 2020. The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java. The Review of Economic Studies 87, 164–203.

[41] Donaldson, D., 2018. Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. American Economic Review 108, 899–934.

[42] Dotsch, R., Wigboldus, D.H., Langner, O., Van Knippenberg, A., 2008. Ethnic out-group faces are biased in the prejudiced mind. Psychological science 19, 978–980.

[43] Dovidio, J.F., Brigham, J.C., Johnson, B.T., Gaertner, S.L., Macrae, C.N., Stangor, C., Hewstone, M., 1996. Stereotypes and Stereotyping. New York: Guilford , 276–319.

[44] Duflo, E., Hanna, R., Ryan, S.P., 2012. Incentives Work: Getting Teachers to Come to School. American Economic Review 102, 1241–1278.

[45] Dutt, Romesh C., 1900. Open Letters to Lord Curzon on Famines and Land Assessments in India. Kegan Paul, Trench, Trubner and Co. Ltd.

[46] Engerman, S.L., Sokoloff, K.L., 1997. Factor Endowments,Institutions,and Differential Paths of Growth Among New World Economies: A View from Economic Historians of the United States, in: Haber, S. (Ed.), How Latin America Fell Behind. Stanford University Press, Stanford.

[47] Faber, B., 2014. Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. Review of Economic Studies 81, 1046–70.

[48] Faguet, J.P., Sánchez, F., Villaveces, M.J., 2020. The Perversion of Public Land Distribution by Landed Elites: Power, Inequality and Development in Colombia. World Development 136.

[49] Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas. International journal of climatology 37, 4302–4315.

[50] Finan, F., Sadouleth, E., de Janvry, A., 2005. Measuring the poverty reduction potential of Land in rural Mexico. Journal of Development Economics 77, 27–51.

[51] Foster, A.D., Rosenzweig, M.R., 2004. Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970–2000. Economic Development and Cultural Change 52, 509–542.

[52] Gage, A.J., Calixte, M.G., 2007. Effects of the physical accessibility of maternal health services on their use in rural Haiti. Population Studies , 271–288.

[53] Gelman, A., Imbens, G., 2019a. Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. Journal of Business and Economic Statistics .

[54] Gelman, A., Imbens, G., 2019b. Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. Journal of Business and Economic Statistics 37, 447–56.

[55] Genicot, Garance and Ray, Debraj, 2017. Aspirations and Inequality. Econometrica 85, 489–519.

[56] Gowda, R., Sridharan, E., 2012. Reforming India's Party Financing and Election Expenditure Laws. Election Law Journal 11.

[57] Green, D., Krasno, J., 1988. Salvation for the Spendthrift Incumbent: Reestimating the Effects of Campaign Spending in House Elections. American Journal of Political Science , 884–907.

[58] Imbens, G., Lemieux, T., 2008. Regression Discontinuity Designs: A Guide to Practice. Journal of Econometrics 142, 615–35.

[59] Iversen, V., Kalwij, A., Verschoor, A., Dubey, A., 2014. Caste Dominance and Economic Performance in Rural India. Economic Development and Cultural Change 62, 423–457.

[60] Iyer, L., 2010. Direct versus Indirect Colonial Rule in India: Long-term consequences. The Review of Economics and Statistics 92, 693–713.

[61] Jacobson, G., 1990. The Effects of Campaign Spending in House Elections: New Evidence for Old Arguments. American Journal of Political Science 9, 321–350.

[62] Jacoby, H.G., 2000. Access to Markets and the Benefits of Rural Roads. The Economic Journal 110.

[63] Jensenius, F.R., Verniers, G., 2017. Studying Indian politics with large-scale data: Indian election data 1961–today. Studies in Indian Politics .

[64] Joppen, C., 1914. Historical Atlas of India: for the use of high schools, colleges, and private students. Longmans, Green, & Company.

[65] Kanuganti, S., Sarkar, A.K., Singh, A.P., Arkatkar, S., 2015. Quantification of accessibility to health facilities in rural areas. Case Studies on Transport Policy , 311–320.

[66] Kelly, Morgan, 2019. The Standard Errors of Persistence. Available at SSRN: https://ssrn.com/abstract=3398303.

[67] Keswell, M., Carter, M., 2014. Poverty and Land Redistribution. Journal of Development Economics 110, 250–261.

[68] Khandker, S.R., Bakht, Z., Koolwal, G.B., 2009. The Poverty Impact of Rural Roads: Evidence from Bangladesh. Economic Development and Cultural Change 57, 685–722.

[69] Knack, S., Keefer, P., 1997. Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. The Quarterly Journal of Economics 112, 1251–1288.

[70] Kubota, J.T., Li, J., Bar-David, E., Banaji, M.R., Phelps, E.A., 2013. The Price of Racial Bias: Intergroup Negotiations in the Ultimatum Game. Psychol Sci. 24, 2498–2504.

[71] Kumar, D., 1982. The Cambridge Economic History of India: c.1757-c.1970. volume 2. Cambridge: Cambridge University Press.

[72] Lee, A., 2019. Land, State Capacity, and Colonialism: Evidence from India. Comparative Political Studies 52, 412–444.

[73] Lee, D., 2008. Randomized experiments from non-random selection in US House elections. Journal of Econometrics , 675–697.

[74] Lee, D., Moretti, E., Butler, M.J., 2004. Do Voters Affect or Elect Policies? Evidence from the U.S. House. Quarterly Journal of Economics 119, 807–59.

[75] Lee, D.S., McCrary, J., Moreira, M.J., Porter, J., 2022. Valid t -Ratio Inference for IV. American Economic Review 112.

[76] Lipton, M., 2009. Land Reform in Developing Countries. Routledge.

[77] Lowes, S., Montero, E., 2021. The Legacy of Colonial Medicine in Central Africa. American economic review 111, 1284–1314.

[78] Mahajan, V., 2020. Modern Indian History. S. Chand Publishing.

[79] Marfatia, A., 2019. Bridging the last mile connectivity gap. URL: https://idronline.org/bridging-the-last-mile-delivery-gap/.

[80] Meiyappan, P., Roy, P., Soliman, A., Li, T., Mondal, P., Wang, S., Jain, A., 2018. India Village-Level Geospatial Socio-Economic Data Set: 1991, 2001. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

[81] Mendola, M., Simtowe, F., 2015. The Welfare Impact of Land Redistribution: Evidence from a Quasi-Experimental Initiative in Malawi. World Development 72, 53–69.

[82] Mohanty, B.B., 2001. Land Distribution among Scheduled Castes and Tribes. Economic and Political Weekly , 3857–3868.

[83] Mu, R., van de Walle, D., 2011. Rural Roads and Local Market Development in Vietnam. Journal of Development Studies , 709–734.

[84] Muralidharan, K., Niehaus, P., Sukhtankar, S., 2016. Building State Capacity: Evidence from Biometric Smartcards in India. American Economic Review 106, 2895–2929.

[85] Muralidharan, K., Sundararaman, V., 2011. Teacher Performance Pay: Experimental Evidence from India. Journal of Political Economy .

[86] Narayanan, R., Dhorajiwala, S., Golani, R., 2019. Analysis of Payment Delays and Delay

Compensation in MGNREGA: Findings Across Ten States for Financial Year 2016–2017. The Indian Journal of Labour Economics 62, 113–133.

[87] Newell, R.S., 1972. Ideology and Realities: Land Redistribution in Uttar Pradesh. Pacific Affairs 45, 220–239.

[88] Nunn, N., 2008. The Long-Term Effects of Africa's Slave Trades. The Quarterly Journal of Economics 123, 139–176.

[89] Nunn, N., Wantchekon, L., 2011. The Slave Trade and the Origins of Mistrust in Africa. American Economic Review 101, 639–647.

[90] Olken, B.A., 2007. Monitoring Corruption: Evidence from a Field Experiment in Indonesia. Journal of Political Economy 115.

[91] Pandey, P., 2010. Service Delivery and Corruption in Public Services: How Does History Matter? American Economic Journal: Applied Economics 2, 190–204.

[92] Porter, G., 2002. Living in a Walking World: Rural Mobility and Social Equity Issues in Sub-Saharan Africa. World Development 9, 321–350.

[93] Prasad, P.H., 1979. Caste and Class in Bihar. Economic and Political Weekly , 481–484.

[94] Press Information Bureau, 2011. Status of MGNREGS in Jharkhand. URL: `https://pib.gov.in/newsite/PrintRelease.aspx?relid=70360`. Accessed on July 15, 2023.

[95] Press Information Bureau, 2020. Land under Irrigation. URL: `https://pib.gov.in/newsite/PrintRelease.aspx?relid=199881`. Accessed on July 15, 2023.

[96] Rasul, I., Rogger, D., 2015. The Impact of Ethnic Diversity in Bureaucracies: Evidence from the Nigerian Civil Service. American Economic Review , 457–461.

[97] Spears, D., 2015. Effects of Sanitation on Early-Life Health: Evidence from a Governance Incentive in Rural India. Unpublished .

[98] Stokes, E., 1959. The English Utilitarians and India .

[99] Stokes, E., 1983. Northern and Central India. Cambridge University Press. volume 2 of *The Cambridge Economic History of India*. p. 36–86.

[100] Tendler, J., 1997. Good government in the Tropics. Johns Hopkins University Press.

[101] The World Bank, 2020. The World Bank Data. https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS?locations=IN. Accessed: 2023-06-19.

[102] Tsai, L.L., 2007. Accountability without Democracy: Solidary Groups and Public Goods provision in rural China. Cambridge University Press.

[103] Vaishnav, M., 2011. The Market for Criminality: Money, Muscle and Elections in India. SSRN Electronic Journal doi:10.2139/ssrn.1928011.

[104] Varshney, A., 2019. Caste and Religion still form the bedrock of people's political understanding in Uttar Pradesh. The Indian Express .

[105] Verniers, G., Jaffrelot, C., 2020. The reconfiguration of India's political elite: profiling the 17th Lok Sabha. Contemporary South Asia , 242–254.

[106] Vishnoi, A., 2020. At Rs 1,450 cr, BJP got 61% funding via Electoral Bonds before LS polls. The Economic Times. URL: https://economictimes.indiatimes.com/news/politics-and-nation/at-rs-1450-cr-bjp-got-61-funding-via-electoral-bonds-before-ls-polls/articleshow/73181670.cms?from=mdr.