

Essays on Urbanisation and Governance in Developing Countries



Karan Nagpal
Jesus College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity Term 2018

Essays on Urbanisation and Governance in Developing Countries

KARAN NAGPAL

Jesus College
University of Oxford

Abstract

This thesis examines the role of the state in shaping economic outcomes in rapidly urbanising developing countries.

Increasing the effectiveness of the state is a major challenge facing most developing countries today. Chapter 1 focusses on one important factor that constrains the state's ability to provide public goods to all citizens: citizens' physical remoteness from their administrators. Using rich data on village India, and a spatial regression discontinuity design, this chapter shows that greater distance to administration reduces a village's access to public goods and worsens economic outcomes. Villages that are more remote from their administrators have fewer roads, schools, health centers and less irrigation. In turn, their residents have lower incomes, fewer assets, less literacy, and are more likely to be employed in agriculture. At least for roads, these effects are not driven by the higher cost of construction in remote villages, but higher cost of monitoring road quality. These results suggest that reducing the distance between the state and its citizens, and changing incentives for bureaucrats, can help to mitigate the large spatial disparities in living standards observed within many developing countries.

In addition to hosting administrative headquarters, cities are also large goods and labour markets, and have been a major driver of recent employment growth in developing countries. Chapter 2 uses newly matched and geocoded data from India's Economic and Population Censuses to study how growing cities affect the economic structure in neighbouring villages. Instrumenting for employment growth in cities using a Bartik (shift-share) instrument, the chapter finds that 10% higher growth in manufacturing employment in the nearest city reduces the growth in manufacturing employment in the village by 1.8%, though it does not affect employment growth in other rural non-farm sectors. The effects are stronger for villages located close to larger towns and cities. Further, the effects are stronger for villages that are less suited for manufacturing activities, such as more remote villages and villages that lack paved roads.

Finally, Chapter 3 examines the effect of property titles on housing investment decisions of low-income urban households, using data from Delhi's resettlement localities. It finds that more secure property titles are not associated with a higher probability of engaging in housing investment, with higher levels of housing investment, or indeed with measures of housing quality. The results are robust to using matching estimators. The null results can be explained by the by the limited differences in perceived eviction risk between titled and non-titled households, and by low reliance on the formal

financial sector by all households. These results suggest that urban property titling programs in developing countries can be more successful if they also expand access to complementary inputs, such as the formal financial sector.

Supervisors This thesis was co-supervised by Prof. Douglas Gollin and Prof. Beata Javorcik.

Total words This thesis contains approximately 35,000 words.

Statement of authorship Chapter 1 is coauthored with Sam Asher (The World Bank) and Paul Novosad (Dartmouth College). I am the primary author of this chapter and am responsible for most of the conceptualization, all of the analysis, and all of the writing. The initial idea for the research question emerged in discussions between all three authors. I developed the intuition for the identification strategy, and Paul Novosad suggested ways to formalise it. Sam Asher and Paul Novosad provided the administrative datasets used in this paper. I researched the context of public administration in India, and designed the empirical analyses for testing the research questions in the paper. I wrote code in Stata, QGIS and Python to implement the empirical analysis. Further, I prepared the conceptual framework for testing different mechanisms using data from India's rural roads project. Finally, I presented the paper at several conferences to elicit feedback, and was responsible for writing the paper.

Chapter 2 is also coauthored with Sam Asher (The World Bank) and Paul Novosad (Dartmouth College). I am the primary author of this chapter and am responsible for a substantial portion of the conceptualization, all of the analysis, and all of the writing. Sam Asher and Paul Novosad provided the initial ideas for the identification strategy, and the administrative datasets to answer the research question. I further developed the conceptual foundations for using the Bartik instrument, investigated different specifications, designed and implemented all the empirical analyses, and wrote the paper.

Chapter 3 is single-authored and is wholly my own work. It is based on my MPhil thesis, but significantly takes forward the empirical analysis and further develops the conceptual questions tested in the thesis.

Corrections This version of the thesis incorporates the minor corrections required for the award of the DPhil degree following the viva voce examination, held on 25 September 2018. The Examiners of this thesis were Prof. Pramila Krishnan (University of Oxford) and Dr. Alexander Moradi (University of Sussex). The corrections were completed on 28 October 2018.

Statement of Originality

Chapters 1 and 2 are based on research carried out by me at the University of Oxford between October 2014 and August 2018. These chapters have not been submitted for a degree at this, or any other, university.

Chapter 3 of the thesis, “Property Titles and Housing Investment Among The Urban Poor,” was submitted in partial fulfilment of the requirements of the MPhil in Economics degree in Trinity Term 2012. At that time, the work was titled “Property rights and investment conditions in Delhi slum localities.” The work has changed significantly during the preparation of the DPhil.

Karan Nagpal

(August 2018)

Acknowledgements

I would like to express my sincere gratitude to my supervisors, James Fenske, Douglas Gollin and Beata Javorcik, for their constant encouragement and support during the DPhil years. I would also like to thank Francis Teal for setting me on this journey during the MPhil, and to Stefano Caria, Stefan Dercon, Julianne Labonne, Simon Quinn, and others at the Centre for the Study of African Economies for asking the right questions.

Finally, a big thank you to my co-authors and mentors, Sam Asher and Paul Novosad, for always believing in me, and for helping me become a better writer of do-files, and of economics papers.

CONTENTS

Introduction	1
1 The Cost of Distance: Geography and Governance in Rural India	8
1.1 Introduction	8
1.2 District administration in India	13
1.3 Data	15
1.3.1 Population Census	15
1.3.2 Socioeconomic and Caste Census	15
1.3.3 PMGSY data	16
1.3.4 GIS data	17
1.4 Empirical Strategy	17
1.4.1 Balance checks	21
1.5 Results	23
1.5.1 Roads	24
1.5.2 Other public amenities	25
1.5.3 Economic impacts	26
1.5.4 Robustness	28
1.6 Policy experiment	30
1.7 Mechanisms	31
1.8 Conclusion	33
Appendix 1.A Additional figures and tables	51
2 The Rural Effects of Urban Growth	67
2.1 Introduction	67
2.2 Data	72
2.3 Empirical Strategy	74
2.3.1 Constructing the Bartik instrument	76
2.4 Results	78
2.4.1 First stage	79
2.4.2 Main results	79
2.4.3 Effects on rural population outcomes	82
2.4.4 Effects by town size	83
2.4.5 Effects by village connectivity	84
2.5 Conclusion	85
3 Property Titles and Housing Investment Among the Urban Poor	98
3.1 Introduction	98
3.2 Context	104
3.3 Data	105
3.3.1 Treatment variable	106

3.3.2	Outcomes	106
3.3.3	Household characteristics	108
3.3.4	Financial information	109
3.4	Empirical Strategy	110
3.4.1	Linear regression	110
3.4.2	Matching estimators	113
3.4.3	Selection on unobserved variables	114
3.4.4	Power analysis	114
3.5	Results	115
3.5.1	Linear regression	115
3.5.2	Matching estimators	118
3.5.3	Selection on unobserved variables	120
3.5.4	Power analysis	121
3.6	Mechanisms	121
3.7	Conclusion	125
Appendix 3.A	Restricting sample to households that have lived in their current house for five years or longer	151
Appendix 3.B	Robustness to using log investment	159
	Concluding remarks	162
	Bibliography	166

INTRODUCTION

URBANISATION AND GOVERNANCE IN DEVELOPING COUNTRIES

Overview This thesis is composed of three paper-length essays that examine the role of the state in shaping economic outcomes in rapidly urbanising developing countries.

Over the last 25 years, urbanisation has accelerated globally, particularly in developing countries (UN Habitat, 2013). This rapid urban growth has led to demands for a commensurate improvement in institutional quality to better manage the urbanisation process (Glaeser, 2014; Brueckner and Lall, 2015), a sharper focus on urban poverty (Marx et al., 2013), and challenged traditional models of urban-rural linkages (Glaeser and Henderson, 2017; Jedwab et al., 2017). It has raised questions such as how governments can empower low-income urban households, and how urban centres influence the surrounding rural economy. This thesis focusses on these issues in the context of a large developing country – India – looking both across the whole country and also within specific cities. In particular, the chapters in this thesis answer the following questions:

1. How does greater distance from administrative centres constrain the state's ability to provide public goods and services?
2. What are the rural effects of urban employment growth?
3. Do stronger property titles in poor urban neighbourhoods lead to greater investment in improving housing quality?

In the rest of this introduction, I will provide a brief description of each chapter.

Chapter 1: The Cost of Distance: Geography and Governance in Rural India This chapter, co-authored with Sam Asher and Paul Novosad, was the basis of my Job Market Paper.

In many developing countries, proximity to urban centres plays a major role in shaping rural outcomes (Fafchamps and Shilpi, 2003; Fafchamps and Shilpi, 2005; Fafchamps and Wahba, 2006; Atkin and Donaldson, 2015; Gunning et al., 2018). One reason for the importance of urban proximity is that towns and cities usually possess large goods and labour markets. Because of frictions such as high transport costs (Feyrer, 2009; Gollin and Rogerson, 2014; Storeygard, 2016) and information asymmetries (Allen, 2014), distance to these markets is often associated with worse economic outcomes.

But in addition to possessing large goods and labour markets, some urban centres also serve as administrative centres, where public officials are based and where decisions about the delivery of public services are taken. Do frictions that restrict the flow of private economic activities over space also matter for the flow of public services over space? This question is particularly relevant in developing countries, where many rural citizens continue to lack access to public goods, despite governments' commitments to universal access (The World Bank, 2015). However, answering this question can be challenging. Because towns can be both administrative centres as well as important goods and labour markets, it is hard to disentangle the effect of remoteness from markets, from the effect of remoteness from administration. Secondly, answering this question requires fairly high-resolution economic data, which is often not available for developing countries.

In this chapter, we isolate the effect of remoteness from administration (or “administrative remoteness”) from other forms of remoteness by implementing a spatial regression discontinuity design. We compare villages located on opposite sides of ad-

ministrative borders in India. We show that these villages are fairly similar in terms of their distances to towns and cities, and share similar demographic and geographic characteristics. But because they are located in different administrative units, these villages can have sharply different distances to their administrative centres. We exploit this variation to investigate the costs of administrative remoteness in India.

To do this analysis, we assemble rich village-level data on incomes, assets, employment structure, and public goods covering the full universe of approximately 600,000 Indian villages. We combine this village-level data with extensive geographic data on the location of towns, administrative centres and administrative borders in India, which we use to construct measures of administrative remoteness for all Indian villages.

We find that administrative remoteness worsens development outcomes in village India. More administratively remote villages have reduced access to public goods provided by the relevant administrative centre. These public goods include paved roads, primary schools, and health centres. For example, administratively remote villages are 1.2 percentage points less likely to have paved roads, and 1 percentage point less likely to have a government primary school. But these villages are no less likely to receive public goods that are provided either directly by the federal government (such as national highways, post offices, or electricity distribution networks), or by local village councils (such as unpaved roads, or water and sanitation facilities). Administrative remoteness also leads to worse economic outcomes. In administratively remote villages, the share of the village workforce engaged in non-agricultural activities is 2.7% lower, average household incomes are about 1% lower, and a lower share of households have good quality housing and valuable assets.

Finally, we examine why administratively remote villages are less likely to receive public goods. Is it because constructing the public infrastructure is more expensive, or is it because monitoring the quality of the public infrastructure is more challenging in these villages? Using detailed project data from India's national rural roads

program¹, we show that paved feeder roads are not more expensive to construct in administratively remote villages, nor do they take longer to construct. However, once constructed, roads in administratively remote villages have a 9% lower probability of undergoing a quality audit. This reduced monitoring probability directly affects whether or not the road is judged as “paved” when census enumerators visit the village a few years later.

Chapter 2: The Rural Effects of Urban Growth This chapter is also co-authored with Sam Asher and Paul Novosad. It continues to examine the relationship between urban and rural areas in developing countries. But instead of focussing on the role of towns as administrative centres, it shifts attention to the importance of towns as large labour markets. In particular, it investigates the effect of exogenous growth in urban employment on the economic structure of neighbouring villages.

Traditionally, economists have thought about urban-rural linkages from the perspective of changes originating in rural areas (Lewis, 1954; Kuznets, 1957; Foster and Rosenzweig, 2007; Michaels et al., 2012). For example, rural agricultural productivity growth has been shown to drive the urbanisation process in different countries (Nunn and Qian, 2011; Bustos et al., 2016). However, with growing global trade in agriculture, and stronger links between developing world cities and global supply chains, the reliance of urban areas on the rural hinterland seems to have reduced (Glaeser, 2014). In such a world, urban growth is more often driven by changes in global and national demand patterns. This chapter studies the effect of such urban employment growth on the employment structure of neighbouring rural areas. More specifically, we evaluate the change in non-farm employment growth in villages when there are exogenous shocks to employment in nearby towns.

There are two challenges to answering this question. The first challenge is data: economic data are typically available at aggregate levels, covering multiple towns and

¹This program is officially called the Pradhan Mantri Gram Sadak Yojna.

their rural catchment areas. It is thus not possible to relate town-specific shocks to the immediate geography of the towns. The second challenge is causal identification: because towns and villages are located in close geographic proximity, it is difficult to isolate exogenous shocks to employment in towns alone.

To deal with the first challenge, we assemble panel data on employment in different industries in the universe of Indian towns and villages in 1998 and 2005. We match this to population census data from 2001 and 2011. Our data covers the universe of all non-farm firms and workers in India, and permit us to match each village to the employment structure and growth patterns of neighbouring towns and cities. For obtaining causal estimates, we construct exogenous shocks to urban employment by using a shift-share (or “Bartik”) instrument variable (Bartik, 1991; Blanchard et al., 1992). For every town, the Bartik instrument is constructed by interacting the baseline employment share of each manufacturing sub-sector, with the change in employment in that sub-sector in all *other* towns in the country.

We find that 10% higher exogenous growth in manufacturing employment in the nearest town reduces the growth rate of manufacturing employment in the village by 1.8%, but has no statistically significant effect on non-farm or service sector employment growth in the village. These results favour the hypothesis that forces of agglomeration increase the returns to manufacturing work in fast-growing towns and cities. Some manufacturing workers in nearby villages relocate to the town, thus driving labour exit from the rural manufacturing sector. These forces are weaker for the services sector, and therefore do not affect employment growth in the rural services sector. Moreover, because more than two-thirds of rural non-farm workers are employed in the services sector, the effects on non-farm employment growth are also not meaningful.

Further, the negative effects of urban employment growth are stronger for villages located close to larger towns and cities, which are precisely the places where we expect the agglomerative forces to be the strongest. The negative effects are also more

pronounced for villages that are particularly less suited for manufacturing industries in the first place, such as villages without paved roads and villages that are located farthest from towns and cities.

Chapter 3: Property Titles and Housing Investment Among the Urban Poor Moving beyond the study of urban-rural linkages, this chapter looks within a particular city, to investigate the effects of a specific policy tool on household outcomes. The policy tool in question is a formal property title. In recent decades, the idea of giving property rights to the urban poor in the form of formal titles has gained traction among economists and policy-makers (de Soto, 2000; Durand-Lasserve et al., 2007). This focus was supported by the initial empirical evidence from Latin America (Field, 2005; Field, 2007; Di Tella et al., 2007; Galiani and Schargrodsky, 2010). However, recent evidence suggests that these results may not generalise (Galiani and Schargrodsky, 2016; Gutierrez and Molina, 2016; Molina et al., 2018). Further, the success of property titles may depend on certain conditions, such as high levels of access to the formal financial sector and large differences in perceived tenure security between titled and non-titled households. Where these conditions are not met, property titles may have limited effects on the household outcomes that policy-makers want to improve.

In this chapter, I use data from 1,376 households in low-income neighbourhoods of Delhi to examine whether more secure property titles lead to higher rates of investment in improving housing quality. I find that more secure property titles are not associated with a higher probability of engaging in housing investment, with higher levels of housing investment, or indeed with measures of housing quality.

However, in the absence of random assignment of property titles, identifying their causal effect on relevant household outcomes can be challenging. I address these challenges in several ways. First, I assume selection on observed variables and include different types of fixed effects, and control for a wide variety of exogenous household characteristics. Second, I show that the results are robust to using propensity

score matching. Third, I consider the possibility that there is selection on unobserved variables. Following Oster (2017), I estimate the effect of property title status on investment decisions under different assumptions on the relative degree of selection on unobserved variables, relative to the degree of selection on observed variables. Fourth, I consider the possibility that my study is under-powered to detect meaningful results, and calculate the minimum effect size that can be detected with the given sample.

Finally, to understand why property titles do not affect investment behaviour in this context, I turn to the theoretical arguments in Besley (1995) and Besley and Ghatak (2010). Property titles can affect investment through three channels: by expanding the opportunities for trade in property markets, by reducing the risk that the property would be expropriated, and by providing collateral that can be used to access credit from the formal financial sector. I consider each of these channels separately, and analyse why they do not apply in the context of my study.

Organisation of the thesis The next three chapters contain the essays described above. These chapters are presented in the form of self-contained journal-style articles. They all include their own sections on literature review, institutional context, empirical methodology, data description, results, discussion of mechanisms, and conclusion.

The last chapter will conclude the thesis, discussing the lessons and implications for future research that result from my work.

CHAPTER 1

THE COST OF DISTANCE: GEOGRAPHY AND GOVERNANCE IN RURAL INDIA

with Sam Asher (The World Bank) and Paul Novosad (Dartmouth College)

1.1 Introduction

In many developing countries, the state has committed to provide public goods such as roads, schools, and health facilities to all citizens. Yet many of them still lack access to these amenities (The World Bank, 2015), even when sufficient resources are available to provide them (Mookherjee, 2015). What constrains the effectiveness of the state in providing universal access to public goods? The economics literature has examined the constraints on both the state's ability to supply public goods (Banerjee et al., 2013; Finan et al., 2017),¹ as well as citizens' ability to demand them (Bardhan, 1984; Alesina et al., 2003; Banerjee et al., 2007).²

We thank seminar participants at the Center for Global Development (Washington DC), the DFID-IZA Workshop (Oxford), the Indian Statistical Institute Annual Conference (Delhi), the Pacific Development Economics Conference (UC Riverside), the CSAE Conference (Oxford), the Nottingham PhD Students Workshop, the IEA Congress (Mexico City), the Indian Ministry of Finance (Delhi), the National Council for Applied Economic Research (Delhi), the National Institute for Public Finance and Policy (Delhi), Ashoka University, the University of Exeter, the Reserve Bank of India (Mumbai), the Indian Institute of Management Bangalore, the Brookings Institution Delhi, the Midwest International Development Economics Conference (Minneapolis), the Annual Bank Conference on Africa (Stanford), and the ADB Conference on Urban Development Economics (Shanghai). We are grateful for discussions with Pranab Bardhan, Julia Bird, Stefano Caria, James Fenske, Doug Gollin, Clement Imbert, Beata Javorcik, Michael Koelle, Julianne Labonne, Simon Quinn, and Guo Xu. All remaining errors are ours.

¹See, for example, Hanna and Wang (2017) on the type of individuals who enter the civil service; Muralidharan and Sundararaman (2011) and Ashraf et al. (2014) on the incentives they face; and Rasul and Rogger (forthcoming) on how the state manages its civil servants.

²On the pernicious effects of elite capture, patronage and clientelism, see, for example, Callen et al. (2013), Anderson et al. (2015), or Burgess et al. (2015).

The state's capacity may be further restricted by the geography of its administration, in particular, by the physical distance between administrators and citizens. The trade literature has documented the importance of transport costs and information frictions in restricting the flow of private economic activities over space (Adamopoulos, 2011; Gollin and Rogerson, 2014; Allen, 2014; Atkin and Donaldson, 2015). Do these frictions also restrict the flow of public services over space? This has been an important motivation for recommending decentralisation as a way of improving public service delivery in developing countries (Bardhan, 2002; Mansuri and Rao, 2013). However, testing this hypothesis is empirically challenging: administrative headquarters, where public officials are based, are often located in towns that also have large goods and labour markets. Distance to administration is therefore correlated with distance to important markets, making it difficult to isolate its effect on rural outcomes.

In this chapter, we test whether distance to administration, what we refer to as "administrative remoteness," matters for the provision of public goods and services, and consequently, whether it affects economic development. We do this by focussing on village distance to district headquarters in India. To recover causal estimates, we employ a spatial regression discontinuity design that compares outcomes for villages located on either side of district borders. These villages are similar in all key respects, but they have substantially different distances to their administrative headquarters. We exploit this variation to examine how – and why – administrative remoteness affects access to public goods, and how these differences, in turn, shape the spatial distribution of poverty and prosperity in rural India.

To do this analysis, we construct a panel dataset covering 600,000 Indian villages, with detailed information on the availability of public goods in each village. We match this to data from a national census that provides household-level information on incomes, assets, and employment structure for 800 million rural individuals. Further, we obtain geo-coordinates for all villages, towns and cities in India. We use them to construct measures of administrative remoteness for all Indian villages, as well as

to calculate their distance to towns and cities. The high spatial resolution and comprehensive coverage enable us to investigate the generalised effects of administrative remoteness in a way that has not been possible before through more aggregate surveys.

We find that administrative remoteness worsens rural access to several public goods that are provided by district administrations in India. Villages that are more distant from their administrative headquarters are less likely to have paved roads, primary schools, health centres, and major irrigation facilities. In turn, a smaller share of their workforce is engaged in the non-farm sector.³ These differences in the employment structure have a bearing on rural living standards. For example, administratively remote villages have lower average monthly incomes.⁴ Further, their residents have worse quality housing, fewer household assets, and lower rates of literacy.

One of the channels driving these results is lower willingness, or ability, of public officials to visit more distant villages. While there are several ways in which public officials can obtain information about the village without travelling there, for some tasks – such as inspecting the quality of public infrastructure – there are few alternatives to physical visits. We confirm this hypothesis using data from India's national rural roads program. We first show that administratively remote villages are no less likely to receive roads under the program, and these roads do not cost more, or take longer to construct, than similar roads in neighbouring villages that are closer to their headquarters. We then investigate whether the quality of the road was audited by national quality monitors employed under the program. This monitoring was intended to be randomly distributed across beneficiary villages. If distance to administration does not matter, we should find no systematic differences in monitoring between administratively remote and proximate villages. However, roads constructed in remote

³Rural employment structure affects rural productivity, because labour productivity in agriculture is lower than in non-farm sectors of the rural economy (Gollin et al., 2014; Restuccia et al., 2008; Caselli, 2005).

⁴We construct a measure of average village monthly income using data from the Socioeconomic and Caste Census 2012 and the National Sample Survey 2011-12, as explained in Section 1.3.2.

villages are 9.4 percent (2.2 percentage points) less likely to be audited under the program, confirming our hypothesis that high transport costs inhibit the state's ability to monitor infrastructure quality in more distant villages.

We consider and rule out four alternative hypotheses. First, governance quality may be correlated with district features such as geographical area, nature of colonial institutions,⁵ or the level of federal budget support.⁶ To account for any factor that changes at the level of the district, we employ district fixed effects throughout our analysis. Second, differential outcomes in remote villages may be driven by other correlates of remoteness that reduce overall access to public goods. We show that administratively remote villages are no less likely to receive public goods provided by other tiers of administration, such as the federal government or village councils. Third, the economic effects of administrative remoteness could be explained by sorting across the district borders, as residents of administratively remote villages relocate to the "better" side of the border. We do not find any evidence of such relocation, consistent with the literature on rural migration in India.⁷ Fourth, provincial borders may restrict rural citizens' access to economic opportunities.⁸ We show that our results are robust to dropping district borders that are also state borders from the sample.

Finally, a policy experiment in which the allocation decision for different public goods is devolved to the sub-district level suggests that such a devolution would considerably lower the costs of administrative remoteness. For example, while an additional kilometre in distance to the district headquarters reduces the probability of access to paved roads by 0.06 percentage points, an additional kilometre in distance to the sub-district centroid reduces the probability of access to paved roads by 0.01 percentage points. In fact, we cannot reject the null that there is no effect of greater dis-

⁵Colonial institutions such as land revenue systems have been shown to affect contemporary public goods provision (Banerjee and Iyer, 2005; Banerjee et al., 2005; Iyer, 2010)

⁶The erstwhile Planning Commission had classified about 250 districts as "backward", which made them eligible for additional support in federal programs.

⁷See, for example, Munshi and Rosenzweig (2016).

⁸For example, Kone et al. (2017) show that the lack of portability of welfare benefits, as well as provincial-level biases in admission to tertiary educational institutions and public sector recruitment, restrict migration across provincial borders in India.

tance to sub-district centroid on access to a number of district-provided public goods.

These results have implications for several strands of the literature. First, these results contribute a new dimension to our understanding of spatial inequality in developing countries. One of the questions investigated by this literature is why remoteness from urban centres is associated with worse economic outcomes. Several papers have documented the negative effects of greater distance to towns, which host large goods and labour markets, on poverty, gains from trade, and structural transformation (Fafchamps and Wahba, 2006; Feyrer, 2009; Michaels et al., 2012; Atkin and Donaldson, 2015; Storeygard, 2016). In this chapter, we argue that towns do not just provide large markets for goods and labour, but also serve as administrative centres. We show that greater distance to such towns reduces rural access to public services, providing another channel through which remoteness from towns worsens rural outcomes.

Further, one of the central premises for decentralisation is that reducing the distance between the state and its citizens can help to improve the efficiency of fiscal expenditure (Oates, 1972; Bardhan, 2002). It has been difficult to test this premise empirically. While there is a growing empirical literature on the benefits of decentralisation (Faguet, 2004; Barankay and Lockwood, 2007; Galiani and Schargrodsky, 2008; Kis-Katos and Sjahrir, 2017), these episodes often accompany widespread political change in the country, making it difficult to disentangle the different components of decentralisation. Our empirical framework allows us to isolate the pure effects of lower distance between citizens and the state, contributing causal evidence to support one of the central premises of decentralisation.

Relatedly, there is a growing literature on the politics of administrative unit proliferation (Grossman and Lewis, 2014; Pierskalla, 2016; Grossman et al., 2017). This literature attributes the surge in the number of sub-national administrative units in developing countries to political economy factors such as creating administrative jobs for political supporters (Green, 2010; Hassan, 2016), ethnic patronage, and clientelism. Our results show that apart from their political consequences, smaller administrative

units can also help improve the delivery of public services and therefore enhance state capacity in developing countries.

Finally, we contribute to a long-standing literature on the provision of public goods in developing countries. This literature has adopted two broad approaches to explain how citizens get access to public goods: bottom-up processes, in which communities organise to demand public goods from the state (Banerjee and Iyer, 2005; Dell, 2010; Anderson et al., 2015; Dell et al., 2017), and top-down interventions, in which the state expands access to public goods throughout the country (Duflo, 2001; Banerjee et al., 2007; Iyer, 2010; Burgess et al., 2015). In this chapter, we provide evidence for an important channel which mediates the effectiveness of both top-down and bottom-up interventions: distance to administration. Greater distance increases costs to the state of supplying the public goods and monitoring their quality. It may also increase costs for communities to organise and make claims on the state.

In the next section (Section 1.2), we explain the context of public administration in India. Then in Section 1.3, we describe the multiple data sources and explain the construction of our main variables. Section 1.4 outlines our empirical strategy and Section 1.5 presents the main results. Section 1.6 outlines the results from a simulated policy experiment. Section 1.7 discusses the mechanisms, and Section 1.8 concludes.

1.2 District administration in India

Public administration in India is divided into several tiers. At the top is the federal government, which is responsible for national public goods such as national security and diplomacy. The federal government also manages organisations that provide public goods through national networks, such as railways, electricity, postal services, and highways.⁹

⁹The railways are managed by the Ministry of Railways, electricity through the Rural Electrification Corporation, postal services through India Post, and highways through the National Highway Authority of India.

Then there are state governments in India's twenty-nine states. State governments are responsible for managing public health and education systems, along with a wide range of social and anti-poverty programs (Second Administrative Reforms Commission, 2009). For many social sectors, the federal and state governments work together: the federal government designs the schemes and provides a significant proportion of the funding, while the state governments take charge of implementing the schemes within their territories. For example, the national rural roads program, the Pradhan Mantri Gram Sadak Yojna (PMGSY), is a federal government program that was launched in 2000. Through this program, the federal government has provided state governments more than US\$ 1.5 billion to construct approximately 400,000 kilometres of all-weather (or paved) access roads to connect more than 185,000 rural habitations to the state and national road network (Asher and Novosad, 2018).

States are further organised into districts. In 2012, India had 640 districts. The district administration is responsible for implementing all federal and state government projects in the district. Unlike the federal and state governments, where an elected executive commands authority over the bureaucracy, district administrations are almost entirely bureaucratic in character (Second Administrative Reforms Commission, 2009).¹⁰ The district administration is headed by the District Collector,¹¹ who often belongs to the elite national cadre of civil servants, the Indian Administrative Service (IAS). The District Collector is considered by many as the chief representative of the Indian state within the district (Kothari, 1970) and has considerable discretion in allocation of federal and state funds to different villages within the district. For example, in the PMGSY program, even though the federal government specified fairly objective eligibility criteria for allocating paved roads to villages, it was the district administration that was responsible for designing the district PMGSY network, after consulting with elected representatives and taking into account special needs for

¹⁰The 73rd and 74th Constitution Amendment Act created district councils, or *zila parishads*, but these have limited powers in most states, with Gujarat and Maharashtra being notable exceptions.

¹¹The Collector is also known as the District Magistrate or the Deputy Commissioner.

marginalised communities (Asher and Novosad, 2018). To facilitate administration and monitoring at local levels, districts are further divided into sub-districts¹² and blocks.

In 1993, the Indian Constitution was amended to create village councils, called *Gram Panchayats*, to serve groups of villages all over the country. These are elected councils that receive federal and state government funds to provide local public goods, such as footpaths and gravel (unpaved) roads, as well as water and sanitation facilities. The village councils are also responsible for identifying welfare beneficiaries for government programs, and for implementing the National Rural Employment Guarantee Act (Chattopadhyay and Duflo, 2004; Anderson et al., 2015).

1.3 Data

To examine the effects of administrative remoteness, we bring together several sources of administrative data identified at the village level. We also obtain geo-coordinates for all towns and villages in India and use these to calculate our distance measures. Below, we describe each dataset in greater detail.

1.3.1 Population Census

Since 1871, the Office of the Registrar General of India (ORGI) has conducted a national population census in the first year of every decade. In this chapter, we use village-level data from the Population Census 2011, which reports demographic characteristics such as population size, age and gender composition, literacy rate, and the share of village population that belongs to the Scheduled Castes (SC) and Scheduled Tribes (ST), which are traditionally marginalised communities in rural India. In addition, the Population Census provides rich information on the availability of various amenities such as different types of roads, schools, health centres, water and sanitation

¹²These are called *tehsils*, *talukas* and *mandals* in different states.

facilities, electricity, post offices, polling booths, and irrigation facilities in the village.

1.3.2 Socioeconomic and Caste Census

In 2012, the Government of India conducted the Socioeconomic and Caste Census (SECC), a one-off census of all households in the country, to collect detailed information on assets, incomes, occupation structure and demographic characteristics at the household level. This information is substantially richer than information collected during the Population Census. The Government of India made the SECC publicly available on the Internet in PDF and Excel formats. To construct a useful dataset from these files, we scraped the Excel and PDF files from the Government's SECC website, parsed the embedded text, and translated the data to English from twelve Indian languages.

In this chapter, we use household level data from the SECC. This includes monthly income of the highest earning member of the household (divided into 3 categories: less than INR 5,000 per month, between INR 5,000 and 10,000 per month, and more than INR 10,000 per month), sources of earnings, measures of housing quality, and the household's ownership of different types of assets. We use the household-level data to generate village outcomes, which we match to village data from the Population Census.

To generate a proxy for average village income, we combine the SECC income data with consumption expenditure data from the 68th round (2011-12) of the National Sample Survey (NSS). We use the NSS data to generate conditional averages of household income. For example, conditional on monthly income being less than INR 5000 per month, the average household income is INR 3,076. Conditional on monthly income being between INR 5,000 and 10,000 per month, the average household income is INR 6,373, and conditional on monthly income being greater than INR 10,000, the average household income is INR 22,357. We use these conditional averages from the NSS to calculate a proxy for average village income. This technique provides us

a measure of average monthly income for every village in India, which is a unique contribution from this dataset.

1.3.3 PMGSY data

We obtain data on the PMGSY program from the administrative software that has been designed for monitoring the program (<http://omms.nic.in>). For each PMGSY road, we observe the estimated and final cost of construction, the date on which construction started, the estimated and final date on which construction was concluded, the length of each road, whether or not the road was subject to quality audits, and quality monitoring reports.

1.3.4 GIS data

We obtained village and town geo-coordinates, GIS data on current and historic district borders, and GIS data on rivers from ML Infomap. We obtained highway GIS data from both OpenStreetMap and the National Highways Authority of India. We use the GIS data to construct straight-line distances between all villages and their neighbouring towns, cities, and administrative headquarters.

1.4 Empirical Strategy

Isolating the effects of administrative remoteness is challenging because administrative headquarters are often located in the largest towns in the district. Therefore, conventional measures of market access – such as distance to towns, cities, and highways – are correlated with distance to administrative headquarters. To address this challenge, our empirical strategy focusses on villages located close to district borders. We show that while distance to towns and cities, and demographic and geographic variables, change smoothly across the border, distance to administrative headquarters changes

sharply at the border.

We illustrate the empirical strategy in Figure 1.1. The polygons in the figure are two districts in the western state of Maharashtra: Ahmadnagar and Beed. Their district headquarters are represented by black diamonds. We focus on the villages close to the Ahmadnagar-Beed border. This border is quite long, approximately 180 kilometres. To ensure that our comparisons are within neighbouring villages, we create equally-spaced segments along the border. In this case, each segment is 20 kilometres long. Each colour represents a different border segment. Within each segment, a filled circle represents a village located on the side of the border that is closer to its administrative headquarters, while a hollow circle represents a village located on the side of the border that is more distant to its headquarters.

As we can observe, the northern and southern ends of the border are closer to the district headquarters of Beed. Therefore, the villages on the Beed side of the border are represented by filled circles, and villages on the Ahmadnagar side are represented by hollow circles. However, the middle section of the border is closer to the district headquarters of Ahmadnagar. Therefore, the villages on the Ahmadnagar side are now represented by filled circles, and villages on the Beed side are represented by hollow circles. Our identifying variation comes from a change in administrative remoteness along the same district border, i.e. within the same district-district dyad.

Following Dell (2010), we estimate the following spatial regression discontinuity equation:

$$y_{v,d,s} = \beta_1 RemoteSide_{d,s} + \beta_2 DistTown_v + f(Location_v) + \beta_3 Geographic_v + \delta_d + \eta_s + \epsilon_{v,d,s} \quad (1.1)$$

where $y_{v,d,s}$ is the outcome of interest for village v located in district d , along segment s of the border between district d and the adjoining district. These outcomes include access to different types of public goods and rural economic indicators, as described

in Section 1.3.

RemoteSide is our binary treatment variable. For each segment s , we have villages on both sides of the district border. For each segment-district group, we find the average distance to the corresponding district headquarters. The side that has the higher average distance to headquarters is “administratively remote”, and hence $RemoteSide_{d,s}$ takes the value 1 for all villages in this segment-district group. Villages on the other side are closer to their administrative headquarters, and hence for them, $RemoteSide_{d,s}$ is 0.

DistTown represents a set of controls for village’s (geodesic) distance in kilometres to the nearest town above different population thresholds. We control for distance to towns whose population in 2011 exceeded 10,000, 50,000, 100,000, and 500,000.

$f(\text{Location}_v)$ is the RD polynomial, which controls for smooth functions of geographic location. Following Gelman and Imbens (2017) and Dell and Olken (2017), $f(\text{Location}_v)$ is a local linear polynomial in latitude and longitude, estimated separately for each administrative border. These linear controls account for any spurious linear trends in outcomes at the border. The results are robust to using other types of RD polynomials, such as a polynomial that includes an interaction term between latitude and longitude, a quadratic polynomial in latitude and longitude, as well as polynomials that include village’s distance to district border.

Geographic refers to geographic controls, such as mean elevation in the village and the village’s distance to a major river. δ_d is the district fixed effect. This controls for any district-specific factors that may affect rural economic outcomes or the provision of public goods. This could include precolonial and colonial institutions such as land revenue systems, or district-specific factors such as average governance quality and district area. The inclusion of the district fixed effect ensures that the coefficient of interest, β_1 , captures the effect of distance from district headquarters that is over and above any district-wide average governance measure. This reassures us that our results are not driven by more “administratively remote” villages being located in

worse-administered districts.

η_s is the border segment fixed effect. District borders can be quite long, ranging from a few kilometres to more than 100 kilometres. We want to ensure that we are comparing villages located in close geographic proximity. Further, as in Figure 1.1, some portions of the border may be close to one headquarters, while other portions are closer to the other headquarters. Therefore, we create equally-spaced segments along the district border. Segment s is the border segment closest to village v .

$\epsilon_{v,d,s}$ is the error term. We cluster the standard errors at the district-segment level, since this is the level of treatment in our analysis (Abadie et al., 2017). The results are robust to allowing spatial correlation in the standard errors, following the method used in Burgess et al. (2016).

The validity of our regression discontinuity design requires a number of key assumptions. These assumptions in turn impose certain restrictions on our sample of villages. First, we compare villages located close to the district border. In our baseline specification, we use a bandwidth of 6 kilometres around the district border. Since there is no widely accepted optimal bandwidth for regression discontinuities in multi-dimensional space, we follow Dell and Olken (2017) and show robustness to several different bandwidths.

Second, we want to ensure that villages on either side of the district border have similar access to the closest town or highway. If this is not the case, the outcomes will be driven by a combination of distance to administration and distance to markets, and it would not be accurate to attribute them to administrative remoteness alone. Geographical features such as rivers, mountains and ridges create barriers (Nunn and Puga, 2012). Hence, we drop borders across which the change in mean elevation is greater than the 90th percentile of all borders in our sample. We further show robustness to dropping administrative borders that coincide with rivers.

Third, we exclude states that are classified as “special category states” by the federal government. These are Jammu and Kashmir, Uttarakhand, Himachal Pradesh,

Assam, Arunachal Pradesh, Manipur, Nagaland, Tripura, Meghalaya, Mizoram and Sikkim. Administration of these states faces special challenges such as “hilly and difficult terrain, low population density or the presence of sizeable tribal population, strategic location along international borders, economic and infrastructural backwardness and non-viable nature of State finances” (The Hindu, 2016). District borders in these states often coincide with geographical barriers such as mountains and ridges, limiting the validity of our identification assumptions. Therefore, while our results are robust to including these states in the sample, we want to be conservative and exclude them from our main sample.

Finally, we show that our results are robust to using an absolute measure of administrative remoteness (distance to district headquarters) rather than the binary treatment variable. To do this, we estimate the following regression equation:

$$y_{v,d,s} = \beta_1 DistanceDistrictHQ_{v,d,s} + \beta_2 DistTown_v + f(Location_v) + \beta_3 Geographic_v + \delta_d + \eta_s + \epsilon_{v,d,s} \quad (1.2)$$

where $DistanceDistrictHQ_{v,d,s}$ is the geodesic distance in kilometres between village v and its district headquarters. All other variable definitions remain the same as in Equation 1.1. This regression specification also allows us to simulate the effects of devolving the public goods allocation decision to the sub-district level in India.

1.4.1 Balance checks

For the control villages to be appropriate counterfactuals for the treatment villages, we require that all controls other than administrative remoteness should change smoothly at the border. If c_1 and c_2 are potential balance outcomes for remote and proximate villages in our sample respectively, x is longitude, and y is latitude, then our empirical design requires that $E[c_1|x, y]$ and $E[c_2|x, y]$ vary continuously at the district border

(Dell et al., 2017).

To test this, we regress each control variable on the treatment dummy, $RemoteSide_{d,s}$, and all the other controls in Equation 1.1. Estimates from these regressions are presented in Table 1.1. The first row shows the first stage estimate, from regressing distance to district headquarters on all the right hand side variables in Equation 1.1. The average treatment size, i.e. the average increase in distance to district headquarters as we move from the administratively proximate side of the border to the administratively remote side, is approximately 15.6 kilometres. This is equal to 37% of the mean distance to district headquarters in our sample. Figure 1.2 shows the discontinuous change in distance to district headquarters at the border (the RD cutoff in our design).

The next four rows present estimates from regressing distance to the nearest town, whose population in 2011 exceeds a certain threshold, on the treatment dummy, $RemoteSide_{d,s}$ and all the other controls in Equation 1.1. If our empirical strategy is valid, we should not be able to reject the null hypothesis that β_1 in these regressions is zero. We find that we cannot reject the null hypothesis that β_1 is equal to zero for village distance to towns with population exceeding 10,000 and 500,000 in 2011. Further, the magnitude of the regression coefficients is very small as compared to the sample mean for these variables: 0.2% for distance to towns with population exceeding 10,000, and 0.03% for distance to towns with population exceeding 500,000.

However, treatment villages are marginally more distant to towns with population in 2011 exceeding 50,000 and 100,000. This is not surprising, since for most villages in our sample, the nearest town in this population range serves as a district headquarter, even though it may not be that village's district headquarter.

For example, consider the villages in yellow in Figure 1.1, located approximately in the middle section of the border. While treatment (administratively remote) and control (administratively proximate) villages are located very close to each other, given a bandwidth of 6 kilometres, they are still, on average, about 6 kilometres apart. For all these villages, the nearest large town is Ahmadnagar, with a population of 350,905 in

2011. For the control yellow villages, located on the Ahmadnagar district side of the border, the average distance to Ahmadnagar town is 18.5 kilometres. This is also the villages' average distance to their district headquarters. For the treatment yellow villages, located on the Beed district side of the border, the average distance to Ahmadnagar town is 22.9 kilometres, about 4.5 kilometres more than the average distance for control villages. However, the difference in treatment (average distance to district headquarters) is much larger, 89.6 kilometres for treatment villages in Beed versus 18.5 kilometres for control villages in Ahmadnagar. Hence, on average, treatment villages are marginally more distant to towns with population greater than 50,000 or 100,000, given our empirical design. What further reassures us is that the magnitudes are a few hundred meters. For towns with population exceeding 50,000, the magnitude is 1.5% of sample mean, and for towns with population exceeding 100,000, the magnitude is 1.9% of sample mean.

To confirm that these statistically significant differences are caused by the average distance between treatment and control villages, we plot these coefficients for several different bandwidths in Figure 1.3 and Figure 1.4 (Figure 1.5 and Figure 1.6 do the same for the coefficient on distance to towns with population exceeding 10,000 and 500,000 respectively). These coefficient plots confirm that the average difference between treatment and control villages in distance to towns with population greater than 50,000 and 100,000 increases with the selected bandwidth around the border.

Figure 1.7 shows how distances to towns with different population sizes change across the district border. The graphs show that distances to towns are continuous at the border, except for the small discontinuous change in distance to towns with population greater than 50,000 and 100,000.

The next five rows of Table 1.1 show that demographic and geographic characteristics are also balanced for treatment and control villages, except for distance to nearest river. Treatment villages are 155 meters more distant to the nearest river, which is 0.3% of the mean distance to nearest river in our sample. We control for distance to nearest

river in all our regressions, and also show robustness of the main results to dropping district borders that coincide with rivers. Figure 1.8 shows that geographic and demographic variables are also continuous at the border, except for the small discontinuous change in distance to the nearest river.

1.5 Results

Administrative remoteness reduces the provision of public goods, decreases the share of non-farm employment, and worsens poverty in rural India. In this section, we first present the main effects on access to paved roads (Section 1.5.1), discussing robustness of the result to different specifications and sample restrictions. We then present results for access to different types of public goods, classified by the tier of administration that has the responsibility to provide them (Section 1.5.2), and discuss the economic effects of administrative remoteness (Section 1.5.3). Finally, in Section 1.5.4, we show that these results are robust to changing the sample restrictions and regression specification choices.

1.5.1 Roads

We first test whether administrative remoteness reduces the probability that a village has a paved road. The first row of Table 1.2 presents regression coefficients from estimating Equation 1.1 for rural access to paved roads. We consider that a village has a paved road if the Village Directory of Population Census 2011 records that the village has an “all-weather road.” Using our preferred specification, we find that villages on the administratively remote side of the border are 1.2 percentage points less likely to have a paved road than control villages. Given that 69% villages in our sample have a paved road, this implies a 1.7 percent reduction in paved road access relative to the sample mean.

Subsequent rows of Table 1.2 show robustness of the main result to different regres-

sion choices. First, Row 2 shows that the statistical significance changes marginally when we allow for spatial correlation in the standard errors. Then, in rows 3-5, we present estimates using different RD polynomials: a linear polynomial in distance to border, a linear polynomial in distance to border and latitude and longitude, and a quadratic polynomial in latitude and longitude. The effect of administrative remoteness on paved road access is similar across all these regressions. Row 6 shows that the administrative remoteness effect is larger when we drop district borders that coincide with rivers, and when we drop state borders from our sample (Row 7). Finally, in Row 8, we observe that this result is robust to restricting the bandwidth to 3 kilometres around the district border.

Since the binary treatment variable is an average of different treatment sizes, we also estimate Equation 1.2 in which we use the absolute measure of administrative remoteness, *DistanceDistrictHQ*, rather than the binary treatment variable, to estimate the impact of each additional kilometre in distance to the district headquarters. This result is presented in the last row of Table 1.2. Moving from the 25th to the 75th percentile of the distance to district headquarters distribution in our sample (from a distance of 25 kilometres to 53 kilometres) reduces the probability that a village has a paved road by 1.84 percentage points, or 2.6 percent of the sample mean.

Figure 1.9 presents the roads result graphically, showing the drop in probability of paved road access for administratively remote villages, to the right of the border discontinuity.

1.5.2 Other public amenities

Having established that administrative remoteness reduces rural access to paved roads, we now investigate whether it affects the provision of other public goods as well. Table 1.3 presents results for a wider set of public amenities, which we classify into 3 categories: those provided by the district administration, those provided by higher tiers of administration such as the federal government, and those provided by village

councils.

First, the administrative remoteness effect is not limited to paved roads alone, but extends to other public goods for which the district administration is responsible. These include government primary schools, mobile health clinics, irrigation facilities, and treated tap water. Remote villages are about 1 percentage point less likely to have a government primary school, 0.3 percentage points less likely to have a mobile health clinic, and 1.1 percentage points less likely to have treated tap water. Further, the share of land irrigated in these villages is 0.7 percentage points lower. These results are robust to dropping state borders from our sample.

However, we cannot reject the null that distance to district headquarters does not matter for access to public goods that are provided by the federal government, such as post offices, polling stations, national highways or electricity. Neither can we reject the null that distance to district headquarters does not matter for access to local amenities for which the village councils are responsible, such as unpaved (gravel) roads, and water and sanitation facilities. The fact that treatment and control villages are not different in their access to public goods provided by other tiers of administration suggests that these effects are not driven by the treatment villages' general remoteness. Rather, the treatment villages lack paved roads, primary schools, and other public goods primarily on account of their greater distance to their district administration.

1.5.3 Economic impacts

We now test whether administrative remoteness, which reduces rural access to public goods, also affects economic outcomes. Table 1.4 presents results for four types of economic outcomes. First, we note that administrative remoteness affects the rural employment structure. We find that administrative remoteness reduces the share of village workforce engaged in non-farm activities, potentially through its effect on access to paved roads and other public goods. The non-farm employment share is 0.7 percentage points lower in treatment villages as compared to control villages. As a

percentage of the sample mean, this is equivalent to a 2.8 percent decrease in non-farm employment share in treatment villages. This result is robust, though smaller in magnitude, when we drop state borders. Figure 1.10 presents this result graphically: there is a drop in non-farm employment share at the district border. A large literature in economics has documented differences in productivity between agriculture and other sectors of the economy (Gollin and Rogerson, 2014). Structural transformation of the economy from agriculture to other economic sectors is key to income growth in developing countries. Our results suggest that administrative remoteness inhibits the structural transformation of the economy away from agriculture.

One explanation for this result could be differences in public sector employment. Greater distance to administrative headquarters could reduce rural residents' access to government jobs and hence drive differences in the employment structure. We find that the share of households that report a government job as the main source of earnings is 0.07 percentage points lower in treatment villages as compared to control villages. This magnitude is a small proportion of the remoteness effect on the non-farm employment share. Further, the difference in the government employment share is smaller (and statistically indistinguishable from zero) when we drop state borders. This is consistent with recent evidence that documents state-level biases in public sector recruitment in India (Kone et al., 2017). Residents of administratively remote villages can find public sector employment in the neighbouring district, but only if the neighbouring district belongs to their state.

Second, administrative remoteness has small effects on household income. Average monthly income in treatment villages is approximately 50 rupees (1% of sample mean) lower than in control villages. However, this result is not robust to dropping state borders. Residents of treatment villages may be commuting to the other side of the border in search of employment opportunities. This ensures equilibrium in observed nominal incomes (except where mobility is restricted, such as state borders), even though commuting costs may be higher for administratively remote rural indi-

viduals.

Third, along with the effect on income, we observe small effects on household assets. For example, the share of households that have a solid wall is 0.9 percentage points lower, the share of households that have a solid roof is 0.4 percentage points lower, and the share of households that own any vehicle is 0.4 percentage points lower in treatment villages. Once again, these magnitudes are fairly small as compared to the outcome mean, and smaller yet when we exclude state borders. The income and asset effects point to the importance of state borders in restricting commuting and labour mobility in rural India.

Finally, administrative remoteness reduces the literacy rate. In treatment villages, the share of literates in the village population is 0.3 percentage points lower than in control villages. This could be driven by the reduced access to government primary schools in treatment villages. However, similar to incomes and assets, when we exclude state borders, the magnitude of the remoteness effect on literacy is smaller. This suggests that even when villages do not have primary schools, their young residents can cross the border to enrol at the nearest village school in the neighbouring district. But this is harder to do when the neighbouring district is in another state.

1.5.4 Robustness

In this section we investigate whether our results are driven by the set of choices we have made while selecting our sample or while specifying the regression discontinuity equation. We find no evidence that that is the case.

One threat to our identification comes from district borders that coincide with rivers. We have assumed that villages on both sides of the border have similar access to market opportunities, but this is not the case when villagers have to cross a river to reach the nearest town. Therefore, we re-estimate our results after dropping all borders that coincide with rivers. In Table [1.A1](#), we show that administrative remoteness reduces access to public goods provided by the district administration even

after restricting our sample to non-river borders. Additionally, now we cannot reject the null that remoteness does not reduce access to a few federal public goods, such as electricity and national highways. Further, Table 1.A2 shows that remoteness reduces non-farm employment share and worsens incomes, assets and literacy across the set of district borders that do not coincide with rivers.

Another concern is that our results are driven by our regression specification choices, such as the type of treatment variable used, the size of the bandwidth around the district border, the length of the border segments, the specification of the RD polynomial of latitude and longitude, and how we cluster the standard errors. We show here that our results are robust to changing each of those decisions.

First, in our preferred specification, the treatment variable, $RemoteSide_{d,s}$, is binary, indicating the villages' *relative* degree of administrative remoteness, relative to neighbouring villages located just across the district border. The results are robust to using an absolute measure of administrative remoteness: the villages' distance to their district headquarters. We present the results from estimating Equation 1.2 in Tables 1.A3 and Table 1.A4. The coefficients in Table 1.A3 represent the change in probability of access to public goods for every 1 kilometre change in distance to district headquarters. In Table 1.A4, the coefficients represent the change in probability of access to public goods for 1 standard deviation change in distance to district headquarters. In our sample, the standard deviation of distance to district headquarters is 23.61 kilometres.

Second, we use a bandwidth of 6 kilometres around the district border in our preferred specification, which allows us to compare villages located in close geographic proximity while also giving us sufficient power to test our hypotheses. We show here that our results are robust to changing the size of the bandwidth. In Section 1.4.1, we have already discussed how the coefficients on distance to towns with population exceeding 10,000, 50,000, 100,000, and 500,000 change when we change the bandwidth from 1 to 9 kilometres around the border. Now, Figure A1 shows how the coefficient on paved roads changes as we change the bandwidth similarly from 1 to 9 kilometres.

Except for small bandwidths of 1 and 2 kilometres, where we do not have sufficient power to reject the null, we find that we can always reject the null that administrative remoteness does not reduce access to paved roads. In a similar way, Figure [A2](#) shows how the coefficient on non-farm employment share change as we change the bandwidth. Though estimates using small bandwidths have large standard errors, we find that we can still always reject the null that administrative remoteness reduces the village's share of non-farm employment.

In the interest of brevity, we do not show similar coefficient plots for all other outcomes variables. However, for an intermediate bandwidth of 3 kilometres, we show that administrative remoteness reduces access to public goods provided by the district administration (except for mobile health clinics, for which we cannot reject the null of no effect), but not to public goods provided by other tiers of administration (Table [1.A5](#)). Further, even with a 3 kilometre bandwidth, administrative remoteness reduces the village's non-farm employment share, and reduces household assets and the literacy rate (Table [1.A6](#)).

Second, we divide each border into segments that are 20 kilometres long to ensure comparison between geographically proximate villages. Tables [1.A7](#) and [1.A8](#) show that our results are robust to changing the length of the border segments to 40 kilometres.

Third, our preferred RD polynomial is linear in latitude and longitude and estimated separately for each district border, following Gelman and Imbens (2017) and Dell and Olken (2017). We test whether the results are robust to using a quadratic polynomial in latitude and longitude. Tables [1.A9](#) and [1.A10](#) show that they are.

Finally, we have clustered our standard errors at the level of district and segment, since that is the unit of treatment in our analysis (Abadie et al., 2017). We test whether our results are robust to clustering the standard errors in 10 x 10 kilometre grid cells to allow for spatial correlation in the error term, as done by Burgess et al. (2016). In Tables [1.A11](#) and [1.A12](#), we show that standard errors obtained by clustering at the

level of grid cells are often smaller than standard errors obtained by clustering at the district-segment level, and our results continue to hold.

1.6 Policy experiment

We use the absolute administrative remoteness results presented in Tables 1.A3 and 1.A4 to simulate the following policy experiment: what would be the consequences of devolving the allocation decisions for different district-provided public goods to the sub-district level? For this, we obtain the geo-coordinates of the centroid of each sub-district, and calculate the distance between the village and the centroid of its sub-district. We then replace distance to district headquarters with the distance to sub-district centroid in Equation 1.2, and estimate the resulting regression equation:

$$y_{v,d,s} = \beta_1 Distance_Subdistrict_Centroid_{v,d,s} + \beta_2 DistTown_v + f(Location_v) + \beta_3 Geographic_v + \delta_d + \eta_s + \epsilon_{v,d,s} \quad (1.3)$$

where $Distance_Subdistrict_Centroid_v$ is the geodesic distance in kilometres between village v and the centroid of its sub-district. All other variables are defined in the same way as they are in Equation 1.1.

The results are presented in Table 1.5. For all public goods except treated water, we cannot reject the null that greater distance to the sub-district centroid does not affect access to the public good, using a 95% confidence interval. The point estimates are also smaller than the point estimates obtained when using distance to district headquarters. For example, while an additional kilometre in distance to the district headquarters reduces the probability of access to paved roads by 0.06 percentage points, an additional kilometre in distance to the sub-district centroid reduces the probability of access to paved roads by only 0.012 percentage points. This analysis suggests that reducing the distance between the village and its administrative headquarters can im-

prove its access to certain public goods.

1.7 Mechanisms

Why does distance to administration reduce the availability of public goods? There are several possible reasons, from the perspectives of both the state (the supplier of public goods) and the citizens (the demanders for public goods).

From the perspective of the state, constructing infrastructure can be expensive. These costs may depend on how far the village is from the administrative headquarters. For example, if the materials for constructing roads are sourced in the headquarters, shipment costs to get them to the village can increase with distance. A related reason is corruption: it may be more difficult for administrators to observe expenditures and output quality in more distant locations, allowing contractors to inflate costs. Given a fixed budget and higher costs in more distant locations, administrators may decide to prioritise nearby locations over distant ones.

Even after infrastructure is constructed, it requires regular maintenance. Roads get filled with potholes, canals require to be desilted. Outside the normal maintenance cycle, there are primarily two ways through which administrators learn about infrastructure quality: one, ground reports from villages, and two, by visiting the villages themselves. When the village is distant from the headquarters, both ground reports and official visits may be less frequent. Even when administrators learn about degradation in quality, it can be more expensive to repair infrastructure due to the shoe-leather cost and corruption channels explained in the previous paragraph.

From the perspective of the citizens, greater distance to administration can weaken their ability to organise and demand public goods from the state. For example, when a village does not have a paved road, its residents can either wait for the state to build a road in the distant future, or they can organise, petition and lobby with politicians and bureaucrats to persuade them to construct a road. Often, villagers will have to

travel frequently to administrative headquarters to make these demands. This can be harder to do when the administrative headquarters is a full day's journey away. The demand and supply channels can, in turn, reinforce each other. For example, in the absence of public goods, citizens can disengage from the political process and acquire a more negative view of the state and its ability to deliver them public goods and services. Krishna and Schober (2014) provides descriptive evidence for such a channel in southern Indian villages.

To test some of these mechanisms, we exploit detailed project data from the Pradhan Mantri Gram Sadak Yojna (PMGSY), a national program launched in 2000 to build paved roads in villages all over the country. Until 2015, the program had been used to construct paved access roads in more than 185,000 rural habitations. These roads were built to identical national standards. District administrations were allocated key responsibilities for designing the PMGSY network in their districts, deciding the priority order for construction, inviting bids from contractors, and awarding them construction contracts, and supervising the construction process through quality inspectors.

We assemble data on the cost for constructing each road, the time taken to build the road, and road characteristics such as length and type of surface. Additionally, we have data on whether the road quality was audited by official quality monitors. We use this data to test how administrative remoteness affects costs and monitoring in national infrastructure projects.

Table 1.6 presents the results. First, administrative remoteness does not affect the cost of building the roads, or the time taken for the construction. The result holds even when we control for the type of surface of the road constructed under the program. This is relevant since different types of surfaces have different average costs. One explanation for the result could be that more expensive roads are located in treatment villages and these are less likely to get constructed. However, we note that remote villages are actually more likely to receive a road under the program. Therefore, administrative remoteness does not seem to affect the cost margin, at least not in the

PMGSY program.

After construction, though, roads in treatment villages are 2.1 percentage points less likely to get audited through quality inspectors. Even though assignment of roads to audit was random, we find that district administrations are significantly less likely to assign auditors to roads constructed in more distant villages. This suggests administrative remoteness affects monitoring costs.

1.8 Conclusion

Many countries have committed to providing universal access to public goods and services that are both intrinsically valuable, as well as important ingredients for private economic activities. Yet, many people continue to live without these public goods, restricting economic growth and exacerbating inequalities in living standards. A long literature in economics has studied the factors that constrain the state's effectiveness in developing countries, studying its administrative machinery and its citizens' ability to make the administration work in their interest.

However, both the state's officials and its citizens are also constrained by the geography of its administration: where the officials are based relative to where the citizens live. In a world of high transport costs and information frictions, this becomes an additional channel through which the state's effectiveness is curtailed.

In this chapter, we estimate the costs of "administrative remoteness," or the citizens' physical distance from their administration. We use a rich dataset on rural public goods and household economic outcomes covering the universe of Indian villages, and exploit spatial discontinuities in distance to administration by comparing villages located across district borders. We show that such villages are similar in all respects except in their distance to administration. We find that administrative remoteness reduces rural access to paved roads, schools, health centres and irrigation facilities, and worsens rural economic outcomes.

While our data provides us a comprehensive picture of the outcomes of the state's activities, we have much more limited information on the state's inputs into development. Using implementation data from India's national rural roads program, we find that administrative remoteness does not affect the cost or duration of road construction in rural habitations. However, roads constructed in more administratively remote villages are less likely to be audited by national quality monitors. Reduced monitoring can lead to the construction of worse roads in rural areas (Olken, 2007), squandering fiscal resources and denying villagers the opportunity to access external markets. However, the relationship between the state and its citizens is correlative. When denied critical public inputs, how do citizens organise to demand public goods from the state? Our data does not allow us to observe and measure citizen actions in response to the under-provision of public goods. Such questions should be the subject of future research.

Few questions have received more attention from social scientists than the constraints on state effectiveness and how to make the states work better in the interest of their citizens. Our results suggest that the spatial organisation of public administration is an important barrier to state effectiveness. Reducing the distance between the citizens and the state can help to improve the state's ability to provide and monitor public goods provision in rural areas and reduce the spatial inequality in living standards observed across many developing countries today. However, policies that reduce distance, such as redrawing administrative borders, or changing the location of administrative headquarters, or creating smaller administrative units, may pose additional fiscal burden. A more cost-effective policy could be to change the incentives facing front-line bureaucrats, and arming them with better data on rural public goods provision. But generating frequent, high-quality data can also be expensive. Developing countries would have to evaluate whether the additional benefits in terms of expanded rural opportunities and reduced spatial disparities are worth these costs.

Table 1.1
Changes at the district border

	β_1	(se)	Sample mean	N
First stage				
Distance to District HQ	15.579	(.223)***	41.87 km	180,331
Distance to towns				
Pop > 10,000	0.047	(0.073)	15.88 km	180,331
Pop > 50,000	0.436	(0.064)***	29.14 km	180,331
Pop > 100,000	0.776	(0.064)***	40.66 km	180,331
Pop > 500,000	-0.031	(0.038)	90.81 km	180,331
Demographic variables				
Population 2011	-3.188	(10.855)	1,397	180,331
Scheduled Caste share	0.027	(0.149)	18.89%	180,331
Scheduled Tribe share	0.195	(0.197)	16.59%	180,331
Geographic variables				
Altitude	0.510	(0.488)	242.6 meters	180,331
Distance to river	0.155	(0.051)***	52.09 kms	180,331
Bandwidth: 6 km				
Geographic polynomial: Linear in latitude and longitude				

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

Notes: This table presents regression coefficients from estimating Equation 1.1. The left hand side variables are listed in Column 1. Column 2 presents the point estimate for β_1 , the coefficient on the binary treatment variable (a village is considered treated if it is located on the more administratively remote side of the district border). Column 3 presents the standard errors, clustered at the level of district-segment. Column 4 presents the sample mean for the left hand side variables, and Column 5 has the number of observations. The regression in the first row includes the full set of controls. Every other regression includes the full set of controls except for the variable on the left hand side. All regressions include a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad, and district and segment fixed effects.

Table 1.2
Effects on paved road access

	β_1	(se)	Sample mean	N	Cluster	Bandwidth	RD Polynomial
Paved road access	-1.232	(.409)*** (.449)***	69.56%	180,264	District-segment Spatial	6 km 6 km	Linear Lat-Long Linear Lat-Long
Different RD polynomials							
Paved road access	-1.137	(.575)**	69.56%	180,264	District-segment	6 km	Distance to border
Paved road access	-1.009	(.600)*	69.56%	180,264	District-segment	6 km	Both
Paved road access	-0.950	(.377)**	69.56%	180,264	District-segment	6 km	Quadratic Lat-Long
Drop riverine borders							
Paved road access	-1.583	(.437)***	69.16%	157,395	District-segment	6 km	Linear Lat-Long
Drop state borders							
Paved road access	-1.654	(.598)***	70.33%	141,621	District-segment	6 km	Linear Lat-Long
Restrict bandwidth							
Paved road access	-1.490	(.743)***	70.67%	76,083	District-segment	3 km	Linear Lat-Long
Probability of new roads							
Paved road access	-1.278	(.692)*	71.45%	82,767	District-segment	6 km	Linear Lat-Long
Intensity of treatment							
Paved road access (25th-75th pct)	-1.843	(.447)***	69.56%	180,264	District-segment	6 km	Linear Lat-Long

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

Notes: This table presents regression coefficients from estimating Equation 1.1 for rural access to paved roads. Paved road access is measured using a binary variable that takes the value 1 if the village is recorded as having an “all-weather road” according to the Village Directory of Population Census 2011. Column 2 presents the point estimate for β_1 , the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. Column 3 presents the standard errors. In all rows except Row 3, these are clustered at the district-segment level. In Row 3, we cluster standard errors allowing for spatial correlation, using 10x10 km grid cells. Column 4 and 5 present the sample mean and sample size respectively. Column 6 states the bandwidth used around the district border. Except for Row 7, we employ a 6 km bandwidth around the district border. In Row 7, we employ a bandwidth of 3 km. Finally, Column 7 states the RD polynomial used. Except for Rows 3-5, this is a linear polynomial in latitude and longitude, estimated separately for each district border. In Row 3, the RD polynomial is a linear polynomial in distance to district border (inversed for the control villages). In Row 4, the RD polynomial is a linear polynomial in distance to district border (inversed for control villages), and latitude and longitude. In Row 5, the polynomial is a quadratic polynomial in latitude and longitude. Row 8, the second last row in the table, restricts the sample to villages that did not have a paved road in 2001. The left hand side variable is the probability that the village did not have a paved road in 2001 but did have one in 2011. The last row of the table, Row 9, presents estimates from regressing paved road access on the distance to district headquarters, rather than the binary treatment variable. The magnitude of β_1 in the last row is the decrease in the probability that a village has a paved road when we go from the 25th percentile (25 kilometres) to the 75th percentile (53 kilometres) of the distance to district HQ distribution in our sample. All regressions include district and border segment fixed effects.

Table 1.3
Rural access to different public goods

	All district borders			Without state borders		
	β_1 (se)	Sample mean	N	β_1 (se)	Sample mean	N
Villages with:						
Amenities provided by districts:						
Paved roads	-1.232 (0.409)***	69.56%	180,264	-1.654 (0.598)***	70.33%	141,621
Govt primary schools	-0.947 (0.257)***	83.10%	179,904	-1.097 (0.283)***	83.05%	141,364
Mobile health clinic	-0.310 (0.092)***	1.84%	180,135	-0.208 (0.094)**	1.57%	141,547
Share of land irrigated	-0.708 (0.231)***	44.76%	180,299	-0.286 (0.262)	46.25%	141,692
Treated water	-1.190 (0.279)***	21.82%	178,606	-1.121 (0.321)***	22.03%	140,252
Amenities provided by federal govt						
Post office	0.021 (0.190)	9.99%	178,608	0.161 (0.199)	10.31%	140,251
Polling station	-0.325 (0.295)	68.21%	178,595	-0.329 (0.309)	70.05%	140,239
National highway	-0.259 (0.172)	5.02%	178,609	-0.211 (0.198)	5.28%	140,252
Electricity	-0.542 (0.368)	55.26%	180,299	0.066 (0.419)	57.28%	141,692
Amenities provided by village councils						
Gravel road	0.273 (0.293)	85.57%	178,569	0.332 (0.339)	85.08%	140,217
Community toilet complex	-0.019 (0.095)	4.95%	178,609	-0.066 (0.109)	2.38%	140,252
Community well	-0.425 (0.370)	56.69%	178,602	-0.458 (0.429)	54.69%	140,252
Tubewell	-0.261 (0.415)	50.49%	178,595	-0.152 (0.471)	52.83%	140,252

Bandwidth: 6 km

Geographic polynomial: Linear in latitude and longitude

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

Notes: This table presents regression coefficients from estimating Equation 1.1 for rural access to different public goods. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. All left hand side variables, except for “share of land irrigated,” are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For “share of land irrigated,” we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 1.4
Effects on rural incomes, assets and employment

	All district borders			Without state borders		
	β_1 (se)	Sample mean	N	β_1 (se)	Sample mean	N
Employment structure						
Non-farm share	-0.733 (0.208)***	26.58%	178,457	-0.577 (0.220)***	26.89%	140,279
Govt salary share	-0.076 (0.043)*	3.75%	171,425	-0.026 (0.049)	3.88%	134,777
Enterprise share	-0.035 (0.055)	1.12%	171,423	-0.078 (0.045)*	1.11%	134,775
Incomes						
Average monthly income	-50.468 (18.844)***	4,893	171,422	-18.566 (20.841)	4,946	134,774
Household assets						
Solid wall share	-0.871 (0.229)***	53.22%	171,422	-0.675 (0.251)***	54.68%	134,774
Solid roof share	-0.384 (0.212)*	44.36%	171,422	-0.100 (0.233)	45.77%	134,774
Vehicle ownership share	-0.365 (0.157)**	19.32%	171,425	-0.273 (0.168)	20.03%	134,777
Literacy						
Literacy rate	-0.261 (0.087)***	55.99%	178,465	-0.156 (0.094)*	57.34%	140,269

Bandwidth: 6 km

Geographic polynomial: Linear in latitude and longitude

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Notes: This table presents regression coefficients from estimating Equation 1.1 for different rural economic outcomes. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. “Nonfarm share” is the share of main workers in the village that do not report their main occupation as cultivators or agricultural labourers, as per the Primary Census Abstract of Population Census 2011. “Govt salary share” is the share of households in the village that report a government salary as their main source of income in the Socioeconomic and Caste Census 2012 (SECC 2012). “Enterprise share” is the share of households in the village that report income from an enterprise as the main source of income in SECC 2012. “Average monthly income” is calculated as explained in Section 1.3.2. “Solid wall share”, “solid roof share”, and “vehicle ownership share” are, respectively, the share of households that have a solid wall, the share of households that have a solid roof, and the share of households that own at least 1 vehicle as per SECC 2012. Finally, “literacy rate” is the share of village population that is literate according to the Primary Census Abstract of the 2011 Population Census. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 1.5
Policy experiment results

Villages with:	β_1	(se)	Sample mean	N
Amenities provided by districts:				
Paved roads	-0.012	(0.024)	69.77%	179,446
Govt primary schools	-0.029	(0.015)*	83.04%	179,150
Mobile health clinic	-0.001	(0.005)	1.72%	179,381
Share of land irrigated	-0.037	(0.014)*	44.78%	179,544
Treated water	-0.032	(0.015)**	22.08%	177,840

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

Notes: This table presents regression coefficients from estimating Equation 1.2 for rural access to different public goods, using village's distance to the centroid of its sub-district. β_1 is the change in access to the public good for 1 kilometre change in the absolute treatment variable, $Distance_Subdistrict_Centroid_v$. All left hand side variables, except for "share of land irrigated," are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For "share of land irrigated," we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

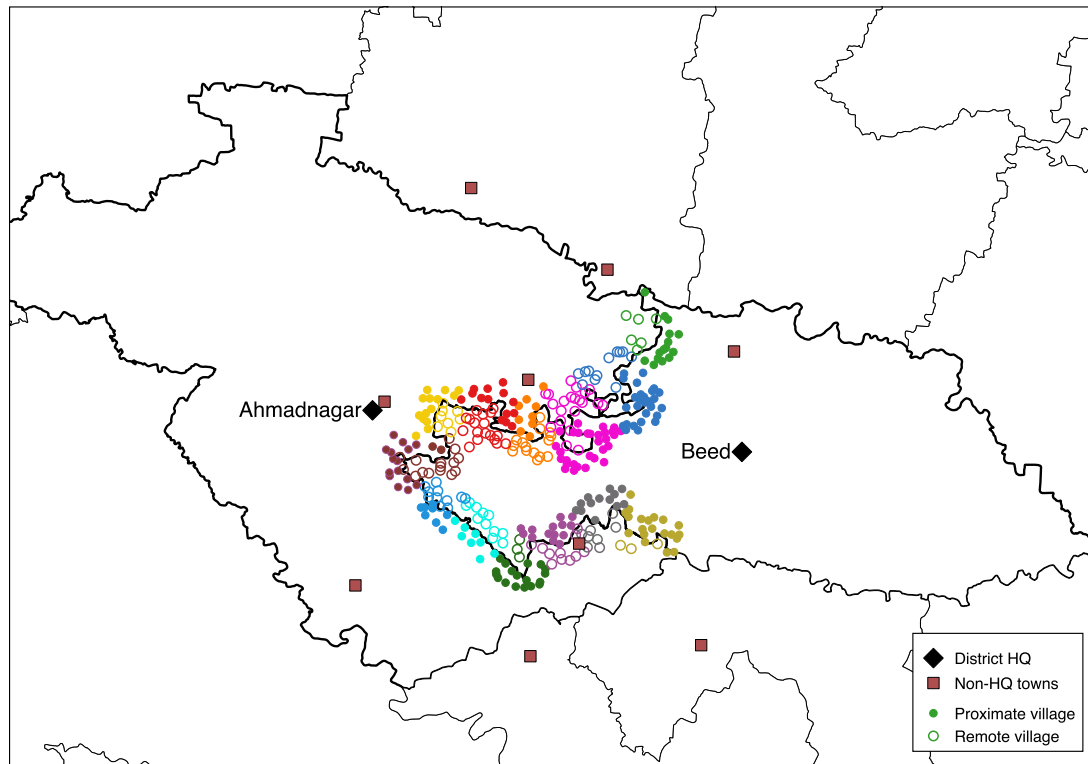
Table 1.6
PMGSY Program Information

	β_1	(se)	Sample mean	N
Placement				
Received PMGSY road	0.655	(.225)**	16.73%	180,299
Project outcomes				
Estimated cost per km	-0.007	(.020)	INR 3.21 million	27,147
Actual cost per km	-0.003	(0.020)	INR 2.49 million	19,858
Estimated duration per km	-1.271	(2.351)	145.8 days	24,052
Actual duration per km	-1.684	(4.951)	270.3 days	22,489
Controlling for surface type				
Estimated cost per km	0.012	(0.022)	INR 3.21 million	22,273
Actual cost per km	-0.012	(0.024)	INR 2.49 million	15,369
Estimated duration per km	-2.381	(2.649)	145.8 days	19,405
Actual duration per km	-6.667	(5.974)	270.3 days	17,925
Monitoring inspection				
Probability of quality audit	-2.192	(0.834)***	23.09%	30,170
Bandwidth: 6 km				
Geographic polynomial: Linear in latitude and longitude				

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

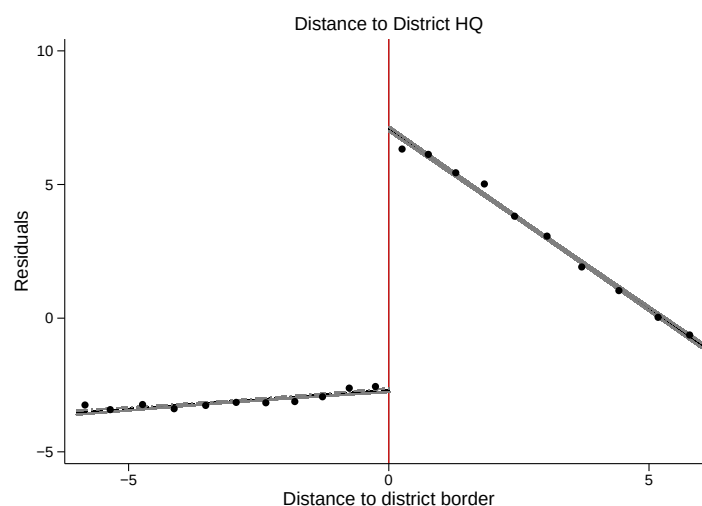
Notes: This table presents regression coefficients from estimating Equation 1.1 for PMGSY program outcomes. β_1 is the coefficient on the binary treatment variable, *RemoteSide_{d,s}*, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. In some specifications, we also control for type of road surface. Cost figures are in millions of Indian rupees and duration figures are in number of days. All regressions include district and border segment fixed effects. We control for geographic location using a linear polynomial in latitude and longitude. Standard errors are clustered at the district-segment level.

Figure 1.1
Illustration of Empirical Strategy



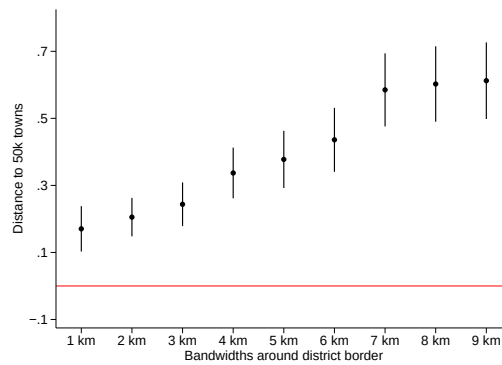
Notes: Example of our empirical strategy, showing Beed and Ahmadnagar districts in Maharashtra, with their district headquarters, border villages, and towns with 2011 population greater than 10,000. Different coloured dots represent different segments of the district border. Filled circles are villages located on the side of the border that is closer to its administrative headquarters, hollow circles are villages located on the side of the border that is more distant to its administrative headquarters.

Figure 1.2
RD plot: first stage



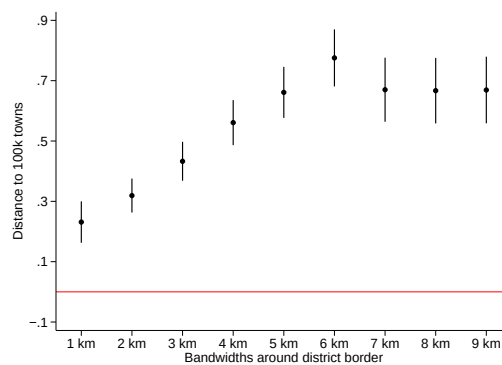
Notes: The figure plots residualised distance to district headquarters, after controlling for all balance variables: distance to towns with 2011 population greater than 10,000, 50,000, 100,000, and 500,000, percent of village population that belongs to Scheduled Castes and Scheduled Tribes, mean elevation, distance to the nearest major river, as well as a linear polynomial in latitude and longitude that is estimated separately for each district border (district-district dyad). Points to the right of zero represent administratively remote villages that are more distant to their district HQ than neighbouring villages on the other side of the border, while points to the left of zero are administratively proximate villages. Each point represents approximately 6,000 observations. A linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages within 6 kilometres of the district border. It excludes villages in special category states, villages along riverine district borders, and villages along district borders with large elevation changes.

Figure 1.3
Distance to towns with population greater than 50,000, using different bandwidths



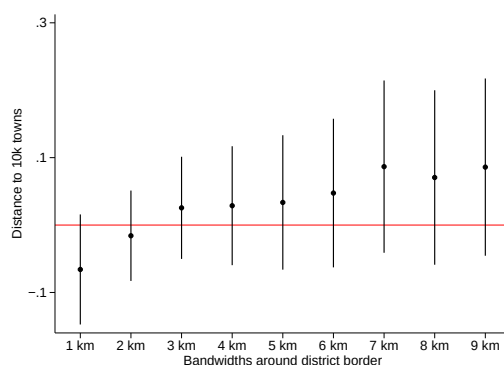
Notes: The figure plots the coefficient β_1 when we regress distance to town with population greater than 50,000 on all right hand side variables in Equation 1.1, except itself, for different bandwidths around the district border.

Figure 1.4
Distance to towns with population greater than 100,000, using different bandwidths



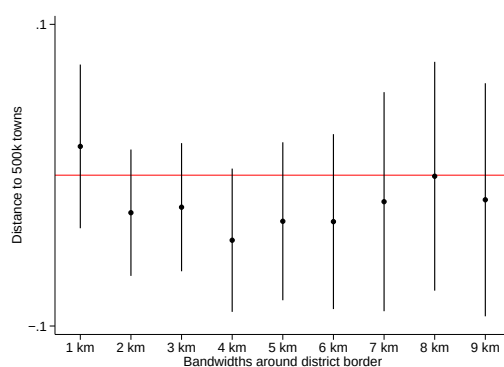
Notes: The figure plots the coefficient β_1 when we regress distance to town with population greater than 100,000 on all right hand side variables in Equation 1.1, except itself, for different bandwidths around the district border.

Figure 1.5
Distance to towns with population greater than 10,000, using different bandwidths



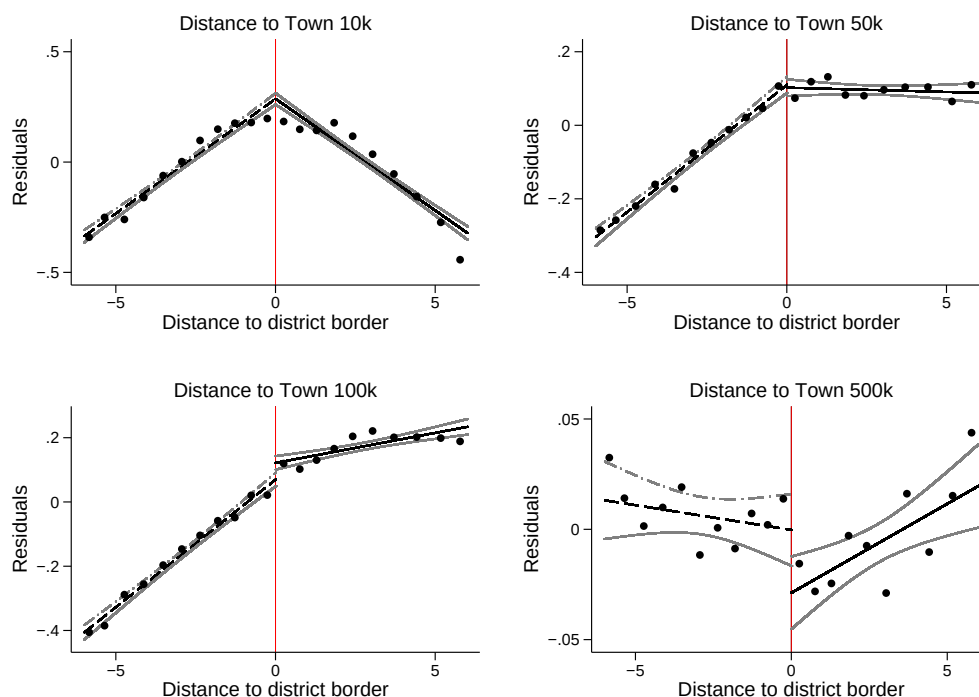
Notes: The figure plots the coefficient β_1 when we regress distance to town with population greater than 10,000 on all right hand side variables in Equation 1.1, except itself, for different bandwidths around the district border.

Figure 1.6
Distance to towns with population greater than 500,000, using different bandwidths



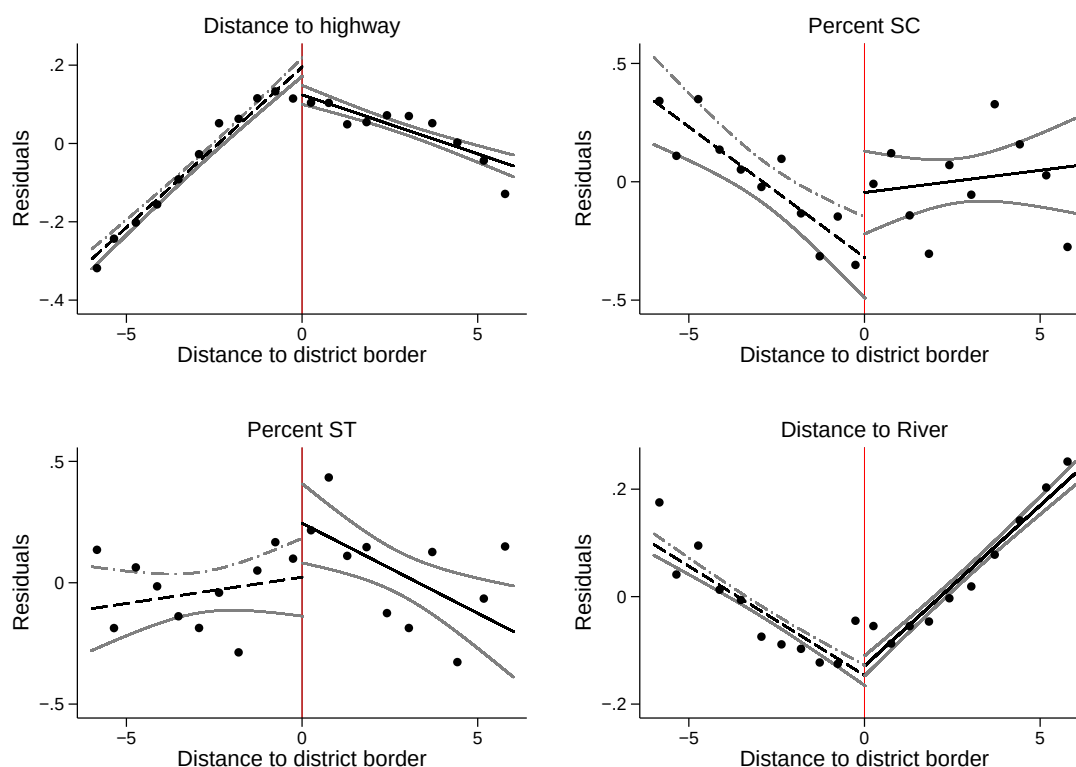
Notes: The figure plots the coefficient β_1 when we regress distance to town with population greater than 500,000 on all right hand side variables in Equation 1.1, except itself, for different bandwidths around the district border.

Figure 1.7
RD plot: balance on distance variables



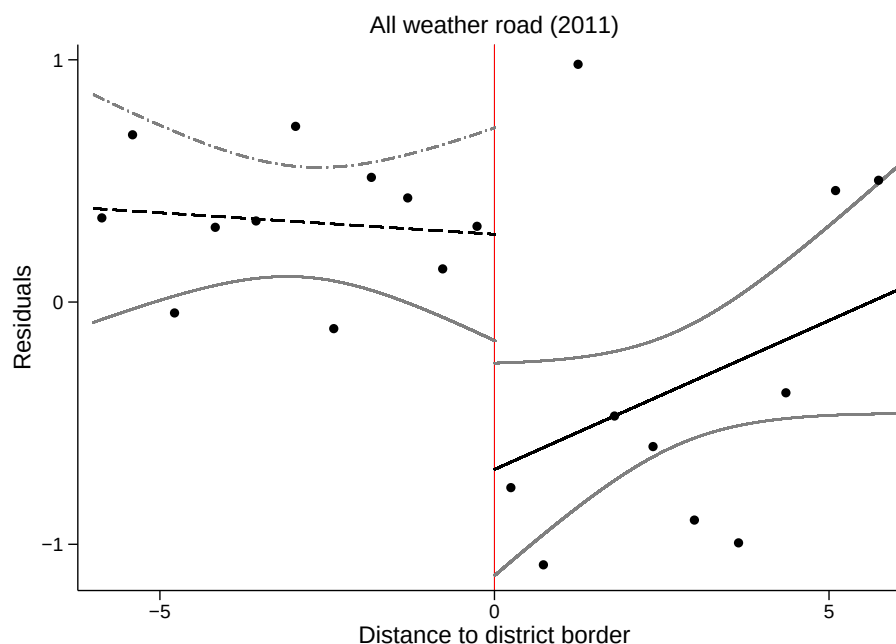
Notes: The figure plots residualised values of distance-to-town balance variables, controlling for all other balance variables other than the one on the left-hand side, as well as a linear polynomial of latitude and longitude that is estimated separately for each district border (district-district dyad). Points to the right of zero are administratively remote villages that are more distant to their district HQ than neighbouring villages on the other side of the border, while points to the left of zero are administratively proximate villages that are closer to their district HQ than neighbouring villages on the other side of the district border. Each point represents approximately 6000 observations. A linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages within 6 kilometres of an intra-state district border. It excludes villages in special category states, villages that lie along riverine district borders, and villages that lie along district borders with large elevation changes.

Figure 1.8
RD plot: balance on other variables



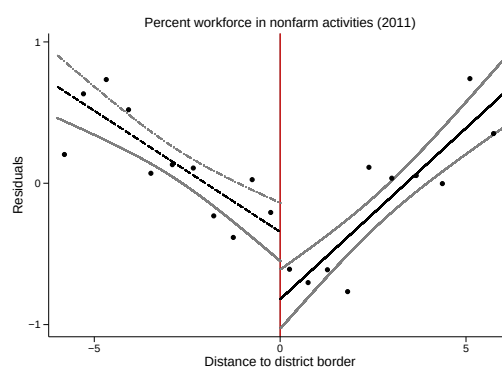
Notes: The figure plots residualised values of balance variables other than distance to towns, controlling for all balance variables other than the one on the left-hand side, as well as a linear polynomial of latitude and longitude that is estimated separately for each district border (district-district dyad). Points to the right of zero are administratively remote villages that are more distant to their district HQ than neighbouring villages on the other side of the border, while points to the left of zero are administratively proximate villages that are closer to their district HQ than neighbouring villages on the other side of the district border. Each point represents approximately 6000 observations. A linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages within 6 kilometres of an intra-state district border. It excludes villages in special category states, villages that lie along riverine district borders, and villages that lie along district borders with large elevation changes.

Figure 1.9
RD plot: access to roads



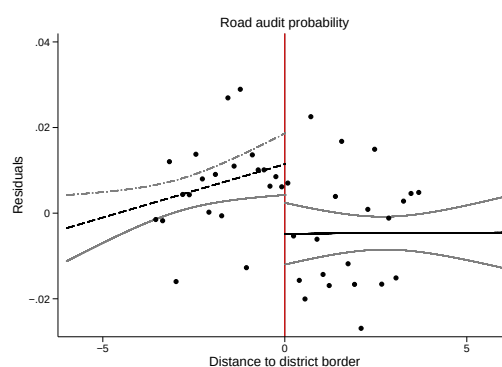
Notes: The figure plots the probability that a village has an all-weather (paved) road according to the Village Directory of Population Census 2011 controlling for all balance variables, as well as a linear polynomial of latitude and longitude that is estimated separately for each district border (district-district dyad). Points to the right of zero are administratively remote villages that are more distant to their district HQ than neighbouring villages on the other side of the border, while points to the left of zero are administratively proximate villages that are closer to their district HQ than neighbouring villages on the other side of the district border. Each point represents approximately 6000 observations. A linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages within 6 kilometres of a district border. It excludes villages in special category states, villages that lie along riverine district borders, and villages that lie along district borders with large elevation changes.

Figure 1.10
RD plot: rural non-farm employment



Notes: The figure plots the residualised share of village workforce engaged in non-agricultural activities according to the 2011 Population Census, controlling for all balance variables, as well as a linear polynomial of latitude and longitude that is estimated separately for each district border (district-district dyad). Points to the right of zero are administratively remote villages that are more distant to their district HQ than neighbouring villages on the other side of the border, while points to the left of zero are administratively proximate villages that are closer to their district HQ than neighbouring villages on the other side of the district border. Each point represents approximately 6,000 observations. A linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages within 6 kilometres of a district border. It excludes villages in special category states, villages that lie along riverine district borders, and villages that lie along district borders with large elevation changes.

Figure 1.11
RD plot: road monitoring



Notes: The figure plots the probability that, conditional on receiving a road under the PMGSY program, the village road was audited by a national quality monitoring inspector, controlling for all balance variables, as well as a linear polynomial of latitude and longitude that is estimated separately for each district border (district-district dyad). Points to the right of zero are administratively remote villages that are more distant to their district HQ than neighbouring villages on the other side of the border, while points to the left of zero are administratively proximate villages that are closer to their district HQ than neighbouring villages on the other side of the district border. Each point represents approximately 1,750 observations. A linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages within 6 kilometres of a district border. It excludes villages in special category states, and villages that lie along district borders with large elevation changes.

Appendix

1.A Additional figures and tables

Table 1.A1
Rural access to public goods, after dropping river borders

	β_1 (se)	Sample mean	N
Villages with:			
Amenities provided by districts			
Paved roads	-1.583 (.437)***	69.16%	157,395
Govt primary schools	-0.840 (0.267)***	82.97%	157,082
Mobile health clinic	-0.376 (0.098)***	1.88%	157,339
Share of land irrigated	-0.688 (0.248)***	44.77%	157,450
Treated water	-0.908 (0.295)***	22.05%	155,928
Amenities provided by central govt			
Post office	-0.017 (0.193)	9.86%	155,927
Polling station	-0.337 (0.301)	67.9%	155,914
National highway	-0.591 (0.180)***	4.86%	155,928
Electricity	-0.677 (0.396)*	54.81%	157,450
Amenities provided by village councils			
Gravel road	0.292 (0.323)	85.62%	155,888
Community toilet complex	0.002 (0.101)	2.25%	155,928
Community well	-0.041 (0.395)	56.93%	155,928
Tubewell	-0.276 (0.445)	50.37%	155,928
Bandwidth: 6 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1.1 for rural access to different public goods. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. All left hand side variables, except for “share of land irrigated,” are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For “share of land irrigated,” we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 1.A2
Rural incomes, assets, employment, after dropping river borders

	β_1 (se)	Sample mean	N
Employment structure			
Non-farm share	-0.852 (0.175)***	25.99%	155,754
Govt salary share	-0.115 (0.035)***	3.68%	149,294
Enterprise share	0.031 (0.048)	1.07%	149,292
Incomes			
Average monthly income	-56.284 (15.820)***	4,872	149,291
Household assets			
Solid wall share	-0.970 (0.192)***	52.81%	149,292
Solid roof share	-0.576 (0.177)***	44.03%	149,292
Vehicle ownership share	-0.538 (0.131)***	19.28%	149,294
Literacy			
Literacy rate	-0.265 (0.075)***	55.48%	155,928
Bandwidth: 6 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1.1 for different rural economic outcomes. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. “Nonfarm share” is the share of main workers in the village that do not report their main occupation as cultivators or agricultural labourers, as per the Primary Census Abstract of Population Census 2011. “Govt salary share” is the share of households in the village that report a government salary as their main source of income in the Socioeconomic and Caste Census 2012 (SECC 2012). “Enterprise share” is the share of households in the village that report income from an enterprise as the main source of income in SECC 2012. “Average monthly income” is calculated as explained in Section 1.3.2. “Solid wall share”, “solid roof share”, and “vehicle ownership share” are, respectively, the share of households that have a solid wall, the share of households that have a solid roof, and the share of households that own at least 1 vehicle as per SECC 2012. Finally, “literacy rate” is the share of village population that is literate according to the Primary Census Abstract of the 2011 Population Census. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 1.A3
Rural access to different public goods using absolute measure of administrative remoteness

Villages with:	β_1	(se)	Sample mean	N
Amenities provided by districts:				
Paved roads	-0.060	(0.015)***	69.77%	179,474
Govt primary schools	-0.033	(0.010)***	83.04%	179,178
Mobile health clinic	-0.005	(0.005)	1.72%	179,409
Share of land irrigated	-0.045	(0.009)***	44.78%	179,572
Treated water	-0.023	(0.011)**	22.08%	177,868

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

Notes: This table presents regression coefficients from estimating Equation 1.2 for rural access to different public goods. β_1 is the change in access to the public good for 1 kilometre change in the absolute treatment variable, $Distance_{DistrictHQ_v}$, the village's distance to its district headquarters. All left hand side variables, except for "share of land irrigated," are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For "share of land irrigated," we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 1.A4
Rural access to different public goods using absolute measure of administrative remoteness

Villages with:	β_1	(se)	Sample mean	N
Amenities provided by districts:				
Paved roads	-1.417	(0.354)***	69.77%	179,474
Govt primary schools	-0.779	(0.236)***	83.04%	179,178
Mobile health clinic	-0.118	(0.118)	1.72%	179,409
Share of land irrigated	-1.063	(0.213)***	44.78%	179,572
Treated water	-0.543	(0.260)**	22.08%	177,868
District Anderson Index	-0.045	(0.007)***	0.0097	179,572

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

Notes: This table presents regression coefficients from estimating Equation 1.2 for rural access to different public goods. β_1 is the change in access to the public good for 1 standard deviation (23.61 km) change in the absolute treatment variable, $DistanceDistrictHQ_v$, the village's distance to its district headquarters. All left hand side variables, except for "share of land irrigated," are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For "share of land irrigated," we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 1.A5
Rural access to public goods, using 3 km bandwidth

	β_1 (se)	Sample mean	N
Villages with:			
Amenities provided by districts			
Paved roads	-1.490 (0.743)***	70.67%	96,539
Govt primary schools	-1.047 (0.299)***	83.30%	96,539
Mobile health clinic	-0.149 (0.108)	1.80%	96,658
Share of land irrigated	-0.597 (0.270)**	44.85%	96,729
Treated water	-1.224 (0.330)***	21.73%	95,846
Amenities provided by federal govt			
Post office	0.097 (0.233)	9.88%	95,845
Polling station	-0.036 (0.336)	68.26%	95,837
National highway	-0.047 (0.216)	4.88%	95,846
Electricity	-0.422 (0.434)	55.00%	96,729
Amenities provided by village councils			
Gravel road	0.540 (0.361)	85.71%	95,830
Community toilet complex	-0.190 (0.120)	2.21%	95,846
Community well	-0.407 (0.427)	56.57%	95,846
Tubewell	0.204 (0.497)	50.17%	95,846
Bandwidth: 3 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1.1 for rural access to different public goods. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. All left hand side variables, except for “share of land irrigated,” are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For “share of land irrigated,” we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 1.A6
Rural incomes, assets, employment, using 3 km bandwidth

	β_1 (se)	Sample Mean	N
Employment structure			
Non-farm share	-0.952 (0.191)***	26.07%	95,640
Govt salary share	-0.069 (0.040)*	3.62%	91,808
Enterprise share	-0.044 (0.066)	1.13%	91,806
Incomes			
Average monthly income	-28.944 (17.834)	4,869	91,805
Household assets			
Solid wall share	-0.742 (0.214)***	52.8%	91,807
Solid roof share	-0.448 (0.189)**	44.06%	91,807
Vehicle ownership share	-0.159 (0.137)	19.3%	91,808
Literacy			
Literacy rate	-0.253 (0.080)***	55.71%	95,846
Bandwidth: 3 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1.1 for different rural economic outcomes. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. "Non-farm share" is the share of main workers in the village that do not report their main occupation as cultivators or agricultural labourers, as per the Primary Census Abstract of Population Census 2011. "Govt salary share" is the share of households in the village that report a government salary as their main source of income in the Socioeconomic and Caste Census 2012 (SECC 2012). "Enterprise share" is the share of households in the village that report income from an enterprise as the main source of income in SECC 2012. "Average monthly income" is calculated as explained in Section 1.3.2. "Solid wall share", "solid roof share", and "vehicle ownership share" are, respectively, the share of households that have a solid wall, the share of households that have a solid roof, and the share of households that own at least 1 vehicle as per SECC 2012. Finally, "literacy rate" is the share of village population that is literate according to the Primary Census Abstract of the 2011 Population Census. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 1.A7
Rural access to public goods, using 40 kilometre segments

	β_1 (se)	Sample mean	N
Villages with:			
Amenities provided by district			
Paved roads	-1.204 (0.438)***	69.57%	172,955
Govt primary schools	-0.993 (0.274)***	83.13%	172,647
Mobile health clinic	-0.152 (0.113)	1.94%	172,888
Share of land irrigated	-0.334 (0.256)	44.33%	173,053
Treated water	-0.868 (0.324)***	23.02%	171,342
Amenities provided by central govt			
Post office	-0.027 (0.189)	9.27%	171,341
Polling station	0.137 (0.305)	68.21%	171,328
National highway	0.066 (0.176)	4.94%	171,342
Electricity	-0.403 (0.464)	56.85%	173,053
Amenities provided by village councils			
Gravel road	-0.111 (0.314)	85.53%	171,302
Community toilet complex	-0.050 (0.104)	2.23%	171,342
Community well	-0.323 (0.400)	55.16%	171,342
Tubewell	0.032 (0.464)	51.36%	171,342
Bandwidth: 6 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1.1 for rural access to different public goods. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. All left hand side variables, except for “share of land irrigated,” are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For “share of land irrigated,” we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 1.A8
Rural incomes, assets, employment, using 40 kilometre segments

	β_1 (se)	Sample mean	N
Employment structure			
Non-farm share	-0.338 (0.185)*	26.60%	171,252
Govt salary share	-0.051 (0.040)	3.70%	164,710
Enterprise share	-0.038 (0.040)	1.16%	164,708
Incomes			
Average monthly income	-56.014 (15.811)***	4,869	164,707
Household assets			
Solid wall share	-0.826 (0.209)***	53.63%	164,706
Solid roof share	-0.193 (0.197)	44.70%	164,706
Vehicle ownership share	-0.349 (0.133)***	19.78%	164,710
Literacy			
Literacy rate	-0.118 (0.077)	56.23%	171,342
Bandwidth: 6 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1.1 for different rural economic outcomes. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. “Nonfarm share” is the share of main workers in the village that do not report their main occupation as cultivators or agricultural labourers, as per the Primary Census Abstract of Population Census 2011. “Govt salary share” is the share of households in the village that report a government salary as their main source of income in the Socioeconomic and Caste Census 2012 (SECC 2012). “Enterprise share” is the share of households in the village that report income from an enterprise as the main source of income in SECC 2012. “Average monthly income” is calculated as explained in Section 1.3.2. “Solid wall share”, “solid roof share”, and “vehicle ownership share” are, respectively, the share of households that have a solid wall, the share of households that have a solid roof, and the share of households that own at least 1 vehicle as per SECC 2012. Finally, “literacy rate” is the share of village population that is literate according to the Primary Census Abstract of the 2011 Population Census. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 1.A9
Rural access to public goods, using quadratic polynomials

	β_1 (se)	Sample mean	N
Villages with:			
Amenities provided by district			
Paved roads	-0.950 (.377)**	69.56%	180,264
Govt primary schools	-0.874 (0.226)***	83.10%	179,904
Mobile health clinic	-0.231 (0.084)***	1.85%	180,135
Share of land irrigated	-0.573 (0.216)***	44.76%	180,299
Treated water	-1.126 (0.261)***	21.82%	178,609
Amenities provided by central govt			
Post office	0.047 (0.161)	9.99%	178,608
Polling station	-0.035 (0.252)	68.21%	178,595
National highway	-0.136 (0.153)	5.02%	178,609
Electricity	-1.020 (0.355)***	55.26%	180,299
Amenities provided by village councils			
Gravel road	0.081 (0.277)	85.57%	178,569
Community toilet complex	0.001 (0.083)	2.18%	178,609
Community well	-0.497 (0.347)	56.69%	178,609
Tubewell	-0.115 (0.387)	50.49%	178,609
Bandwidth: 6 km			
Geographic polynomial: Quadratic in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1.1 for rural access to different public goods. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. All left hand side variables, except for “share of land irrigated,” are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For “share of land irrigated,” we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 1.A10
Rural incomes, assets, employment, using quadratic polynomials

	β_1 (se)	Sample mean	N
Employment structure			
Non-farm share	-1.094 (0.157)***	26.58%	178,425
Govt salary share	-0.075 (0.033)**	3.75%	171,388
Enterprise share	-0.073 (0.038)*	1.12%	171,386
Incomes			
Average monthly income	-28.945 (13.545)***	4,893	171,385
Household assets			
Solid wall share	-0.801 (0.169)***	53.22%	171,384
Solid roof share	-0.177 (0.162)	44.36%	171,384
Vehicle ownership share	-0.424 (0.110)***	19.32%	171,388
Literacy			
Literacy rate	-0.239 (0.066)***	55.99%	178,609
Bandwidth: 6 km			
Geographic polynomial: Quadratic in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1.1 for different rural economic outcomes. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. “Nonfarm share” is the share of main workers in the village that do not report their main occupation as cultivators or agricultural labourers, as per the Primary Census Abstract of Population Census 2011. “Govt salary share” is the share of households in the village that report a government salary as their main source of income in the Socioeconomic and Caste Census 2012 (SECC 2012). “Enterprise share” is the share of households in the village that report income from an enterprise as the main source of income in SECC 2012. “Average monthly income” is calculated as explained in Section 1.3.2. “Solid wall share”, “solid roof share”, and “vehicle ownership share” are, respectively, the share of households that have a solid wall, the share of households that have a solid roof, and the share of households that own at least 1 vehicle as per SECC 2012. Finally, “literacy rate” is the share of village population that is literate according to the Primary Census Abstract of the 2011 Population Census. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 1.A11
Rural access to public goods, using spatially clustered standard errors

	β_1 (se)	Sample mean	N
Villages with:			
Amenities provided by districts:			
Paved roads	-1.232 (0.449)***	69.56%	180,264
Govt primary schools	-0.929 (0.275)***	83.10%	179,904
Mobile health clinic	-0.312 (0.098)***	1.84%	180,135
Share of land irrigated	-0.710 (0.250)***	44.76%	180,299
Treated water	-1.180 (0.299)***	21.82%	178,609
Amenities provided by federal govt			
Post office	0.042 (0.213)	9.99%	178,608
Polling station	-0.295 (0.322)	68.21%	178,595
National highway	-0.255 (0.199)	5.02%	178,609
Electricity	-0.535 (0.402)	55.26%	180,299
Amenities provided by village councils			
Gravel road	0.277 (0.317)	85.57%	178,569
Community toilet complex	-0.017 (0.112)	4.95%	178,609
Community well	-0.419 (0.395)	56.69%	178,609
Tubewell	-0.256 (0.457)	50.49%	178,609
Bandwidth: 6 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1.1 for rural access to different public goods. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. All left hand side variables, except for “share of land irrigated,” are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For “share of land irrigated,” we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered using 10x10 km grid cells to allow for spatial correlation in standard errors.

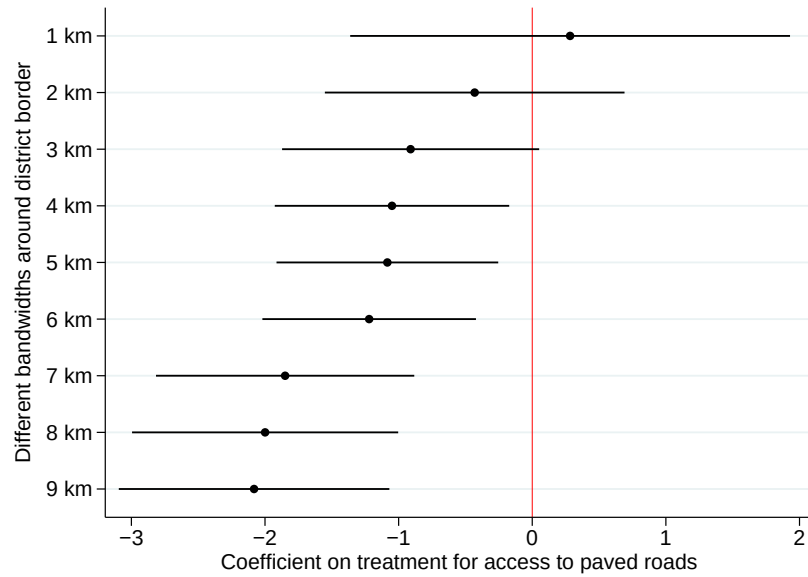
Table 1.A12
Rural incomes, assets, employment, using spatially clustered standard errors

	β_1 (se)	Sample mean	N
Employment structure			
Non-farm share	-0.729 (0.182)***	26.58%	178,425
Govt salary share	-0.076 (0.038)**	3.75%	171,388
Enterprise share	-0.035 (0.051)	1.12%	171,386
Incomes			
Average monthly income	-50.457 (17.323)***	4,893	171,385
Household assets			
Solid wall share	-0.871 (0.197)***	53.22%	171,384
Solid roof share	-0.384 (0.176)**	44.36%	171,384
Vehicle ownership share	-0.365 (0.134)***	19.32%	171,388
Literacy			
Literacy rate	-0.261 (0.077)***	55.99%	178,609
Bandwidth: 6 km			
Geographic polynomial: Linear in latitude and longitude			

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

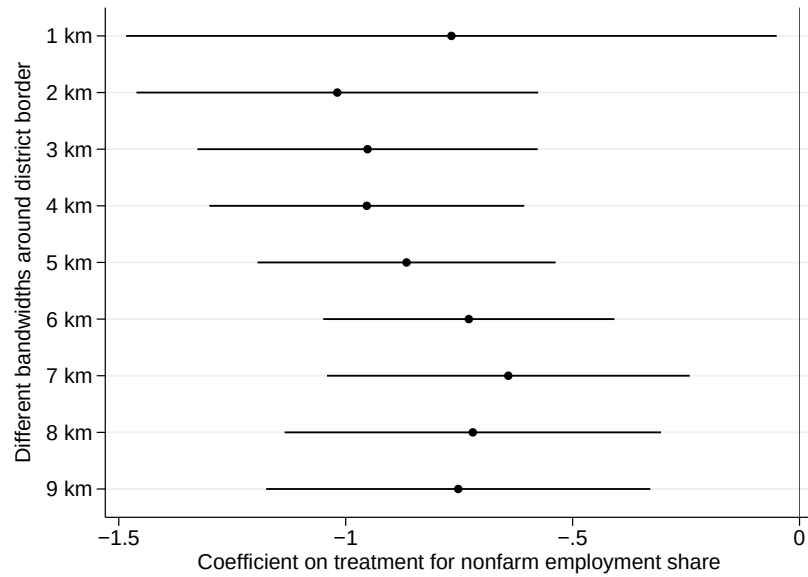
Notes: This table presents regression coefficients from estimating Equation 1.1 for different rural economic outcomes. β_1 is the coefficient on the binary treatment variable, *RemoteSide_{d,s}*, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. “Nonfarm share” is the share of main workers in the village that do not report their main occupation as cultivators or agricultural labourers, as per the Primary Census Abstract of Population Census 2011. “Govt salary share” is the share of households in the village that report a government salary as their main source of income in the Socioeconomic and Caste Census 2012 (SECC 2012). “Enterprise share” is the share of households in the village that report income from an enterprise as the main source of income in SECC 2012. “Average monthly income” is calculated as explained in Section 1.3.2. “Solid wall share”, “solid roof share”, and “vehicle ownership share” are, respectively, the share of households that have a solid wall, the share of households that have a solid roof, and the share of households that own at least 1 vehicle as per SECC 2012. Finally, “literacy rate” is the share of village population that is literate according to the Primary Census Abstract of the 2011 Population Census. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered using 10x10 km grid cells to allow for spatial correlation in standard errors.

Figure A1
Differences in paved road access, using different bandwidths



Notes: This figure plots the coefficient β_1 obtained from estimating Equation 1.1 for village access to paved roads, using a number of different bandwidths around district border.

Figure A2
Differences in non-farm employment share, using different bandwidths



Notes: This figure plots the coefficient β_1 obtained from estimating Equation 1.1 for village nonfarm employment share, using a number of different bandwidths around district border.

CHAPTER 2

THE RURAL EFFECTS OF URBAN GROWTH

with Sam Asher (The World Bank) and Paul Novosad (Dartmouth College)

2.1 Introduction

There is a close economic relationship between urban and rural areas in all countries. While urban locations provide large goods and labour markets that bring together workers and firms to produce manufactured goods and services, villages provide labour and agricultural surpluses which make urban production possible. These linkages have formed the core of structural transformation models (Lewis, 1954; Kuznets, 1957; Foster and Rosenzweig, 2007; Michaels et al., 2012), in which growth in rural agricultural productivity drives the urbanisation process in the country.¹ However, with growing world trade in agriculture, and stronger links between developing world cities and global supply chains, the reliance of urban areas on the rural hinterland has reduced (Glaeser, 2014). In such a world, urban growth is more often driven by changes in global and national demand patterns. How does this urban growth affect the proximate rural economy? Is rapid employment growth in urban areas complementary to rural employment growth, or does it drive labour exit from the villages? Understanding these urban-to-rural linkages will help developing countries design

This version August 2018. We thank seminar participants at the CSAE Workshop (Oxford), and the CSAE Conference (Oxford). All remaining errors are our own.

¹See, for example, Nunn and Qian (2011) and Bustos et al. (2016) for empirical evidence on the effect of changes in the agricultural sector on urbanisation. Recent empirical papers have investigated the effect of other exogenous rural shocks, such as climate change (Henderson et al., 2017), exports of natural resources (Gollin et al., 2016), deforestation (Richards and VanWey, 2015), and discoveries of mineral resources (Fafchamps et al., 2016) on the urbanisation process in developing countries.

policies that can enable rural areas to capitalise on the recent success of urban areas.

In this chapter, we study the effect of urban employment growth on the employment structure in neighbouring rural areas. More specifically, we evaluate the change in non-farm employment growth in villages when there are exogenous shocks to employment in neighbouring towns. There are two major challenges to answering this question. The first challenge is data: economic data are typically available at aggregate levels, covering multiple towns and their rural catchment areas. It is thus not possible to relate town-specific shocks to the immediate geography of the towns. The second barrier is causal identification: because towns and villages are located in close geographic proximity, it is difficult to isolate exogenous shocks to employment in towns alone.

We resolve the first challenge by assembling a panel dataset on employment in different industries in the universe of Indian towns and villages in 1998 and 2005. We match this to population census data from 2001 and 2011. To investigate the geography of urban-rural shocks, we further obtain geo-coordinates for these towns and villages. Our data cover the universe of all non-farm firms and workers in India, and permit us to match each village to the employment structure and growth patterns of its nearest urban location.

To obtain causal estimates for the effect of urban employment growth on the rural employment structure, we construct exogenous shocks to urban employment in India using a shift-share (or “Bartik”) instrument variable (Bartik, 1991; Blanchard et al., 1992). For every town, the Bartik instrument is constructed by interacting the baseline employment share of each manufacturing industry, with the change in employment in that industry in all *other* towns in the country. We use the predicted growth in town-level manufacturing employment to instrument for actual growth in town-level manufacturing employment between 1998 and 2005.

We find that higher exogenous growth in manufacturing employment in the nearest town has no statistically significant effect on non-farm employment growth in the

village, though the direction of the effect is negative. When we decompose rural non-farm employment into manufacturing and service sector employment, we find that higher exogenous growth in manufacturing employment in the nearest town reduces the growth rate of manufacturing employment in the village. The IV estimate suggests an elasticity of -0.180, i.e. 10% higher growth in manufacturing employment in the nearest town reduces the growth rate of village manufacturing employment by 1.8%. These results favour the hypothesis that forces of agglomeration in towns drive labour exit from the manufacturing sector of surrounding rural areas.

Another way in which urban growth can affect the rural economy in nearby villages is by increasing the rural demand for services, such as consumer services and retail (Cali and Menon, 2012). This increased demand for services can lead to higher employment growth in the services sector. However, we find no statistically significant effect of urban manufacturing employment growth on employment growth in the services sector in neighbouring villages.

The Bartik instrument is valid, conditional on the covariates, if the baseline employment shares in the nearest town, interacted with national-level trends, are not correlated with changes in rural employment. We deal with three specific threats to this identifying assumption. First, we include district fixed effects throughout our analysis. These fixed effects absorb the effect of confounding factors such as local policies, governance quality, and institutions on rural and urban employment outcomes. Second, we control for the size of towns and villages at baseline, since small and large settlements may specialise in different types of industries. Third, since the Bartik instrument is constructed as the weighted average of baseline employment shares, we test whether these baseline employment shares are correlated with potential town-level confounding variables. We find that our results continue to hold even after controlling for these town-level confounding variables. Therefore, we argue that conditional on the covariates and the district fixed effects, the Bartik instrument isolates shocks to urban employment, and we can exploit these shocks to estimate the effect of urban

employment growth on employment outcomes of neighbouring villages.

To test whether or not the Bartik instrument suffers from a “weak instruments” problem, we report the value of the Kleibergen-Paap rk Wald F-statistic for each of our IV regressions. This statistic is almost always substantially larger than 10, the conventional threshold for rejecting the null that the instruments are weak. This suggests that our instrumental variable is not a “weak instrument.”

When we evaluate the rural effects of urban employment growth separately by the size of the nearest town, we find that the negative effects of urban growth are driven by villages located close to the larger towns in our sample. The larger towns and cities are precisely the places where we expect the agglomeration forces of manufacturing industries to be the strongest (Ryan, 2012). 10% higher growth in manufacturing employment in the nearest town, when the nearest town is larger than the median town in our sample, reduces the growth rate of village manufacturing employment by 2.24%. For villages located close to smaller towns, manufacturing employment growth in the nearest town has no statistically significant effect on either non-farm employment growth, or manufacturing employment growth.

Finally, we find that the negative effects of urban manufacturing employment growth are more pronounced for villages that are ill suited for manufacturing industries in the first place. This includes villages without paved roads, and villages that are most distant from their nearest town. For example, for the sub-sample of villages with paved roads in the 2001 Population Census, we find no statistically significant relationship between urban manufacturing employment growth and rural non-farm or manufacturing employment growth. But for villages without paved roads, manufacturing employment growth in the nearest town reduces the rate of both non-farm and manufacturing employment growth. Similarly, when we examine heterogeneity by distance to the nearest town, the magnitude of the effect is the largest for villages located farthest from a town or city.

Is manufacturing employment growth in villages lower because workers previ-

ously employed in manufacturing migrate out of the village, or do these workers start commuting to the nearest town? This question is difficult to answer because the Economic Census reports the number of jobs in a particular location, but not the residential location of the workers. We nevertheless attempt an approximate answer to this question by examining the effect of urban employment growth during 1998–2005 on village population growth rate during 2001–2011 (the closest years for which we have village demographic data). We find no statistically significant effect of urban employment growth on rural population growth, suggesting that village residents are increasingly commuting to the nearest town for work. We also find no effect on the change in the village's employment–population ratio between 1998 and 2005.

These results contribute to several strands of literature. First, by investigating the effect of urban growth on the rural employment structure, we contribute new results to the extensive literature on structural transformation. Most papers in this literature have studied the consequences of changes originating in rural areas – such as growth in agricultural productivity – on urbanisation and structural transformation in developing countries. A number of recent empirical contributions have followed this causal chain, by looking at the effects of new crops (Nunn and Qian, 2011), new agricultural technologies (Bustos et al., 2016), climate change (Henderson et al., 2017), natural resource exports (Gollin et al., 2016), deforestation (Richards and VanWey, 2015), and mineral resource discoveries (Fafchamps et al., 2016).

The set of papers that examines the consequences of urban changes on the rural economy is relatively smaller, and has often not established causal links between urban growth and rural structure. Most of these papers find that higher rates of urbanisation lead to better rural outcomes. For example, Kochar (2004) argues that higher wages in urban areas increase schooling in rural areas. Cali and Menon (2012) finds that a higher rate of urbanisation in Indian districts is correlated with a lower rate of poverty. Fafchamps and Shilpi (2005) finds that villages located close to urban centres in Nepal have more diversified and market-oriented activities. However, most of these

papers only establish a correlation between urban and rural outcomes. In addition to the challenges with identification, these papers work with data from more aggregate economic units. For example, Cali and Menon (2012) use district-level data from India. But with an average of ten towns per district, they are not able to map out the geography of urban-to-rural spillovers. We take the literature forward by proposing an identification strategy that has not previously been used to study this question, and by using high-resolution economic data from the universe of Indian towns and villages. While we cannot rule out that higher urban employment growth leads to better rural outcomes, our results add an important caveat to the general story in this literature: that urban growth can lead to labour exit from the surrounding rural economy. What this means for rural incomes and productivity remains to be assessed.

Secondly, by examining the importance of roads and urban proximity in transmitting urban shocks to the rural economy, we contribute new results to a growing literature that studies the channels through which transport investments shape rural economic outcomes (Atkin and Donaldson, 2015; Storeygard, 2016; Asher and Novosad, 2018). Our results here are consistent with a model in which the existence of paved roads makes the village more attractive for manufacturing firms. Therefore, even if there is higher manufacturing employment growth in the neighbouring towns and cities, manufacturing employment in the village does not decrease. Finally, by using Bartik instruments to study the effect of urban employment growth on the rural employment structure in a large developing country, we suggest a new application for a widely used instrument (Blanchard et al., 1992; Moretti and Thulin, 2013; Diamond, 2016).

The rest of the chapter is structured as follows. In Section 2.2, we explain our data sources. In Section 2.3, we outline the empirical strategy followed in this chapter. Section 2.4 presents and discusses the results, and Section 2.5 concludes.

2.2 Data

We bring together three sources of data in this chapter: the Population Census, the Economic Census and geographic coordinates for all towns and villages covered by the two censuses. The Population Census reports the total population, demographic characteristics and public goods for all towns and villages in the country. We use Population Census data from 2001 and 2011. The Economic Census enumerates all non-farm establishments in the country, with basic information such as number of employees and industrial classification. Its strength lies in its comprehensiveness and rich detail on the firms' industrial classification. In this chapter, we use Economic Census data from 1998 and 2005.

We obtain geo-coordinates for all towns and villages in India and match this to our combined Economic–Population Census dataset. We use these geo-coordinates to calculate the geodesic distance from each village to several different towns in the area. This allows us to define, for each village, the nearest town, as well as the nearest town above different population thresholds.

While defining towns and villages in this chapter, we follow the administrative classification of a settlement as village or town. For the towns in our sample, we calculate the employment–population ratio (using employment data from Economic Census 1998 and Population Census 2001) and drop towns with employment–population ratio smaller than the 1st percentile (0.002) and greater than the 99th percentile (0.581). We also trim the sample of villages at the 1st and the 99th percentile using village population in 2001. This leaves us with 4,723 towns and approximately 440,000 villages in our sample.

Table 2.1 shows the summary statistics for the towns and villages in our sample. The average town in our sample has a population of 33,691 in 2001, with a total non-farm workforce of 4,651 workers. 1,291 of these workers are employed in the manufacturing sector. During 1998–2005, the manufacturing workforce in these towns grew

by approximately 20%.

The average village in our sample has a population of 1,110 in 2001, with a non-farm workforce of 45 workers according to the Economic Census 1998. 14 of these workers are employed in the manufacturing sector, and 31 in the services sectors. The non-farm workforce grew by 42% during 1998-2005, and the manufacturing workforce by 18.5%. The average village in our sample is 12 kilometres from its nearest town, and 53.1% of the villages had a paved road in 2001.

2.3 Empirical Strategy

We want to study the effect of employment growth in the manufacturing sector in towns on employment growth in the neighbouring villages. For each village v in our sample, we observe employment growth in the nearest town t . Essentially, we want to estimate the following regression equation:

$$\begin{aligned} \Delta \ln Employment_{v,1998-05}^{Sector} = & \beta_1 \Delta \ln Employment_{t,1998-05}^{Manufacturing} + \beta_2 \ln Employment_{v,1998}^{Sector} \\ & + \beta_3 \ln Employment_{t,1998}^{Manufacturing} + \delta + \epsilon_v \end{aligned} \quad (2.1)$$

where $\Delta \ln Employment_{v,1998-05}^{Sector}$ is the log change in employment in village v in *Sector* between 1998 and 2005, where *Sector* could be the whole non-farm sector, or its specific components such as manufacturing or services. $\Delta \ln Employment_{t,1998-05}^{Manufacturing}$ is the log change in manufacturing sector employment in town t between 1998 and 2005, where town t is the closest town to village v . $\ln Employment_{v,1998}^{Sector}$ is the log of baseline employment in *Sector* in village v . $\ln Employment_{t,1998}^{Manufacturing}$ is the log of baseline employment in the manufacturing sector in town t . δ is the district fixed effect, which accounts for the effects of local policies and institutions. Finally, ϵ_v is the error term. To allow for spatial correlation in the error term, we construct 0.1 x 0.1 degree grid cells

and cluster our standard errors at the level of these grid cells.

β_1 is then the correlation between manufacturing employment growth in town t and employment growth in *Sector* in village v . Since our outcome variables are expressed in logs, β_1 is the elasticity of village employment growth with respect to town employment growth. In other words, manufacturing employment growth of 1% in town t is associated with employment growth of $\beta_1\%$ in the non-farm (or manufacturing or services) sector in village v .

The baseline employment levels allow us to account for the effect of initial town and village size on employment growth rates. For example, we might expect villages with low employment levels at baseline to have higher growth rates because of a small base. Further, we include district fixed effects to account for differences in institutions and government policies at the district level. Despite these controls and fixed effects, we cannot attribute a causal interpretation to the β_1 parameter, since it is likely to suffer from omitted variable bias. For example, since town t and village v are located in close geographic proximity, there are several confounding factors that could together drive employment growth in both towns as well as nearby villages.

To isolate exogenous shifts in urban employment demand, we construct a shift-share (or Bartik) instrument, following Moretti and Thulin (2013). For each town t , this instrument is essentially a weighted average of national growth in different manufacturing sub-sectors in all *other* towns in the country, with the weights coming from the employment share of each manufacturing sub-sector in town t at baseline.

The first-stage regression equation for this instrumental variable (IV) strategy is:

$$\begin{aligned} \Delta \ln Employment_{t,1998-05}^{Manufacturing} &= \mu_1 \Delta \ln Bartik Employment_{t,1998-05}^{Manufacturing} \\ &+ \mu_2 \ln Employment_{t,1998}^{Manufacturing} + \mu_3 \ln Employment_{v,1998}^{Sector} + \delta + \varepsilon_v \end{aligned} \quad (2.2)$$

where $\Delta \ln Employment_{t,1998-05}^{Manufacturing}$ is the log change in manufacturing sector employ-

ment in town t between 1998 and 2005, where town t is the *nearest* town for atleast one village in our sample. $\Delta \ln \text{BartikEmployment}_{t,1998-05}^{\text{Manufacturing}}$ is the log change in manufacturing sector employment in town t between 1998 and 2005 as predicted by the Bartik instrument. As before, $\ln \text{Employment}_{t,1998}^{\text{Manufacturing}}$ is the log of baseline employment in manufacturing sector in town t , $\ln \text{Employment}_{v,1998}^{\text{Sector}}$ is the log of baseline employment in *Sector* in village v in our sample, and δ is the district fixed effect. ϵ_v is the error term, again clustered at the level of 0.1×0.1 degree grid-cells.

The reduced form equation is:

$$\begin{aligned} \Delta \ln \text{Employment}_{v,1998-05}^{\text{Sector}} = & \beta_1 \Delta \ln \text{BartikEmployment}_{t,1998-05}^{\text{Manufacturing}} \\ & + \beta_2 \ln \text{Employment}_{v,1998}^{\text{Sector}} + \beta_3 \ln \text{Employment}_{t,1998}^{\text{Manufacturing}} + \delta + \epsilon_v \end{aligned} \quad (2.3)$$

where $\Delta \ln \text{Employment}_{v,1998-05}^{\text{Sector}}$ is the log change in *Sector* employment in village v between 1998 and 2005, where *Sector* could be the entire non-farm sector, or manufacturing or services sectors. In some specifications, we estimate Equation 2.3 for village population growth rates and change in village employment–population ratios. $\Delta \ln \text{BartikEmployment}_{t,1998-05}^{\text{Manufacturing}}$ is the log change in manufacturing employment in town t between 1998 and 2005, as predicted by the Bartik instrument, where town t is the nearest town to village v . $\ln \text{Employment}_{v,1998}^{\text{Sector}}$ is the log of baseline employment in *Sector* in village v , and $\ln \text{Employment}_{t,1998}^{\text{Manufacturing}}$ is the log of baseline manufacturing employment in town t . δ is the district fixed effect. ϵ_v is the error term, clustered at the level of 0.1×0.1 degree grid-cells.

The IV estimate is the ratio of the reduced form coefficient, β_1 , and the first-stage coefficient, μ_1 . 1% growth in manufacturing employment in town t leads to $\beta_1/\mu_1\%$ growth in employment in the relevant sector in village v , controlling for baseline employment in both the town and the village, and accounting for district-level variation in policies and institutions through the district fixed effects. We describe the construc-

tion of the Bartik variable, $\Delta \ln Bartik Employment_{t,1998-05}^{Manufacturing}$, in the next section.

2.3.1 Constructing the Bartik instrument

We start by aggregating national employment in all the manufacturing sub-sectors in the Economic Census dataset. These correspond to National Industrial Classification (NIC) codes 19-122 and 150. $Employment_National_{1998}^{subsector}$ is the aggregate of a manufacturing sub-sector's employment in all towns in the sample:

$$Employment_National_{1998}^{subsector} = \sum_t Employment_{t,1998}^{subsector} \quad (2.4)$$

where t denotes a town in the sample

For each town t , we exclude its own employment in the particular manufacturing sub-sector:

$$Employment_in_all_other_towns_{t,1998}^{subsector} = Employment_National_{1998}^{subsector} - Employment_{t,1998}^{subsector} \quad (2.5)$$

Now, for each town, we evaluate the log change in employment in the given sub-sector in all *other* towns during 1998-2005:

$$\begin{aligned} \Delta \ln OtherPlaces Employment_{t,1998-05}^{subsector} &= \ln Employment_in_all_other_towns_{t,2005}^{subsector} \\ &\quad - \ln Employment_in_all_other_towns_{t,1998}^{subsector} \end{aligned} \quad (2.6)$$

The predicted level of employment for a manufacturing sub-sector in town t in 2005 is:

$$\begin{aligned} \text{BartikEmployment}_{t,2005}^{\text{subsector}} &= \text{Employment}_{t,1998}^{\text{subsector}} \cdot \\ &\quad (1 + \Delta \ln \text{OtherPlacesEmployment}_{t,1998-05}^{\text{subsector}}) \end{aligned} \quad (2.7)$$

We then aggregate predicted manufacturing employment in town t in 2005 across all the manufacturing sub-sectors:

$$\text{BartikEmployment}_{t,2005}^{\text{Manufacturing}} = \sum_{\text{subsector}} \text{BartikEmployment}_{t,2005}^{\text{subsector}} \quad (2.8)$$

Finally, we calculate predicted employment change between 1998-2005 for the manufacturing sector in town t :

$$\begin{aligned} \Delta \ln \text{BartikEmployment}_{t,1998-05}^{\text{Manufacturing}} &= \ln \text{BartikEmployment}_{t,2005}^{\text{Manufacturing}} \\ &\quad - \ln \text{Employment}_{t,1998}^{\text{Manufacturing}} \end{aligned} \quad (2.9)$$

We use the log change in predicted manufacturing employment in a town as an instrument for the actual log change in manufacturing employment in that town between 1998 and 2005.

2.4 Results

We find that higher growth rate of manufacturing employment in towns reduces the growth rate of manufacturing employment in neighbouring villages, but has no effect on employment growth rates in the services sector, or the non-farm sector as a whole. In this section, we start by testing the strength of the first stage regression (Section 2.4.1). We then evaluate the effect of manufacturing employment growth in the nearest town on employment growth in different rural sectors (Section 2.4.2). We further investigate the effects of urban growth on rural population outcomes (Sec-

tion 2.4.3). Finally, we investigate the effects separately for towns of different sizes (Section 2.4.4), and villages with different levels of connectivity to the nearest town (Section 2.4.5).

2.4.1 First stage

We first test whether the predicted employment growth rate obtained using the Bartik instrument is informative about the actual employment growth rate between 1998 and 2005. To do this, we estimate the first-stage Equation 2.2. The results are presented in Table 2.2. The first column shows the first-stage results when rural nonfarm employment growth is the outcome variable; the second column shows the first-stage results when rural manufacturing employment growth is the outcome variable; and the third column shows the first-stage results when rural service sector employment growth is the outcome variable. All regressions control for baseline values of the urban and rural employment outcomes, and include district fixed effects to account for the effects of local policies and institutions.

We find that the estimated elasticity between the predicted and the actual growth rate of urban manufacturing employment is 0.339. This implies that 10% higher value of the predicted urban manufacturing employment growth rate is correlated with a 3.39% higher value of the actual urban manufacturing employment growth rate. This coefficient is precisely estimated.

Figure 2.1 illustrates this relationship graphically, again showing a strong relationship between the actual and predicted growth rate of urban manufacturing employment in India, controlling for baseline urban manufacturing employment and absorbing the district fixed effects. We can therefore conclude that the Bartik instrument is informative for the actual growth rate of manufacturing employment in urban India.

2.4.2 Main results

We now evaluate the relationship between urban manufacturing employment growth and employment growth in the neighbouring villages. Table 2.3 presents both the OLS and the IV estimates for the three main rural outcomes: non-farm employment growth, manufacturing employment growth, and service sector employment growth between 1998 and 2005. The OLS results suggest a positive and statistically significant relationship between urban manufacturing employment growth and employment growth in rural manufacturing and services sectors, and therefore, employment growth in the rural non-farm sector as a whole. However, the OLS results are likely to be biased upwards: towns and villages located in close geographic proximity experience similar local employment shocks, and therefore it is not surprising that urban and rural employment growth rates are strongly correlated.

The IV estimates tell a different story. These estimates suggest that exogenous shocks to growth in manufacturing employment in the nearest town *reduce* the growth rate of employment in the rural manufacturing sector. The estimated elasticity is -0.18: 1% higher growth in manufacturing employment in the nearest town reduces the growth rate of rural manufacturing employment by 0.18%. It also reduces the growth rate of employment in the rural services sector, and in the rural non-farm sector as a whole; however these effects are not statistically significant. Further, rural employment growth rates for all three sectors are negatively correlated with the corresponding baseline rural employment levels: villages with fewer jobs in 1998 experience a larger increase in employment between 1998 and 2005, suggesting a degree of employment convergence in rural India.

The Kleibergen-Paap F-statistic from these IV regressions is around 63. This is substantially larger than 10, which is the conventional threshold for rejecting the null hypothesis of weak instruments. Therefore, our analysis does not suffer from the weak instruments problem.

To further test the validity of the Bartik instruments, we calculate the mean em-

ployment shares at baseline of each urban manufacturing industry in our sample. For the five industries with the highest mean employment shares at baseline,² we estimate the correlation coefficients between the baseline employment shares and a set of confounding variables observed at the town level at baseline. These confounding variables include the male percentage of the urban workforce, the percentage of the urban workforce employed in government companies, the percentage of the urban workforce that belongs to the Scheduled Castes, and the percentage of the urban workforce that belongs to the Scheduled Tribes. All these variables are observed at baseline, i.e. in the 1998 Economic Census. The correlations are presented in Table 2.4.

While for most industries, the baseline employment share is not correlated with the potential confounding variables, there are statistically significant correlations in some instances. To address this issue, we control for the baseline values of the potential town-level confounding variables in all our analyses. These town-level controls are in addition to the district fixed effects, which already absorb the effects of any local variation in governance quality, federal and state policies, as well as pre-colonial and colonial institutions. The district fixed effects as well as the town-level baseline controls strengthen the claims of exogeneity in favour of the Bartik instrument.

Table 2.5 presents the IV estimates for the main rural employment outcomes, after controlling for the baseline values of the potential confounding variables. We find that the additional controls do not change the results qualitatively. We still cannot reject the null that higher manufacturing employment growth in the nearest town does not affect employment growth in the rural non-farm or services sectors. However, higher manufacturing employment growth in the nearest town does reduce the employment growth rate in the rural manufacturing sector. The IV estimate, -0.178, is very close to the previous estimate of -0.18. This implies that 10% higher growth in manufacturing employment in the nearest town reduces the growth rate of rural manufacturing employment by 1.78%. Finally, the value of the Kleibergen-Paap F-statistic is still around

²These are manufacture of wearing apparel, repair of personal and household goods, manufacture of grain mill products, preparation of textile fibres, and manufacture of refractory ceramic products.

58, suggesting that the Bartik instrument does not suffer from a weak instruments problem.

These results are consistent with the following hypothesis: agglomeration increases the returns to manufacturing employment in the towns. As the manufacturing sector grows in the towns, it attracts manufacturing workers from nearby rural areas; some of these workers relocate their employment to the towns to benefit from agglomeration economies. This could be one reason why we observe lower growth in manufacturing employment in villages located close to faster-growing towns. Because agglomeration forces are stronger for manufacturing industries than for services industries (Ryan, 2012), we observe negative effects of urban growth on the former but not on the latter. The absence of negative effects on rural services employment growth could dampen the effects on rural non-farm employment growth, since more than two-thirds of rural non-farm workers are employed in the services sector.

2.4.3 Effects on rural population outcomes

When the nearest town experiences high employment growth in its manufacturing sector, do rural manufacturing workers migrate out of the village, or do they continue living in the village but start commuting to the nearest town? The answer to this question matters for the extent to which urban growth can lead to prosperity in the neighbouring villages.

There are two challenges to answering this question. First, there are no Indian datasets that track both the workers' residential and work location. The Economic Census records the number of jobs in the village (and town), while the Population Census records the number of people living in the village. The second challenge is that these censuses are collected in different years, therefore it is hard to even make aggregate arguments about the effect of urban employment growth on rural population size.

Despite these limitations, we attempt to answer this question in two ways. First, we

generate the employment–population ratio for the village, using data on the number of non-farm workers in the village from the Economic Census 1998 and 2005, and normalising with village population numbers from 2001. We calculate the change in the employment–population ratio between 1998 and 2005. The first column of Table 2.6 presents the IV estimates for the effect of higher manufacturing employment growth in the nearest town on the change in the rural employment–population ratio. These estimates suggest that urban manufacturing employment growth has no effect on the employment–population ratio of neighbouring villages.

Secondly, we estimate the effect of urban manufacturing employment growth on village population growth between 2001 and 2011. The second column of Table 2.6 presents the IV estimates, suggesting that higher employment growth in the manufacturing sector in the nearest town during 1998–2005 has no statistically significant effect on population growth in the village during 2001–11. Our data unfortunately does not allow us to distinguish between workers in the manufacturing sector and in other non-farm sectors in the Population Census. Nevertheless, these results suggest that workers previously employed in rural manufacturing sectors do not migrate out of the village, but possibly start commuting to the nearest town when jobs grow there due to exogenous demand shocks.

2.4.4 Effects by town size

In this section, we investigate how the effects vary depending on the size of the village’s nearest town. We divide our sample of towns into two halves, based on the town’s population in the 2001 Population Census. The median town population in 2001 was 13,125. Towns with population smaller than this number are “small towns,” and towns with population larger than this number are “large towns.” Table 2.7 presents the IV estimates separately for villages close to “small” and “large” towns in our sample.

For “small” towns, we cannot reject the null that higher manufacturing employ-

ment growth in the nearest town has no effect on rural non-farm or manufacturing employment growth. The Kleinbergen-Papp F-statistic for this regression is also very low (11.55), suggesting that the Bartik instrument may suffer from a weak instruments problem for this sub-sample.

For “large” towns, we cannot reject the null that higher manufacturing employment growth in the nearest town does not affect rural non-farm employment growth. However, we do find statistically significant negative effects of higher manufacturing employment growth in the nearest town on rural manufacturing employment growth. The IV estimate suggests an elasticity of -0.224: 10% higher growth in manufacturing sector employment in the nearest “large” town leads to a 2.24% decline in the growth rate of manufacturing employment in the village. Moreover, the Kleinbergen-Papp F-statistic is very high for this sub-sample (around 75), minimising the risk of a weak instruments problem.

Therefore, the negative rural effects of urban manufacturing employment growth are more pronounced when the nearest town is “large.” This supports our hypothesis that agglomerative forces in towns attract manufacturing workers from nearby villages. Since these agglomerative forces are stronger in large towns (Ryan, 2012), the negative effects on rural manufacturing employment growth are also larger for villages located close to larger towns, as compared to villages located close to smaller towns in our sample.

2.4.5 Effects by village connectivity

In this section, we investigate how the effects vary depending on the village’s connectivity to external markets. We look at two dimensions of connectivity: whether or not the village has a paved road, and how remote or proximate the village is to its nearest town.

We first divide the sample into villages with and without paved roads, according to the 2001 Population Census. Table 2.8 presents IV estimates for the effect of higher ur-

ban manufacturing employment growth on employment growth in the rural non-farm and manufacturing sectors. For both outcomes, we find statistically significant negative effects of urban manufacturing growth for the sample of villages that did not have a paved road in 2001.³ These are precisely the villages that are least suited for manufacturing activities, since they lack good connections to external markets. Hence, when the manufacturing sector grows fast in neighbouring towns, manufacturing workers in these villages are much more likely to migrate out of these villages.

We then divide the sample of villages into four sub-samples based on their distance to the nearest town. The first sub-sample (quartile) includes villages located upto 4.08 km to the nearest town. The second sub-sample includes villages located between 4.08 and 7.5 km to the nearest town. Villages in the third sub-sample are located between 7.5 and 12.33 km to the nearest town, and villages in the fourth sub-sample are located more than 12.33 km from their nearest town (the maximum value is 31.69 km).

Table 2.9 presents the IV estimates. For both rural non-farm and manufacturing employment growth, we find that the effects of higher urban manufacturing growth are restricted to the fourth sub-sample, the villages located more than 12.33 km from their nearest town. Similar to the villages without paved roads, these villages are most remote from their nearest large market, and therefore least suited for manufacturing industries. Growth in nearby towns is therefore much more likely to lead to labour exit from these villages, as compared to villages located in close geographic proximity to their nearest town or city.

2.5 Conclusion

As cities in developing countries reduce their reliance on the rural hinterland, economists will have to start investigating the effects of exogenous growth in urban employment on the economic structure of the surrounding villages. Does urban growth lead to

³The F-statistic is 34.47, suggesting that we do not have a weak instruments problem.

complementary growth in non-farm employment in nearby villages, or does it drive labour exit from the villages? This is important for the design of public policies in developing countries, to enable rural areas to capitalise on the recent growth of developing world cities. However, these questions are difficult to answer. First, towns and villages located in close geographic proximity experience the same employment shocks. Second, economic data is often available for large administrative units, making it difficult to distinguish between changes in urban and rural employment.

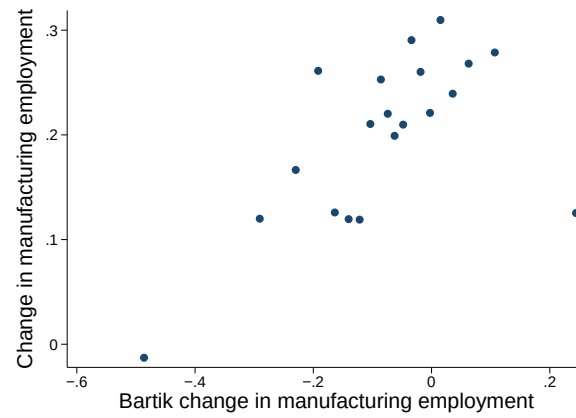
In this chapter, we investigate the effect of growth in manufacturing employment in the nearest town on employment growth rates in different sectors in the village. To isolate exogenous changes in urban employment, we construct a Bartik (shift-share) instrument that exploits employment changes in all *other* towns in the country. Further, we bring newly matched employment data for the universe of Indian towns and villages, which allows us to unpack the geography of urban-rural linkages.

We find that while higher manufacturing employment growth in the nearest town has no statistically significant effect on non-farm employment growth in the village, it does reduce manufacturing employment growth in the village. This result is consistent with the following hypothesis: agglomeration increases the returns to manufacturing activities in urban areas. As the urban manufacturing sector grows, some rural manufacturing workers relocate to the nearest town or city, thus driving labour exit from the villages. Further, we find that these effects are more pronounced for villages located close to larger towns and cities, which is precisely where we expect the agglomeration forces to be the strongest. The effects are also larger for villages that are ill-suited for manufacturing activities in the first place, such as villages without paved roads, and villages located far from their nearest town or city.

While we are able to show the negative effects of urban manufacturing growth for rural manufacturing employment, our data do not allow us to unpack the consequences of this reduced growth for the rural economy. The absence of statistically significant effects on village population growth rates suggest that manufacturing workers

do not migrate out of the village, but start commuting to the nearest town for work. This would allow some gains from urban growth to be absorbed into the rural economy. What this means for rural incomes, assets, and productivity should be the subject of future research.

Figure 2.1
First stage binscatter



Notes: This figure shows the binscatter for the relationship between predicted and actual growth in urban manufacturing employment in our sample, where the predicted growth is constructed using the Bartik shift-share instrument. Each of the 20 bins represents approximately 236 towns in the sample.

Table 2.1
Summary statistics

	N	Mean
Town outcomes		
Population (2001)	4723	33691.0 (98119.2)
Non-farm employment (1998)	4723	4650.9 (13112.4)
Log change in non-farm employment (1998-05)	4723	0.225 (0.695)
Manufacturing employment (1998)	4723	1291.3 (4380.3)
Log change in manufacturing employment (1998-05)	4723	0.199 (0.954)
Village outcomes		
Population (2001)	440,049	1110.3 (1158.5)
Non-farm employment (1998)	440,049	45.47 (138.7)
Log change in non-farm employment (1998-05)	440,049	0.421 (1.676)
Manufacturing employment (1998)	440,049	14.44 (89.86)
Log change in manufacturing employment (1998-05)	440,049	0.185 (1.420)
Services employment (1998)	440,049	31.03 (86.83)
Log change in rural services employment (1998-05)	440,049	0.438 (1.562)
Distance to nearest town	440,049	12.12 (9.462)
Paved Road (2001)	440,049	0.531 (0.499)

Notes: This table presents summary statistics for the towns and villages in our sample. Employment data from 1998 and 2005 comes from the respective Economic Censuses. Population data comes from the Population Census 2001.

Table 2.2
First stage: urban manufacturing employment growth 1998–2005

	Rural nonfarm employment growth	Rural manufacturing employment growth	Rural services employment growth
Bartik change in urban manufacturing employment	0.339 (0.043)***	0.339 (0.043)***	0.339 (0.043)***
Baseline log urban manufacturing employment	-0.264 (0.006)***	-0.264 (0.006)***	-0.264 (0.006)***
Baseline log rural non-farm employment	0.004 (0.001)***		
Baseline log rural manufacturing employment		0.003 (0.001)**	
Baseline log rural services employment			0.004 (0.001)***
Outcome Mean	13.35%	13.35%	13.35%
Fixed effects	District	District	District
Baseline controls	Yes	Yes	Yes
N	443,394	443,394	443,394

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

Notes: This table presents regression estimates from estimating the first-stage equation, Equation 2.2, growth in urban manufacturing employment between 1998 and 2005. Standard errors, clustered at the level of 0.1×0.1 degree grid-cells, are reported in parentheses below the point estimates.

Table 2.3
OLS and IV results for the main outcomes

Panel A. OLS results

	Rural nonfarm employment growth	Rural manufacturing employment growth	Rural services employment growth
Log change in urban manufacturing employment	0.034 (0.006)***	0.039 (0.005)***	0.025 (0.006)***
Baseline log urban manufacturing employment	0.027 (0.003)***	0.021 (0.003)***	0.024 (0.003)***
Baseline log rural non-farm employment	-0.512 (0.003)***		
Baseline log rural manufacturing employment		-0.518 (0.003)***	
Baseline log rural services employment			-0.502 (0.003)***
Outcome Mean	41%	18%	43%
Fixed effects	District	District	District
Baseline employment level controls	Yes	Yes	Yes
Baseline confounding controls	No	No	No
N	443,509	443,509	443,394

Panel B. IV results

	Rural nonfarm employment growth	Rural manufacturing employment growth	Rural services employment growth
Log change in urban manufacturing employment	-0.127 (0.080)	-0.180 (0.069)***	-0.084 (0.074)
Baseline log urban manufacturing employment	-0.016 (0.021)	-0.038 (0.018)**	-0.005 (0.020)
Baseline log rural non-farm employment	-0.511 (0.003)***		
Baseline log rural manufacturing employment		-0.518 (0.003)***	
Baseline log rural services employment			-0.501 (0.003)***
Kleibergen-Paap F statistic	63.03	62.99	63.02
Outcome Mean	41%	18%	43%
Fixed effects	District	District	District
Baseline employment level controls	Yes	Yes	Yes
Baseline confounding controls	No	No	No
N	443,394	443,394	443,394

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

Notes: In this table, Panel A presents regression estimates from estimating the OLS equation, Equation 2.1, and Panel B presents IV estimates using the two stage least squares method, for the main outcome variables: rural nonfarm employment growth, rural manufacturing employment growth, and rural services employment growth between 1998 and 2005. All regressions control for baseline values of the urban and rural employment levels, and include district fixed effects. Standard errors, clustered at the level of 0.1 x 0.1 degree grid-cells, are reported below the point estimates.

Table 2.4
Correlations between industries with high baseline employment shares
and potential confounding variables

	Baseline confounders			
	Percent male	Percent government	Percent SC	Percent ST
Mean value (EC 1998)	87.06%	21.5%	5.17%	2.30%
Top 5 industry categories at baseline:				
Manufacture of wearing apparel	0.407 (0.241)*	0.072 (0.022)**	-0.163 (0.031)***	0.027 (0.045)
Repair of personal & household goods	-0.105 (0.131)	0.056 (0.019)*	0.106 (0.030)***	-0.023 (0.047)
Manufacture of grain mill products	-0.030 (0.151)	-0.022 (0.021)	-0.057 (0.043)	-0.027 (0.032)
Preparation of textile fibre	-0.071 (0.172)	-0.110 (0.023)***	-0.011 (0.041)	-0.004 (0.068)
Manufacture of ceramic products	-0.101 (0.053)*	-0.027 (0.044)	-0.012 (0.020)	-0.053 (0.031)

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

Notes: This table presents estimates of the correlation between the baseline employment industries of the five industries with the highest employment shares, and baseline values of potential confounding variables: percentage of workforce which is male, percentage workforce employed in government companies, percentage workforce that belongs to the Scheduled Castes, and the percentage workforce that belongs to the Scheduled Tribes. All regressions control for baseline values of urban and rural employment, and include district fixed effects. Standard errors, clustered at the level of 0.1×0.1 degree grid-cells, are reported below the point estimates.

Table 2.5
IV estimates for the main outcomes, controlling for potential confounders

	Rural nonfarm employment growth	Rural manufacturing employment growth	Rural services employment growth
Log change in urban manufacturing employment	-0.123 (0.084)	-0.178 (0.072)**	-0.080 (0.078)
Baseline log urban manufacturing employment	-0.016 (0.023)	-0.038 (0.019)*	-0.005 (0.021)
Baseline log rural non-farm employment	-0.512 (0.003)***		
Baseline log rural manufacturing employment		-0.518 (0.003)***	
Baseline log rural services employment			-0.502 (0.003)***
Kleibergen-Paap F statistic	58.26	58.23	58.25
Outcome Mean	41%	18%	43%
Fixed effects	District	District	District
Baseline employment level controls	Yes	Yes	Yes
Baseline confounding controls	Yes	Yes	Yes
N	443,394	443,394	443,394

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

Notes: This table presents IV estimates using the two stage least squares method for the main outcome variables: rural nonfarm employment growth, rural manufacturing employment growth, and rural services employment growth between 1998 and 2005. All regressions control for baseline values of urban and rural employment, and include district fixed effects. They also control for the baseline (1998) values of the potential confounding variables: percentage of workforce which is male, percentage workforce employed in government companies, percentage workforce that belongs to the Scheduled Castes, and the percentage workforce that belongs to the Scheduled Tribes. Standard errors, clustered at the level of 0.1 x 0.1 degree grid-cells, are reported below the point estimates.

Table 2.6
Effect on rural population outcomes

	Emp-Pop Ratio	Pop Growth Rate
Log change in urban manufacturing employment	0.0004 (0.008)	0.016 (0.012)
Baseline log urban manufacturing employment	0.001 (0.002)	0.005 (0.003)
Baseline rural employment–population ratio	-0.512 (0.188)***	
Baseline log rural population		-0.060 (0.001)***
Kleibergen-Paap F statistic	67.30	67.28
Outcome Mean	0.011	15.80%
Fixed effects	District	District
Baseline employment level controls	Yes	Yes
Baseline confounding controls	Yes	Yes
N	430,000	430,000

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

Notes: This table presents IV estimates using the two stage least squares method for the change in the employment–population ratio between 1998 and 2005, and the population growth rate between 2001 and 2011. All regressions control for baseline values of urban manufacturing employment and the relevant rural population outcome, and include district fixed effects. They also control for the baseline (1998) values of the potential confounding variables: percentage of workforce which is male, percentage workforce employed in government companies, percentage workforce that belongs to the Scheduled Castes, and the percentage workforce that belongs to the Scheduled Tribes. Standard errors, clustered at the level of 0.1×0.1 degree grid-cells, are reported below the point estimates.

Table 2.7
Effect on rural employment growth by size of the nearest town

<i>Panel A. Effect on rural non-farm employment growth</i>		
	Small towns	Large towns
Log change in urban manufacturing employment	-0.090 (0.187)	-0.105 (0.085)
Baseline log urban manufacturing employment	-0.046 (0.087)	0.025 (0.020)
Baseline log rural non-farm employment	-0.518 (0.004)***	-0.513 (0.004)***
Kleibergen-Paap F statistic	11.54	75.28
Outcome Mean	39.51%	42.38%
Fixed effects	District	District
Baseline employment level controls	Yes	Yes
Baseline confounding controls	Yes	Yes
N	216,573	226,821

<i>Panel B. Effect on rural manufacturing employment growth</i>		
	Small towns	Large towns
Log change in urban manufacturing employment	-0.031 (0.154)	-0.224 (0.072)***
Baseline log urban manufacturing employment	-0.014 (0.072)	-0.020 (0.017)
Baseline log rural manufacturing employment	-0.526 (0.004)***	-0.518 (0.004)***
Kleibergen-Paap F statistic	11.55	75.19
Outcome Mean	16.53%	19.47%
Fixed effects	District	District
Baseline employment level controls	Yes	Yes
Baseline confounding controls	Yes	Yes
N	216,573	226,821

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

Notes: This table presents IV estimates using the two stage least squares method, for rural non-farm and rural manufacturing employment growth between 1998 and 2005, based on the size of the nearest town. "Small towns" are towns with 2001 population less than the median town population in our sample, and "large towns" are towns with 2001 population greater than the median town population in our sample. Median town population in 2001 is 13,125. All regressions control for baseline values of urban and rural employment, and include district fixed effects. They also control for the baseline (1998) values of the potential confounding variables: percentage of workforce which is male, percentage workforce employed in government companies, percentage workforce that belongs to the Scheduled Castes, and the percentage workforce that belongs to the Scheduled Tribes. Standard errors, clustered at the level of 0.1×0.1 degree grid-cells, are reported below the point estimates.

Table 2.8
Effect on rural employment growth by access to paved roads

Panel A. Effect on rural non-farm employment growth

	Without roads	With roads
Log change in urban manufacturing employment	-0.216 (0.111)*	-0.059 (0.087)
Baseline log urban manufacturing employment	-0.046 (0.031)	-0.004 (0.022)
Baseline log rural non-farm employment	-0.599 (0.004)***	-0.549 (0.003)***
Kleibergen-Paap F statistic	34.47	65.56
Outcome Mean	36.18%	47.27%
Fixed effects	District	District
Baseline employment level controls	Yes	Yes
Baseline confounding controls	Yes	Yes
N	199,907	225,673

Panel B. Effect on rural manufacturing employment growth

	Without roads	With roads
Log change in urban manufacturing employment	-0.232 (0.094)**	-0.108 (0.079)
Baseline log rural manufacturing employment	-0.613 (0.004)***	-0.519 (0.003)***
Baseline log urban manufacturing employment	-0.058 (0.026)**	-0.022 (0.020)
Kleibergen-Paap F statistic	34.47	65.51
Outcome Mean	12.70%	23.59%
Fixed effects	District	District
Baseline employment level controls	Yes	Yes
Baseline confounding controls	Yes	Yes
N	199,907	225,673

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

Notes: This table presents IV estimates using the two stage least squares method, for rural non-farm and rural manufacturing employment growth between 1998 and 2005, for villages with and without access to paved roads, as per the 2001 Population Census. All regressions control for baseline values of urban and rural employment, and include district fixed effects. They also control for the baseline (1998) values of the potential confounding variables: percentage of workforce which is male, percentage workforce employed in government companies, percentage workforce that belongs to the Scheduled Castes, and the percentage workforce that belongs to the Scheduled Tribes. Standard errors, clustered at the level of 0.1×0.1 degree grid-cells, are reported below the point estimates.

Table 2.9
Effect on rural employment growth by distance to the nearest town

Panel A. Effect on rural non-farm employment growth

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Log change in urban manufacturing employment	-0.007 (0.152)	-0.139 (0.130)	-0.147 (0.130)	-0.399 (0.192)**
Baseline log urban manufacturing employment	0.005 (0.044)	-0.020 (0.033)	-0.010 (0.032)	-0.092 (0.054)*
Baseline log rural non-farm employment	-0.558 (0.005)***	-0.523 (0.004)***	-0.502 (0.005)***	-0.478 (0.005)***
Kleibergen-Paap F statistic	16.70	36.95	35.06	16.32
Outcome Mean	49%	42.12%	36.77%	35.89%
Fixed effects	District	District	District	District
Baseline employment level controls	Yes	Yes	Yes	Yes
Baseline confounding controls	Yes	Yes	Yes	Yes
N	110,343	110,768	110,956	111,327

Panel B. Effect on rural manufacturing employment growth

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Log change in urban manufacturing employment	-0.031 (0.143)	-0.181 (0.109)*	-0.161 (0.110)	-0.468 (0.171)***
Baseline log urban manufacturing employment	0.001 (0.041)	-0.040 (0.028)	-0.028 (0.027)	-0.117 (0.048)**
Baseline log rural manufacturing employment	-0.553 (0.005)***	-0.525 (0.004)***	-0.508 (0.005)***	-0.495 (0.005)***
Kleibergen-Paap F statistic	16.68	36.93	35.02	16.33
Outcome Mean	22.43%	18.13%	15.87%	15.57%
Fixed effects	District	District	District	District
Baseline employment level controls	Yes	Yes	Yes	Yes
Baseline confounding controls	Yes	Yes	Yes	Yes
N	110,343	110,768	110,956	111,327

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

Notes: This table presents IV estimates using the two stage least squares method, for rural non-farm and rural manufacturing employment growth between 1998 and 2005, based on village's distance to the nearest town. Quartile 1 is 0–4.08 km, Quartile 2 is 4.08–7.50 km, Quartile 3 is 7.50–12.33 km, and Quartile 4 is 12.33–31.69 km. All regressions control for baseline values of urban and rural employment, and include district fixed effects. They also control for the baseline (1998) values of the potential confounding variables: percentage of workforce which is male, percentage workforce employed in government companies, percentage workforce that belongs to the Scheduled Castes, and the percentage workforce that belongs to the Scheduled Tribes. Standard errors, clustered at the level of 0.1×0.1 degree grid-cells, are reported below the point estimates.

CHAPTER 3

PROPERTY TITLES AND HOUSING INVESTMENT AMONG THE URBAN POOR

3.1 Introduction

Assigning well-defined property rights has long been considered an important ingredient in recipes for economic growth and prosperity (Smith, 1776; North, 1981). In recent decades, the idea of giving property rights to the urban poor in the form of formal titles has gained traction among economists and policymakers (de Soto, 2000; Durand-Lasserve et al., 2007).¹ By reducing eviction risk, expanding opportunities for trade, and providing collateral, formal property titles are expected to encourage investment by the urban poor and thereby improve their economic outcomes (Baharoglu, 2008; Besley and Ghatak, 2010). While the early empirical literature supported these views, at least in the context of Latin America (Field, 2005; Field, 2007; Di Tella et al., 2007; Galiani and Schargrodsky, 2010),² recent papers have cast doubts on whether these results generalise in Latin America (Galiani and Schargrodsky, 2016; Gutierrez and Molina, 2016; Molina et al., 2018). In other parts of the world, with lower access to the formal financial sector and higher perceived tenure security, it is possible that prop-

This version August 2018. I thank Micro Home Solutions for making their household survey data available for this research. I am grateful to Amrit Amirapu, Cheryl Doss, Paolo Falco, James Fenske, Doug Gollin, Beata Javorcik, Francis Teal, Andrew Zeitlin, and two anonymous Oxford examiners for their helpful feedback on this chapter. I am also grateful to participants at the CSAE Conference 2018, the Oxford CSAE Workshop, and the QEH Informal Development Studies Workshop for their comments and suggestions. All remaining errors are my own.

¹Several countries have launched urban land titling programs in recent years, including Angola, Egypt, Ghana, Malawi, Senegal, South Africa, Turkey, Afghanistan, Cambodia, India, Indonesia, Laos, Philippines, Argentina, Brazil, Bolivia, Colombia, Ecuador, Mexico, Honduras, Paraguay, and Peru. See Durand-Lasserve et al. (2007) for more details.

²See Deininger and Binswanger (1999) for a review of an even earlier empirical literature.

erty titles may not have significant effects on household outcomes. But this remains an open empirical question.

In this chapter, I investigate the effect of possessing formal property titles on housing investment decisions by the urban poor in Delhi. I study two household decisions in particular: whether it has invested in improving the quality of its current residential structure over the preceding five years, and the amount it has invested in housing improvements during the same period. I also study the relationship between property titles and the quality of the households' current residential structure. I focus on housing because it remains one of the most valuable physical assets for the urban poor (Alston et al., 1999), and because investment in improving the quality of housing is a common reason for borrowing and a major motivation for household savings by the urban poor (Mehra et al., 2012).

To examine this question, I use cross-sectional data from 1,376 households in three resettlement localities in Delhi. Variation in property title status arises from the process through which these localities were settled. The government evicted households living in centrally located slums in Delhi and "resettled" them in these localities. The resettled households were given formal property titles, which provided legal protections against eviction. Further, these titles could be used as collateral for obtaining credit from formal financial institutions. But the titles did not confer any transfer rights (Government of National Capital Territory of Delhi, 2006; Dupont, 2008; Bhan, 2009).³ Nevertheless, some resettled households sold their properties through unregistered transactions to new buyers (Bhan and Shivanand, 2013). The households that purchased these properties did not receive any legal documents to support their claims to the property, and hence do not have formal property titles. The empirical design in this chapter compares households that were officially resettled and continue to live in these localities, with households that came in later through unregistered transactions.

³Formal sector lenders (banks and microfinance companies) are willing to accept the title as collateral, even though they cannot repossess the property. Lenders believed that because they retain physical possession of the document while the loan is outstanding, it exerts a psychic cost on the borrowers (Mehra et al., 2012).

I find that possessing formal property titles is not correlated with the decision to invest in housing improvements or with current housing quality. While a simple regression of the investment decision on property title status yields a “naive” average treatment effect of 22.3 percentage points, this result seems to be driven by the construction of new houses in Savda Ghevra, which was established less than five years ago. This construction activity is not really influenced by property title status; rather, it represents new households moving in and building on the plot of land allocated to them by the government. Restricting the sample to households that have been in their current house for five years or longer, I find that property titles are no longer correlated with the decision to invest in housing improvements. I find similar null results for the correlation between property titles and the amount of investment in housing improvements during the preceding five years.

However, in the absence of random assignment of property titles, identifying their causal effect is challenging. Households with formal titles are arguably different in several ways from households that do not possess these titles. For example, titled households are the ones that decided not to leave the resettlement localities after being settled there by the government. Further, non-titled households had to purchase their property, whereas the titled households received a plot of land from the government. Therefore, a simple comparison between the investment behaviour or housing quality of titled and non-titled households will suffer from selection bias.

I address these challenges in several ways. First, assuming selection on observed variables, I use linear regression analysis to control for observed household characteristics that could affect selection into treatment. In this analysis, I employ locality fixed effects, since the three localities were established at different points of time and have different levels of access to public infrastructure. Further, since 81% of the households have migrated to Delhi, I use fixed effects for the households’ home state. These fixed effects allow me to control for differences in language, culture, proximity to home state, and in some cases, occupation choices and informal credit networks (Alpa et

al., 2018). Moreover, I control for several observed household characteristics that are not potential outcomes of property titles status. Following Field (2005), I control for the size of the household at the time of moving to the current house, as well as for age, age squared, and gender of the household head. For the subset of migrant households, I additionally control for the number of years the household had been in Delhi when it moved to its current location, as well as its reason for migrating to Delhi. Even after controlling for these observed variables, I continue to find that property titles are not correlated with the households' decisions to invest in improving the quality of their residential structure.

It is possible that the benefits from property titles are offset by features that make non-titled households more likely to invest. For example, since non-titled households moved to their current location more recently than titled households, they may have a greater need to invest in improving the quality of their residential structure.⁴ To test this, I control for the number of years the household has lived in its current house, but the results do not change. In any case, this concern is partially addressed by the fact that there are no statistically significant differences in the housing quality of titled and non-titled households.

Another possibility is that non-titled households are richer than titled households; after all, the former purchased their current house, as opposed to receiving it from the government during the resettlement process. If that were the case, the additional financial resources could compensate for the lack of titles. However, I find no statistically significant correlation between property title status and household consumption expenditure or assets.⁵

In addition to using linear regression analysis, I employ propensity score matching to address selection on observed variables. This technique essentially constructs a set of non-titled households that is as similar as possible to the set of titled households in

⁴Such a situation could arise if the houses coming on the market were more dilapidated than the average house in the area.

⁵I cannot control for consumption or assets in the main regression as these are potential outcomes of property titles.

terms of the observed variables, to ensure that the estimates are not biased by selection into (or out of) property title status. The matching estimator also shows that there is no statistically significant effect of property title status on households' investment decisions or the quality of their residential structure.

I further consider the possibility of selection on unobserved variables. While I cannot, by definition, observe these variables, I can infer their importance from the relationship between the treatment and the observed variables, and by making assumptions regarding the shared covariance properties of the observed and unobserved variables and the property titles treatment (Altonji et al., 2005; Oster, 2017). In particular, I assess the extent to which the degree of selection on unobserved variables will have to differ from the degree of selection on observed variables, for my estimates of a null effect to mask a true non-zero effect. I find that the relative degree of selection on unobserved variables has to be "very high" (twice or more) for my null effects to mask a true non-zero effect of property title status on the decision to invest in housing improvements.

Finally, it is possible that the study is under-powered to detect the treatment effect of property titles on the outcomes of interest. I use power analysis to calculate the minimum effect that can be detected given the sample size and the variability in the outcome variables. For example, for the decision to invest, I find that the minimum effect that can be detected with a 95% confidence interval and power of 0.8 is a 10.2 percentage point (45.33%) increase in the probability of investing in housing improvements. Since I find no statistically significant effect, I can conclude that the true effect of property titles on the investment decision is unlikely to be greater than 10.2 percentage points (45.33%). As a comparison, in Peru, formal property titles were associated with a 68% increase in the rate of housing improvements (or "renovations") (Field, 2005).

To understand why property titles do not affect investment behaviour in this context, I turn to the theoretical arguments expounded in Besley (1995) and Besley and

Ghatak (2010). Property titles can affect investment through three channels: by expanding the opportunities for trade in property markets, by reducing the risk that the property would be expropriated, and by providing collateral that can be used to access credit from the formal financial sector. The formal property titles that I study here did not confer transfer rights to their possessors (Bhan, 2009; Bhan and Shivanand, 2013), shutting down the first channel. Secondly, in India, the difference in perceived eviction risk between titled and non-titled households is smaller than the difference in eviction risk in Latin America (PRIndex Analytical Report, 2017). Eviction also seems to be politically infeasible (Mehra et al., 2012), thereby closing the second channel. Finally, I show that a very small proportion of the sample households have access to the formal financial sector, and that property title status is not positively correlated with access to formal or informal sources of lending. In any case, most households depend on their own stock of savings to finance their investment in improving housing quality. Therefore, titled households are not able to reap the benefits of using their titles as collateral. These factors help to explain why we do not observe any effects of property title status on investment outcomes.

This chapter contributes to several strands of literature. Most importantly, it contributes to the literature documenting the effects of more secure property titles on a range of economic outcomes, such as investment in improving the quality of agricultural land (Besley, 1995), fertility (Field, 2003), housing investment (Field, 2005; Collin et al., 2015), labour supply (Field, 2007), credit access (Field and Torero, 2006), and children's education (Galiani and Schargrodsky, 2010). Most of these papers have exploited quasi-natural or natural experiments and estimated significant positive effects of formal property titles.

However, these results may not generalise. Recent evidence suggests that property titles have small or null effects on investment among the poorest households (Molina et al., 2018), and high costs of registering property transactions can force several households to allow their plots to revert to informality in the long-run (Galiani and Schar-

grotsky, 2016; Gutierrez and Molina, 2016). The results in this chapter support this view. I investigate the impact of property titles 10-20 years after they were awarded, and find that they have no long-run effect on housing investment or housing quality. The sample of households in this chapter consists of low-income households with severely constrained access to formal sector lending, and I show that for these households property titles do not seem to matter.

Relatedly, this chapter also contributes to a growing literature on the economics of urban poverty (Marx et al., 2013) and informality in the developing world (La Porta and Shleifer, 2014). One of the key questions examined by this literature is how developing countries can address the challenges posed by slums and provide adequate housing to the urban poor (Glaeser et al., 2015; Brueckner and Lall, 2015; Barnhardt et al., 2017). The resettlement localities studied in this chapter were part of the government's efforts at clearing slums from central Delhi. By studying the effects of property titles in this setting, this chapter generates insights on how governments should manage the resettlement process for the benefit of the urban poor.

The rest of the chapter is organised as follows. In Section 3.2, I provide background on the resettlement localities and the nature of property titles, followed by a description of the data in Section 3.3. Section 3.4 explains the empirical strategy, Section 3.5 presents the main results, and Section 3.6 discusses the mechanisms. Finally, Section 3.7 concludes.

3.2 Context

In this section, I discuss how Delhi's resettlement localities were established, and the nature of the property titles that were given to the resettled households.

In 1956, the Delhi government passed the Slum Clearance Act to provide a legal framework for clearing squatters and slums from public land located in the central parts of the city. This act provided for settling the slum households in "resettlement

localities,” which were established on the outskirts of the city. The three resettlement localities studied here were established at different points of time: Mangolpuri was established in 1976, Narela in 1996, and Savda Ghevra in 2005-06 (Dupont, 2008; Mehra et al., 2012; Bhan and Shivanand, 2013). When established, each place was located at the fringe of the main city. Since 1976, however, the main city has expanded outwards and Mangolpuri is no longer a fringe location. Narela and Savda Ghevra are still distant from the main city. The geographic location of the three areas is shown in Figure 3.1. Because of these differences in location relative to the main city, and differences in the availability of public infrastructure, it is important to account for locality fixed effects in my empirical analysis.

The households were given a form of legal title, called ‘possession slip,’ over the plots they were allocated in the new localities. While the possession slips insured the household against eviction or expropriation and could be used as collateral, they did not confer transfer rights to the holders. Nevertheless, a proportion of the relocated households sold their plots through a ‘power of attorney’ (Bhan, 2009). While this type of document is commonly used for property transactions in India, it has to be registered in a court of law to be legally valid. In this case, since the underlying transaction violated the terms of the possession slips, the power of attorney documents could not be registered and hence did not constitute legally valid titles. In what follows, I define possession slips to be formal property titles, and I will compare housing investment decisions made by households that have possession slips with decisions made by households holding an unregistered power of attorney.

The distinction between possession slip and power of attorney is important because banks, microfinance companies, and other financial institutions accept only the possession slips as collateral. While the loan is outstanding, the financial institution retains physical ownership of the document. This exerts a psychic pressure on the borrower to repay the loan and reclaim the document, which is often the borrower’s most valuable asset. Moreover, even though the financial institution cannot repossess

the plot in the event of a default, it can initiate legal proceedings against the borrower. Financial institutions do not accept the unregistered power of attorney document as collateral, thereby denying households that hold such documents access to formal sources of credit (Mehra et al., 2012).

Therefore, the variation in property title status comes from two distinct processes – the Government’s policy to assign non-transferrable property titles only once, and the decision of some households to purchase *de facto* ownership through unregistered property transactions. I will exploit this variation to investigate how more secure property titles are related to households’ investment decisions and housing quality.

3.3 Data

I use data from a cross-sectional survey of 1,376 households located in three resettlement localities in Delhi: Mangolpuri, Narela, and Savda Ghevra. It was collected as part of a Michael and Susan Dell Foundation (MSDF) study on the housing finance market in India. In this section, I discuss the main variables used in the analysis and how they have been constructed. Table 3.1 provides summary statistics for the main variables by locality.

3.3.1 Treatment variable

The treatment variable, property title status, is defined as a binary variable that takes the value 1 if the household has a possession slip and 0 otherwise. This is because the possession slip insures the household against eviction and is accepted as collateral by formal financial institutions. Overall, 84.2% households possess property titles. Table 3.1 shows that this varies from 56.1% in Mangolpuri to 95-97% in Savda Ghevra and Narela. To ensure that my results are not driven by the fact that a large proportion of households in Narela and Savda Ghevra have property titles, I will show that all results are robust to restricting the sample to Mangolpuri households.

3.3.2 Outcomes

I study the effect of property titles on three outcomes: whether or not the household has invested in home improvements during the preceding five years, the amount of investment in home improvements during the preceding five years, and its current housing quality.

The first outcome, whether or not the household has invested in home improvements during the preceding five years, is a binary variable. In the questionnaire, the households are asked, “Have you undertaken any construction in your house in the past 5 years?” As shown in Table 3.1, 41.2% of the households in the sample answered in the affirmative. This average ranges from about 16.6% in Mangolpuri and Narela, to almost 88% in Savda Ghevra. The large value for Savda Ghevra is not surprising since it was established most recently among the sample resettlement localities, only 5-6 years prior to the survey. The average household in Savda Ghevra has lived in its current house for just over four years; 93% of the Savda Ghevra households moved to their current house within the last 5 years. Therefore, it is possible that the high affirmative response to the housing investment question reflect construction of the house itself, as opposed to investment in improving the quality of the existing residential structure, which is what I am interested in. To ensure that this is not driving the results, in Appendix 3.A I show that the results are robust to restricting the sample to households that have been in their current house for five years or longer.

The second outcome is the amount of investment made by the household in improving its housing quality in the preceding five years. As Table 3.1 shows, among the set of households that have invested, the average value of the investment is approximately INR 100,000 (USD 2,240),⁶ and the variation across localities is not substantial. This value is approximately 17 times the average monthly household consumption expenditure.

To ensure that the results are not driven by outliers, I will use the Inverse Hyper-

⁶Based on the exchange rate of INR 44.6351 per 1 USD prevailing on 1 October 2010.

bolic Sine (IHS) transformation of the investment amount (Bellemare et al., 2013).⁷ The IHS transformation is useful since for almost 60% of the sample, housing investment in the preceding five years is zero. Using log investment would require adding an arbitrarily small number, say 1, to the investment value, which is not satisfactory. Therefore, while I show robustness of the results to using log investment in Appendix 3.B, the preferred specification is to use the IHS value of housing investment.

The third outcome is an index measuring the quality of the current residential structure. The dataset provides six measures of housing quality: (1) whether the house has a separate room that serves as a kitchen; (2) whether the roof is solid (made from cement, slab, or bricks); (3) whether the walls are solid (made from bricks or cement); (4) whether the floor is solid (made from cement, tile, or marble); (5) whether the house has its own toilet; and (6) whether the house receives piped water. All measures reflect the result of the household's private investment in improving the quality of its residential structure. Descriptive statistics for these indicators are presented in Table 3.2.

I aggregate these measures in two ways. The first is the aggregate of the standardised value of each housing quality indicator. This index is simple and intuitive, but places an equal weight on each indicator. The second index aggregates the values using the procedure developed in Anderson (2008). This procedure weights each indicator by the inverse of the covariance matrix, thus placing a greater weight on indicators that are less correlated with the other indicators. Both these measures of housing quality are correlated and give me the same results. Table 3.2 shows the values of these indices for all three localities. It is worth noting that Mangolpuri has the highest average values of the housing quality indices.

⁷The IHS transformation of x , formally denoted $\text{arsinh}(x)$, is defined such that: $IHS(x) = \ln(x + \sqrt{x^2 + 1})$

3.3.3 Household characteristics

In my analysis, I control for several observed household characteristics that are not potentially driven by the households' property titles status. Following Field (2005), I control for the size of the household at the time of moving to the current residence, as well as for age, age squared, and gender of the household head. I do not control for the current household size as fertility may be affected by property title status, as demonstrated in Field (2003). I control for age and gender of the household head because both these factors can affect the household's ability to raise resources to finance housing investment. For the subset of migrant households, I also control for the number of years the household had been in Delhi when it moved to its current residence, as well as the stated reason for migrating to Delhi. Descriptive statistics for these variables are presented in Table 3.1.

3.3.4 Financial information

I use two aspects of the household's financial position in my analysis. The first is monthly household consumption expenditure. I have data on the households' monthly expenses for food, education, electricity, gas and other utilities, entertainment, mobile phone bills, rent, and interest expenses, and annual expenditure on household goods, medical expenses, travel, and social functions. I convert the annual expenses to monthly estimates, and aggregate the monthly expenses to calculate the household's monthly consumption expenditure. As Table 3.1 shows, households have an average monthly consumption expenditure of INR 5807 (USD 130). Households in Mangolpuri have higher consumption expenditures than households in Narela, and Savda Ghevra households have the lowest consumption expenditures. In the analysis, I will use log of the monthly consumption expenditure.⁸

The second proxy for the households' financial position is an asset index. I have

⁸The IHS transformation is not required since no household has zero monthly consumption expenditure.

data on whether the household owns a bicycle, a motorcycle, a car, a television, a washing machine, livestock, and jewellery. Similar to the housing quality index, I aggregate this data in two ways: first, a simple aggregate of the normalised values of the six asset indicators, which weights possession of these assets equally; and second, aggregating the six asset indicators using the procedure developed by Anderson (2008), which weights the assets using the inverse of the covariance matrix.

The descriptive statistics for the individual assets and the asset indicators are shown in Table 3.3. Similar to the average monthly consumption expenditure, Mangolpuri households have the highest values for the asset indices. Most Mangolpuri households own a television, motorcycles are also common, and 1.7% households in Mangolpuri even own a car. These assets are much less common in Narela and Savda Ghevra, where a larger proportion of the households relies on bicycles.

3.4 Empirical Strategy

Evaluating the effect of property titles on housing investment decisions and housing quality is challenging. In the absence of randomization, property title status may be correlated with observed and unobserved variables that also affect these outcomes. To address this issue, I begin by assuming that the set of titled and non-titled households differ only based on observed variables (selection on observables). I deal with selection on observables in two ways: linear regression, described in Section 3.5.1, and matching techniques, described in Section 3.5.2. Then, in Section 3.5.3, I consider selection on unobserved variables. In particular, I estimate the ratio of selection on unobserved variables to selection on observed variables that would be required to attribute the entire effect of property titles to selection bias. Finally, in Section 3.5.4, I evaluate whether the study is adequately powered to detect meaningful treatment effects.

3.4.1 Linear regression

For the uncensored outcomes – whether the household has engaged in housing investment in the preceding five years, and the housing quality index – I estimate the following OLS regression equation:

$$y_{i,l,s} = \beta_1 \text{PropertyTitle}_{i,l,s} + \beta_2 \text{HouseChars}_{i,l,s} + \beta_3 \text{Migration}_{i,l,s} + \delta_l + \eta_s + \epsilon_{i,l,s} \quad (3.1)$$

where $y_{i,l,s}$ is the outcome of interest for household i in locality l ; s is its home state. $\text{PropertyTitle}_{i,l,s}$ is the treatment variable, which takes the value 1 if the household has formal title for its residence, and 0 otherwise. $\text{HouseChars}_{i,l,s}$ are household characteristics that are either pre-determined or not potential outcomes of property title status. $\text{Migration}_{i,l,s}$ includes migration-related variables. δ_l is the locality fixed effect, and η_s is the home-state fixed effect. Finally, $\epsilon_{i,l,s}$ is the error term in the regression equation. I use heteroskedasticity-robust standard errors throughout the analysis.

I control for household characteristics that are pre-determined or not potential outcomes of property title status. I follow Field (2005) in choosing these household characteristics: size of the household at the time of moving to the current residence, as well as age, age squared, and gender of the household head. For the subset of migrant households, I also control for the number of years the household had been in Delhi when it moved to its current residence, as well as its reason for migrating to Delhi. For 48% of the households, financial distress compelled them to migrate to Delhi, while for 47% of the households, the reason was economic opportunity (“jobs”). The remaining households migrated for reasons related to education or family conflict.

I include locality fixed effects because the three localities in my sample were established at different points of time and have different levels of access to public infrastructure. The home-state fixed effects allow me to control for differences in language,

culture, proximity to the home state, and possibly also differences in occupation choice and informal credit networks.

However, the OLS framework is not appropriate for studying the relationship between property title status and the amount of investment in the preceding five years. Since 59% of the sample households have not invested in improving the quality of housing in the preceding five years, we have a corner solution for the amount of investment. The OLS regression does not take this into account, which can yield incorrect coefficients. Therefore, to investigate this relationship, I estimate the following Tobit regression model (Tobin, 1958):

$$\begin{aligned} \operatorname{arsinh}(\operatorname{Investment}_{i,l,s}) = & \beta_1 \operatorname{PropertyTitle}_{i,l,s} + \beta_2 \operatorname{HouseChars}_{i,l,s} \\ & + \beta_3 \operatorname{Migration}_{i,l,s} + \delta_l + \eta_s + \epsilon_{i,l,s} \end{aligned} \quad (3.2)$$

where $\operatorname{arsinh}(\operatorname{Investment}_{i,l,s})$ is the Inverse Hyperbolic Sine (IHS) transformation of the amount of housing investment in the preceding five years.⁹ The other variables have the same definition as in Equation 3.1.

In addition to the observed variables included in $\operatorname{HouseChars}_{i,l,s}$ and $\operatorname{Migration}_{i,l,s}$, there may be other variables that are correlated with both property title status and the outcome variables. One among these is the duration for which the household has lived in its current residence. This is positively correlated with property title status, since the titled households were the first households to settle in any given resettlement locality; the non-titled households entered later. It may also be correlated with investment decisions, for example if the houses that the non-titled households purchased were in worse condition than the houses of the titled households, who chose not to sell. Therefore, I add length of residence in the current house as an additional control in Equations 3.1 and 3.2 and test whether it affects the results.

⁹In Appendix 3.B, I show that the results are robust to using $\log(\operatorname{Investment}_{i,l,s})$ instead of the IHS transformation.

A related hypothesis is that the observed value of investment actually represents construction of the house itself, rather than investment in improving its quality. This may be especially true for households that have lived in their current house for fewer than five years. To address this challenge, I restrict the sample to all households that have lived in their current house for five years or longer.

Another variable that could be correlated with both property title status and the outcome variables is the households' financial position. For example, since non-titled households have purchased their property, they may be financially stronger than titled households, who received it from the government. This negative correlation between property title status and financial position could offset any theoretical benefits from the titles. I have data on two proxies for household financial position: monthly household consumption expenditure and an index of household assets. To test the correlation between property titles and financial position, I estimate Equation 3.1 using these two proxies as outcome variables.

3.4.2 Matching estimators

In addition to regression analysis, I use propensity score matching to address selection on observed variables (Rosenbaum and Rubin, 1983). This technique deals with selection bias by restricting the comparison of investment behaviour within selected households with similar exogenous characteristics. In other words, it compares titled households to non-titled households that have a similar "propensity" to possess formal titles, based on the observed characteristics.

I first estimate a logit regression of property title status on observed household characteristics. I use the following set of matching variables: (1) size of household at the time of moving to the current location; (2) age of the household head; (3) age squared of the household head; and (4) gender of the household head. This yields the predicted probability – the propensity score – that a given household possesses a formal property title. In the second stage, I use an Epanechnikov kernel function

to construct a set of non-titled households that has the closest propensity score to the titled households. I can impose the additional requirement that the matched households belong to the same locality and home-state “cell.” I implement the procedure both ways: by imposing exact matching within locality and home-state cell, and not imposing it.

In all matching estimations, I use bootstrap with 50 replications to obtain the standard errors. Further, following the bias-correction method proposed by Abadie and Imbens (2011), I account for regression adjustment with respect to the matching variables after matching. I implement the matching procedure in Stata 15 using the ‘kmatch’ command (Jann, 2017).

For the subset of migrant households, I have additional pre-treatment characteristics: the reason why they migrated to Delhi, and the number of years they had been in Delhi when they moved to their current location. I implement the matching procedure separately for the subset of migrant households. In addition to the matching variables mentioned in the previous paragraph, I also match on the number of years the household had been in Delhi when it moved to its current location. Moreover, now I can impose the additional requirement that the matched households come from the same locality, home-state and “reason for migration” cell. Once again, I implement the procedure both ways: by imposing exact matching within these cells, and not imposing it.

3.4.3 Selection on unobserved variables

In previous sections, I have described my empirical strategy for addressing selection on the observed variables. Without random assignment of property titles to the households, it is more challenging to address selection on unobserved variables, since by definition they are not observed. However, I can make inferences based on the relationship between the treatment and the observed variables (Altonji et al., 2005), and by exploiting the shared covariance properties between the observed and unobserved

variables and the treatment (Oster, 2017). In particular, I assess the extent to which the degree of selection on unobserved variables will have to differ from the degree of selection on observed variables, for my estimates of a null effect to mask a true non-zero effect of property titles on the key outcomes of interest. I implement this procedure in Stata 15 using the ‘psacalc’ command (Oster, 2013).

3.4.4 Power analysis

Finally, after considering selection on observed and unobserved variables, I consider the possibility that my analysis is under-powered to detect meaningful treatment effects. To put another way, I calculate the minimum effect for each outcome that can be detected using a 95% confidence interval and power equal to 0.8, given the size of the treatment and control groups and the standard deviation of the outcome variable. I implement this in Stata using the ‘power twomeans’ command.

3.5 Results

Property titles seem to have no statistically significant effect on housing investment decisions or housing quality in Delhi’s resettlement localities. I show this by first assuming selection on observed variables and using linear regression analysis to study the relationship between property titles and investment decisions (Section 3.5.1). Continuing with the selection on observed variables assumption, I then discuss the results obtained from implementing propensity score matching in Section 3.5.2, where I also discuss the balance tests for the matching variables and the density distributions. Next, in Section 3.5.3, I estimate the effect of property title status on the decision to invest in housing improvements assuming different values for the relative degree of selection on unobserved variables, relative to the degree of selection on observed variables. Finally, in Section 3.5.4, for each outcome variable, I present the minimum effect size that could have been detected given the sample size and the standard deviation

of the outcome of interest.

3.5.1 Linear regression

I first test whether property titles are associated with a higher probability of investing in housing improvements during the preceding five years. Table 3.4 presents regression coefficients from estimating Equation 3.1 for the decision to invest in housing. The outcome variable takes the value 1 if the households has invested in improving the quality of its housing in the preceding five years, and 0 otherwise. Column 1 of Panel A presents the “naive” average treatment effect – a simple comparison of the outcome mean for the titled and non-titled households. The coefficient is large and statistically significant. Titled households are 22.3 percentage points more likely to have invested in housing in the preceding five years. This is more than 50% of the sample mean. However, much of this comes from variation across localities. Adding a locality fixed effect in Column 2 of Panel A reduces the coefficient to 1.6 percentage points, and I can no longer reject the null that the coefficient is equal to 0, even at a 90% confidence interval.¹⁰ In Column 3, I also include fixed effects for the households’ home state, and the result does not change. Similarly, allowing for selection on observed household characteristics by controlling for these characteristics (in Column 4) and controlling for households’ migration histories (in Column 5) does not change the results. Finally, I present the regression coefficients for the subsample of Mangolpuri households in Panel B of Table 3.4. The results are robust to restricting the sample to Mangolpuri.

Then, I test whether property titles are associated with measures of housing quality, such as the normalised and the “Anderson” versions of the housing quality index (HQI). The construction of these indices has been explained in Section 3.3.4. Tables 3.5

¹⁰In fact, the result in Column 1 is driven by households who have moved to their current residence within the last 5 years (most of them in Savda Ghevra), and therefore may reflect new construction rather than home improvements. We can see this from Column 1 of Table 3.A1, where I have restricted the sample to households who have been in their current house for 5 years or longer.

and 3.6 present regression coefficients from estimating Equation 3.1 for the normalised and “Anderson” versions of the HQI respectively. Column 1 of Panel A presents the “naive” average treatment effect – a simple comparison of the HQI mean for titled and non-titled households. The HQI is negatively correlated with property title status: the HQI values for households with titles are about 1 standard deviation lower than the HQI values for non-titled households.¹¹ However, adding locality fixed effects (Column 2, Panel A) reduces the magnitude of the coefficient and it is no longer statistically significantly different from zero. This null result continues to hold when I add home-state fixed effects (Column 3, Panel A), controls for household characteristics (Column 4, Panel A), and controls for the households’ migration histories (Column 5, Panel A). Therefore, assuming selection on observed variables, I can conclude that HQI values for titled and non-titled households are not (statistically) significantly different.

Even if property titles are not correlated with the probability of investing in housing improvements or with housing quality, they may still affect the amount of investment. This can be the case, for example, if households can use the formal titles as collateral and borrow from formal financial sector lenders, such as banks and micro-finance companies. To test this, I estimate a Tobit regression (Equation 3.2) of the amount of investment in housing improvements in the preceding five years on property title status. As described in Section 3.3.2, I use the IHS transformation of the amount of investment.¹² Table 3.7 presents the regression coefficients from this estimation. Similar to its relationship with the decision to invest in housing improvements and the housing quality index, property title status has a “naive” average treatment effect on the amount of investment, but this relationship does not survive the addition of locality and home-state fixed effects, and controls for household and migration characteristics.

Next, I consider whether the absence of correlation between property titles and

¹¹This is also the case when I restrict the sample to households who have been in their current house for five or more years, as shown in Tables 3.A2 and 3.A3.

¹²In Section 3.B, I show that the results are robust to using $\log(\text{investment} + 1)$ rather than $IHS(\text{investment})$.

housing improvement decisions could be explained by differences in the duration for which the households have lived in their current residence. To do this, I additionally control for this duration in Equations 3.1 and 3.2. The results from this estimation are presented in Table 3.8. These results suggest that even after controlling for how long the households have lived in their current residence, in addition to the other observed covariates, property titles are not related to housing investment decisions or housing quality. Moreover, I cannot reject the null that the coefficient for “duration in current house” is equal to zero.

I also consider whether titled and non-titled households are systematically different in terms of their financial capabilities. I do this by regressing monthly household consumption expenditure and household asset indices on property title status and the full set of observed variables in Equation 3.1. The results are presented in Table 3.9. Property titles are not correlated with any of these proxies for the households’ financial capabilities, thus ruling out the possibility that non-titled households have more financial resources to invest in improving the quality of their residential structure.¹³

3.5.2 Matching estimators

In this section, I present the results from propensity score matching to study the relationship between property titles and housing investment decisions. I obtain propensity scores by estimating a logit regression of property title status on the observed variables. Given the propensity scores, I use an Epanechnikov kernel function to match a given titled household to a non-titled household with a similar propensity score.

Figure 3.2 shows the density distributions for the propensity scores of the unmatched (“raw”) sample and the matched sample, where the matched sample has been constructed using an Epanechnikov kernel function. In the left panel, we can compare the propensity score density distribution of titled and non-titled households

¹³This is also true when I restrict the sample to households who have been in their current house for five years or longer, as shown in Table 3.A5.

in the raw (unmatched) sample. As expected, titled households have a higher ‘propensity’ to possess property titles, as compared to non-titled households. In the right panel, we can compare the propensity score density distribution of the titled and non-titled households in the matched sample. Now these density distributions follow each other much more closely. In fact, except for the region around a propensity score of 0.9, the two distributions almost overlap with each other. Figure 3.3 shows the density distributions for propensity scores for the unmatched and matched samples in Mangolpuri. Here we observe that the density distributions for titled and non-titled households in the matched sample track each other very closely.

This is confirmed by the balance tests presented in Table 3.10. Panel A shows the balance tests for the full sample, Panel B for the Mangolpuri sub-sample. In Column 1, I present the results of balance tests obtained by imposing exact matching within locality and home-state cells, and in Column 2, I present the results of balance tests obtained by not imposing this exact within-cell matching.

With exact matching, all observed variables except the gender of the household head are balanced. Without exact matching, all observed variables are balanced, i.e. I cannot reject the null hypothesis that the titled and non-titled households have the same means for the observed variables. Since all observed variables are balanced without exact matching within cells, I will follow this matching strategy to evaluate the effect of property titles on key outcomes of interest. Nevertheless, I will also show that the results are robust to following the exact matching strategy.

Table 3.11 shows how the matched sample compares with the unmatched, or raw, sample. I compare the standardised differences in the means of the observed variables between the titled and non-titled households in the unmatched sample with the standardised differences in the matched sample. As expected, the standardised differences in the means for the titled and non-titled households are smaller in the matched sample.

Now I can use this matched sample to evaluate the effect of property titles on dif-

ferent outcomes of interest. I present these results in Table 3.12. For all of the outcomes of interest – whether the household has invested in home improvements during the preceding five years, the amount of investment in home improvements during the preceding five years, or the housing quality indices – I cannot reject the null that the effect of property titles on the relevant outcome is zero. This is the case with and without exact matching with home-state and locality cells. These results are robust to restricting the sample to Mangolpuri households (as shown in Panel B of Table 3.12) and to restricting the sample to households who have lived in their current residence for five years or longer (as shown in Table 3.A7).

As an additional robustness check, I restrict the sample to households that have migrated to Delhi. For these households, I have additional information on pre-treatment outcomes, namely, the number of years the household had been in Delhi when it moved to its current location, and the household's stated reason for migrating to Delhi. Table 3.13 presents the balance tests for the observed variables with and without exact matching within cells,¹⁴ and Table 3.14 compares the difference in means of the observed variables for the titled and non-titled households within the unmatched (raw) and matched sample.

As we can see from Table 3.15, adding the additional observed variables for the set of migrant households does not change the previous conclusion – property titles are not driving differences in the key outcomes of interest.

3.5.3 Selection on unobserved variables

In addition to the observed variables considered above, there may be other unobserved variables with which property titles and the outcome variables are correlated. Even though I cannot observe these variables, I can make inferences based on the relationship between the treatment and the observed variables (Altonji et al., 2005), and

¹⁴Now the matching is within cells constituted from locality, households' home state, and reason for migrating to Delhi.

by exploiting the shared covariance properties between the observed and unobserved variables and the treatment (Oster, 2017). In particular, I assess the extent to which the degree of selection on unobserved variables will have to differ from the degree of selection on observed variables, for my estimates of a null effect to mask a true non-zero effect of property titles on the key outcomes of interest.

The results are presented in Table 3.16. Property titles may have a statistically significant positive effect on the investment decision only when the degree of selection on unobserved variables is twice as large as the degree of selection on observed variables. Under all other reasonable values (1, or less than 1) for the relative degree of selection, property titles either reduce the probability of investment or have no effect on the investment outcomes.

3.5.4 Power analysis

It is possible that I do not find any effect of property titles on investment decisions because the study is under-powered to find any meaningful effects, given the size of the sample and the variability in the outcome variables. Therefore, I use power analysis to calculate the size of the minimum effect of property title status that can be detected for each outcome variable, given the study parameters. Since I do not find any statistically significant effect of property title status, I can conclude that the true effect is unlikely to be larger than the minimum detectable effect.

The minimum detectable effects are presented in Table 3.17. For the decision to invest in housing improvements, I find that the minimum effect size that can be detected in this study is 10.2 percentage points, or 45.33% of the control group mean (22.5%). Since I find no statistically significant effect, I can conclude that the true effect is unlikely to be larger than 45.33%. This is smaller than the effect size (68%) estimated for Peru in Field (2005) for the same variable. For the Mangolpuri sub-sample, the corresponding minimum detectable effect is 10.3 percentage points, or 63.58% of the control group mean (16.2%).

For the housing quality index, I find that the minimum effect size that can be detected with this study is 0.2 standard deviations. Since I find no effect of property title status on the housing quality index, I can conclude that the true effect of property titles on the index is unlikely to be larger than 0.2 standard deviations.

3.6 Mechanisms

Why do property titles seem to have no effect on decisions related to housing improvements? Besley (1995) lists three main reasons why property titles can be expected to enhance a household's investment incentives: by expanding the opportunities for trade in property markets, by reducing the risk that the property would be expropriated, and by providing collateral that can be used to access credit from the formal financial sector.¹⁵ Since property titles do not seem to have much of an effect on decisions related to housing improvements in Delhi's resettlement localities, I examine why these explanations may not be relevant in this context.

First, I consider the expansion of opportunities for trade. Without well-defined property titles, potential property buyers face the risk that the selling household does not actually own the property. Therefore, households that want to sell their properties are restricted to a small set of potential buyers, who may be related to them through village or kinship networks. This can reduce the market value of the property. Since households will be able to realise a smaller value for their property, this may reduce their incentives for investing in its improvement. Conversely, with well-defined titles, the set of potential buyers is no longer constrained, and households may be able to realise a higher market value, thereby strengthening their incentives for investment.

However, the titles in Delhi's resettlement localities came with an explicit restriction that the households could not alienate, or sell, the property (Bhan, 2009; Bhan and Shivanand, 2013). Therefore, we should not expect that the opportunities-for-trade

¹⁵Besley and Ghatak (2010) adds a fourth reason: by reducing the need to invest in guarding the property.

channel will strengthen the investment incentives of titled households. It is possible that the restriction on transferring titles may have reduced the value for which the households sold their properties through unregistered “power of attorney” documents, but without a counter-factual group of households who did not face this restriction, I cannot establish this empirically.

Secondly, well-defined property titles can help to insure the households against future eviction, either by the state or by non-state actors. Indeed, in Latin America, titles seem to have increased investment activity precisely by lowering eviction or expropriation risk (Field, 2005; Galiani and Schargrodsky, 2010). Unfortunately, the survey did not ask the households specific questions about their perceived risk of eviction. However, cross-country survey evidence suggests that differences in perceived eviction risk between titled and non-titled households in India are smaller than similar differences in eviction risk in Latin America. For example, when asked about their “worry of losing [their] right to current property...against will in the next 5 years,” a larger proportion of households in Colombia expressed such a worry as compared to households in India (PRIndex Analytical Report, 2017). The difference in the responses of those with formal property documents and those without was also smaller in India than in Colombia. In Delhi’s resettlement localities, households believe that the risk of eviction by the state is low because a large proportion of these households have already been evicted once by the state (Mehra et al., 2012).¹⁶

Finally, it is possible that property titles enhance households’ ability to access cheap credit from the formal financial sector (banks and microfinance institutions) by providing a valuable source of collateral. Even though households are not allowed to alienate the property titles, formal financial institutions still accept them as collateral. Their reasoning is that the physical, original title document is one of the most valuable possessions of these households. Because the financial institution takes possession of

¹⁶As noted by the authors in Mehra et al. (2012), “From the residents’ perspective in the newer settlements, it was clear that politically and practically another eviction does not seem possible and provision of facilities such as schools, dispensaries, and even a Mother Dairy (brand selling dairy products) booth all serve as signals of permanency of their settlement.”

the original title document while the loan is outstanding, the borrowers experience a psychic cost if they default on the loan amount, and stand to lose the original document.¹⁷

However, despite the collateral role of property titles, I find no evidence of their effect on households' investment decisions. Here I explore three reasons for this: (1) access to the formal financial sector is generally very low; (2) property titles do not explain households' access to the formal and informal sources of credit; and (3) investments in improving house quality are driven by household savings, rather than debt anyway.

Access to, and usage of, the formal financial sector is very low in these resettlement localities. Table 3.18 presents some descriptive statistics. Fewer than 60% households have a bank account, and only 20% have taken out a life insurance policy. Very few households have previously taken debt from a formal financial sector institution, such as bank (less than 2%) or a microfinance company (less than 1%). Finally, among households that report having any outstanding debt, 76% of this debt is owed to informal sector creditors, such as moneylenders, family members, employers, and in the form of goods on credit. Only 24% is owed to formal sector creditors such as banks, chit funds and microfinance companies.

Further, property titles are not positively correlated with access to formal or informal sources of lending. I measure access to formal sources of lending by creating an index based on whether the household has a bank account, a life insurance policy, and whether it has taken debt from a bank, a chit fund or a microfinance company. I measure access to informal sources of lending by creating an index based on whether the household has taken debt from moneylenders, family members, employers or in the form of goods on credit. Both indices are created using the method developed in Anderson (2008). These are fairly imperfect measures of access to formal and informal

¹⁷Source: author's discussions with banks and microfinance institutions operating in the resettlement localities, and also anecdotal evidence reported in Micro Home Solutions' report on these resettlement localities (Mehra et al., 2012).

credit, but provide us an indicative picture. To study the correlation between these indices and property titles, I estimate Equation 3.1 using these indices as outcome variables. The results are reported in Table 3.19. Property title status is negatively correlated with both these indices, though only the effect on the formal credit index is statistically significantly different from zero.

Finally, I show that households' decisions to invest in housing improvements during the preceding five years are better explained by their current stock of savings, rather than outstanding debt. To do this, I regress the decision to invest in housing improvements and the amount of investment on total savings and total debt, using the full set of controls and fixed effects specified in Equation 3.1. As we can see from Table 3.20, the total stock of savings is a better predictor of both the investment decision and the amount of investment than the total outstanding debt. Therefore, it is not surprising that in this context, property titles do not seem to be correlated with households' investment decisions.

3.7 Conclusion

Among policy-makers and commentators, the idea of giving formal property titles to the urban poor has long been considered a panacea: it is expected that these formal titles will unlock the "dead capital" of the urban poor and enable them to invest their way out of poverty (de Soto, 2000). It is not without reason: theoretically, property titles can enhance the investment opportunities of the poor (Besley, 1995; Besley and Ghatak, 2010), and in Latin America, there is ample empirical evidence that they seem to have achieved these goals (Field, 2005; Galiani and Schargrodsky, 2010). But do these effects generalise, and what should governments do to enhance these effects?

That is the question I seek to answer in this chapter, using cross-sectional data from Delhi's resettlement localities. I find that in these localities, formal property titles do not affect either the households' decision to invest in housing improvements, or

the amount of investment in housing improvements. They also have no effect on the quality of the households' current residential structure.

There are several econometric challenges in estimating the causal effect of property titles without randomly assigning these titles. I address these issues by first assuming selection on a set of observed household characteristics and employing a number of fixed effects in a linear regression framework. Further, I use propensity score matching to restrict comparisons within a group of titled and non-titled households that are very similar in terms of exogenous characteristics. Both the regression and the matching estimations yield the same result: property titles do not affect decisions related to investments in improving housing quality, or indeed housing quality itself. In addition, I show that these null results are not driven by the stronger financial position of non-titled households, and are not driven by the duration for which the households have lived in their current residence.

I consider two other possibilities: that there is selection on unobserved variables, and that the study is under-powered to detect any meaningful treatment effects. Using the method developed in Oster (2017), I show that the degree of selection on unobserved variables would have to be twice as large as the degree of selection on observed variables for my null results to mask a true non-zero effect of property titles on the housing investment decision. In terms of power, I find that even if the study is under-powered, the true treatment effect of property titles is unlikely to be large, and is certainly smaller than the effect estimated by Field (2005) for Peru.

Why are property titles not effective in Delhi's resettlement localities? While the given data is not appropriate to answer this question convincingly, I provide some evidence that in the absence of substantial differences in the perceived risk of eviction between titled and non-titled households, property titles do not affect investment decisions because they do not confer any transfer rights to their holders, and because households have very low levels of access to the formal financial sector.

These results suggest that if governments want to use property titling programs

to improve the economic lives of the urban poor, they should also promote well-functioning property markets through which the urban poor can transfer their properties at low transaction costs. Such markets would require good regulation, to ensure that the urban poor, who belong to some of the most vulnerable sections of society, are not defrauded of their property. Secondly, governments should work to expand access to the formal financial sector, so that households can leverage the titles as collateral and obtain low-cost credit from formal sector lenders. Once again, good regulation would be required to ensure that the urban poor do not fall into credit traps.

While the results in this chapter should temper the enthusiasm of property titling advocates, and foreground the importance of complementary inputs highlighted in the previous paragraph, we have to remember that these results are constrained by the cross-sectional nature of the study. If governments in India or elsewhere are really interested in rolling out large urban titling programs, they should invest in experimental work to rigorously investigate the effects of property titles on the desired household outcomes. Such experimental work can also be leveraged to understand which complementary inputs and institutions can maximise the benefits from property titles in developing countries.

Table 3.1
Summary statistics for main variables

	Full sample	Mangolpuri	Narela	Savda Ghevra
Share of households...				
...with property titles	0.842 (0.365)	0.561 (0.497)	0.975 (0.156)	0.956 (0.206)
...engaged in housing investment in last 5 years	0.412 (0.492)	0.166 (0.373)	0.167 (0.373)	0.878 (0.328)
Housing investment in last 5 years (Rs)	99,274 (119,895) (N=567)	105,143 (155,822) (N=70)	86,275 (108,354) (N=80)	100,783 (115,154) (N=417)
Monthly household expenditure (Rs)	5807.0 (3075.1)	6757.2 (3231.4)	5814.9 (2509.9)	4956.9 (3207.2)
Percent households with any savings	0.534 (0.499)	0.715 (0.452)	0.521 (0.500)	0.387 (0.488)
Total savings (Rs)	8972.9 (26691.8)	19934.4 (44287.6)	5250 (10908.5)	3019.5 (7227.7)
Years in current house	12.08 (10.47)	25.02 (9.738)	8.563 (2.329)	4.173 (2.083)
Percent migrated to Delhi	0.809 (0.393)	0.724 (0.447)	0.875 (0.331)	0.817 (0.387)
Years since migration	28.30 (12.19) (N=1084)	36.13 (12.98) (N= 294)	25.59 (10.07) (N=407)	25.17 (10.94) (N=383)
Household size	5.499 (1.924)	5.530 (2.025)	5.554 (1.802)	5.415 (1.952)
Household size at move	4.177 (1.926)	2.831 (1.503)	4.546 (1.725)	4.996 (1.827)
Household head male	0.827 (0.378)	0.850 (0.357)	0.798 (0.402)	0.836 (0.371)
Household head age	46.91 (12.00)	51.49 (12.01)	44.12 (10.71)	45.66 (12.06)
Observations	1376	421	480	475

Notes: This table presents summary statistics – mean and standard deviations – for the main variables used in the analysis, by each locality. For variables with some missing observations, the number of non-missing observations is specified separately in the row below the standard deviations.

Table 3.2
Summary statistics for housing quality

	Full sample	Mangolpuri	Narela	Savda Ghevra
House quality index (normalised)	-0.0337 (0.997)	1.064 (0.419)	-0.626 (0.747)	-0.408 (0.773)
House quality index (Anderson)	0.0993 (1.020)	1.091 (0.585)	-0.423 (0.853)	-0.252 (0.839)
Kitchen	0.267 (0.443)	0.565 (0.496)	0.135 (0.343)	0.137 (0.344)
Solid roof	0.619 (0.486)	0.948 (0.223)	0.371 (0.484)	0.579 (0.494)
Solid wall	0.873 (0.333)	0.957 (0.203)	0.821 (0.384)	0.851 (0.357)
Solid floor	0.742 (0.438)	0.865 (0.343)	0.652 (0.477)	0.724 (0.447)
House toilet	0.425 (0.495)	0.945 (0.228)	0.148 (0.355)	0.244 (0.430)
Piped water	0.308 (0.462)	0.971 (0.167)	0.0208 (0.143)	0.0105 (0.102)
Observations	1376	421	480	475

Notes: This table presents summary statistics – mean and standard deviations – for different measures of housing quality available in the dataset. These variables are used to construct two housing quality indices (HQI): (1) HQI (normalised) which is the sum of the normalised values of the different measures of housing quality; and (2) HQI (Anderson) which aggregates the different measures of housing quality using the procedure developed in Anderson (2008).

Table 3.3
Summary statistics for household assets

	Full sample	Mangolpuri	Narela	Savda Ghevra
Asset index (normalised)	-0.003 (0.971)	0.560 (1.188)	-0.226 (0.661)	-0.277 (0.795)
Asset index (Anderson)	0.091 (1.087)	0.635 (1.369)	-0.153 (0.758)	-0.145 (0.897)
Bicycle	0.271 (0.445)	0.352 (0.478)	0.133 (0.340)	0.339 (0.474)
Motorcycle	0.107 (0.309)	0.264 (0.441)	0.044 (0.205)	0.032 (0.175)
Car	0.005 (0.071)	0.017 (0.128)	0 (0)	0 (0)
TV	0.826 (0.380)	0.945 (0.228)	0.835 (0.371)	0.709 (0.454)
Washing machine	0.093 (0.291)	0.230 (0.422)	0.033 (0.180)	0.032 (0.175)
Livestock	0.007 (0.085)	0.010 (0.097)	0.006 (0.079)	0.006 (0.079)
Jewelry	0.624 (0.485)	0.698 (0.460)	0.652 (0.477)	0.528 (0.500)
Observations	1376	421	480	475

Notes: This table presents summary statistics – mean and standard deviations – for different household assets in the dataset. These variables are used to construct two asset indices: (1) Asset index (normalised) which is the sum of the normalised values of the different household assets; and (2) Asset index (Anderson) which aggregates the different household assets using the procedure developed in Anderson (2008).

Table 3.4
OLS results for decision to invest in housing improvements

<i>Panel A. For the full sample</i>					
	(1)	(2)	(3)	(4)	(5)
Property title	0.223 (0.032)***	0.016 (0.032)	0.014 (0.033)	0.020 (0.033)	-0.012 (0.042)
Outcome Mean	0.41	0.41	0.41	0.41	0.43
Locality fixed effects	No	Yes	Yes	Yes	Yes
Home state fixed effects	No	No	Yes	Yes	Yes
Household characteristics	No	No	No	Yes	Yes
Migration controls	No	No	No	No	Yes
N	1376	1376	1376	1376	1063

<i>Panel B. For the Mangolpuri sub-sample</i>					
	(1)	(2)	(3)	(4)	
Property title	0.007 (0.037)	-0.000 (0.037)	0.009 (0.038)	-0.038 (0.053)	
Outcome Mean	0.17	0.17	0.17	0.19	
Home state fixed effects	No	Yes	Yes	Yes	
Household characteristics	No	No	Yes	Yes	
Migration controls	No	No	No	Yes	
N	421	421	421	276	

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

Notes: This table presents OLS regression coefficients from estimating Equation 3.1 for whether or not the household has engaged in any housing investment in the preceding five years. Panel A contains results for the full sample of households. The locality fixed effects specification in Column 2 adds indicator variables for Mangolpuri, Narela and Savda Ghevra. The home-state fixed effects specification in Column 3 adds indicator variables for the households' home state. In Column 4, I add controls for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. In Column 5, I add controls for migration-related variables, which include reason for migrating to Delhi and the number of years the household had been in Delhi when it moved to the current residence. Column 5 has fewer observations because 19% of the sample are originally from Delhi and therefore no migration information is available for them. Panel B contains results for the sub-sample of households who live in Mangolpuri. The columns contain almost the same specifications as in Panel A, except for the locality fixed effects specification. Robust standard errors are reported below the coefficients.

Table 3.5
OLS results for housing quality index

<i>Panel A. For the full sample</i>					
	(1)	(2)	(3)	(4)	(5)
Property title	-0.998 (0.057)***	0.036 (0.044)	0.029 (0.044)	0.035 (0.045)	0.027 (0.060)
Outcome Mean	-0.034	-0.034	-0.034	-0.034	-0.117
Outcome SD	0.997	0.997	0.997	0.997	1.008
Locality fixed effects	No	Yes	Yes	Yes	Yes
Home state fixed effects	No	No	Yes	Yes	Yes
Household characteristics	No	No	No	Yes	Yes
Migration controls	No	No	No	No	Yes
N	1376	1376	1376	1376	1063

<i>Panel B. For the Mangolpuri sub-sample</i>					
	(1)	(2)	(3)	(4)	
Property title	0.051 (0.041)	0.044 (0.042)	0.035 (0.044)	0.006 (0.052)	
Outcome Mean	1.064	1.064	1.064	1.095	
Outcome SD	0.419	0.419	0.419	0.413	
Home state fixed effects	No	Yes	Yes	Yes	
Household characteristics	No	No	Yes	Yes	
Migration controls	No	No	No	Yes	
N	421	421	421	276	

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

Notes: This table presents OLS regression coefficients from estimating Equation 3.1 for the housing quality index (HQI) generated by aggregating the normalised values of six indicators of housing quality (whether the house has a separate kitchen, whether the roof is solid, whether the walls are solid, whether the floor is solid, whether house has a toilet, and whether it receives piped water). Panel A contains results for the full sample of households. The locality fixed effects specification in Column 2 adds indicator variables for Mangolpuri, Narela and Savda Ghevra. The home-state fixed effects specification in Column 3 adds indicator variables for the households' home state. In Column 4, I add controls for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. In Column 5, I add controls for migration-related variables, which include reason for migrating to Delhi and the number of years the household had been in Delhi when it moved to the current residence. Column 5 has fewer observations because 19% of the sample are originally from Delhi and therefore no migration information is available for them. Panel B contains results for the sub-sample of households who live in Mangolpuri. The columns contain almost the same specifications as in Panel A, except for the locality fixed effects specification. Robust standard errors are reported below the coefficients.

Table 3.6
OLS results for housing quality index following Anderson (2008)

<i>Panel A. For the full sample</i>					
	(1)	(2)	(3)	(4)	(5)
Property title	-0.892 (0.061)***	0.046 (0.055)	0.036 (0.055)	0.042 (0.056)	0.067 (0.071)
Outcome Mean	0.099	0.099	0.099	0.099	0.021
Outcome SD	1.020	1.020	1.020	1.020	1.037
Locality fixed effects	No	Yes	Yes	Yes	Yes
Home state fixed effects	No	No	Yes	Yes	Yes
Household characteristics	No	No	No	Yes	Yes
Migration controls	No	No	No	No	Yes
N	1376	1376	1376	1376	1063

<i>Panel B. For the Mangolpuri sub-sample</i>					
	(1)	(2)	(3)	(4)	
Property title	0.082 (0.058)	0.070 (0.059)	0.059 (0.061)	0.062 (0.074)	
Outcome Mean	1.091	1.091	1.091	1.128	
Outcome SD	0.585	0.585	0.585	0.585	
Home state fixed effects	No	Yes	Yes	Yes	
Household characteristics	No	No	Yes	Yes	
Migration controls	No	No	No	Yes	
N	421	421	421	276	

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

Notes: This table presents OLS regression coefficients from estimating Equation 3.1 for the housing quality index (HQI) generated by aggregating six indicators of housing quality (whether the house has a separate kitchen, whether the roof is solid, whether the walls are solid, whether the floor is solid, whether house has a toilet, and whether it receives piped water) using the procedure developed in Anderson (2008). Panel A contains results for the full sample of households. The locality fixed effects specification in Column 2 adds indicator variables for Mangolpuri, Narela and Savda Ghevra. The home-state fixed effects specification in Column 3 adds indicator variables for the households' home state. In Column 4, I add controls for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. In Column 5, I add controls for migration-related variables, which include reason for migrating to Delhi and the number of years the household had been in Delhi when it moved to the current residence. Column 5 has fewer observations because 19% of the sample are originally from Delhi and therefore no migration information is available for them. Panel B contains results for the sub-sample of households who live in Mangolpuri. The columns contain almost the same specifications as in Panel A, except for the locality fixed effects specification. Robust standard errors are reported below the coefficients.

Table 3.7
Tobit results for the IHS of housing investment amount

Panel A. For the full sample

	(1)	(2)	(3)	(4)	(5)
Property title	7.285 (1.198)***	0.474 (1.072)	0.434 (1.063)	0.656 (1.060)	-0.402 (1.281)
Outcome Mean	4.817	4.817	4.817	4.817	4.956
Locality fixed effects	No	Yes	Yes	Yes	Yes
Home state fixed effects	No	No	Yes	Yes	Yes
Household characteristics	No	No	No	Yes	Yes
Migration controls	No	No	No	No	Yes
N	1376	1376	1376	1376	1063

Panel B. For the Mangolpuri sub-sample

	(1)	(2)	(3)	(4)
Property title	0.338 (2.525)	0.045 (2.480)	0.835 (2.503)	-2.207 (2.879)
Outcome Mean	1.892	1.892	1.892	2.166
Home state fixed effects	No	Yes	Yes	Yes
Household characteristics	No	No	Yes	Yes
Migration controls	No	No	No	Yes
N	421	421	421	276

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

Notes: This table presents Tobit regression coefficients from estimating Equation 3.2 for the inverse hyperbolic sine (IHS) transformation of the amount of housing investment in the preceding five years. Panel A contains results for the full sample of households. The locality fixed effects specification in Column 2 adds indicator variables for Mangolpuri, Narela and Savda Ghevra. The home-state fixed effects specification in Column 3 adds indicator variables for the households' home state. In Column 4, I add controls for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. In Column 5, I add controls for migration-related variables, which include reason for migrating to Delhi and the number of years the household had been in Delhi when it moved to the current residence. Column 5 has fewer observations because 19% of the sample are originally from Delhi and therefore no migration information is available for them. Panel B contains results for the sub-sample of households who live in Mangolpuri. The columns contain almost the same specifications as in Panel A, except for the locality fixed effects specification. Robust standard errors are reported below the coefficients.

Table 3.8
Regressions controlling for duration in the current house

Panel A. For the full sample

	Decision to invest	Housing investment (IHS)	HQI (normalised)	HQI (Anderson)
Property title	0.027 (0.033)	0.865 (1.068)	0.046 (0.046)	0.051 (0.057)
Duration in current house	-0.002 (0.002)	-0.079 (0.064)	-0.004 (0.003)	-0.003 (0.003)
Outcome Mean	0.412	4.817	-0.034	0.099
Outcome SD	0.492	5.799	0.997	1.019
Locality fixed effects	Yes	Yes	Yes	Yes
Home state fixed effects	Yes	Yes	Yes	Yes
Household characteristics	Yes	Yes	Yes	Yes
N	1376	1376	1376	1376

Panel B. For the Mangolpuri sub-sample

	Decision to invest	Housing investment (IHS)	HQI (normalised)	HQI (Anderson)
Property title	0.007 (0.038)	0.701 (2.513)	0.042 (0.045)	0.064 (0.062)
Duration in current house	0.001 (0.002)	0.079 (0.173)	-0.005 (0.003)	-0.003 (0.004)
Outcome Mean	0.166	1.892	1.064	1.091
Outcome SD	0.373	4.282	0.419	0.585
Locality fixed effects	Yes	Yes	Yes	Yes
Home state fixed effects	Yes	Yes	Yes	Yes
Household characteristics	Yes	Yes	Yes	Yes
N	421	421	421	421

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

Notes: This table presents OLS regression coefficients obtained from estimating Equation 3.1 for the decision to invest in improving housing quality and housing quality indices, and Tobit coefficients obtained from estimating Equation 3.2 for the amount of investment in improving housing quality during the preceding five years. In these regressions, I additionally control for the households' duration of stay in their current house. Panel A contains results for the full sample of households, Panel B contains results for the sub-sample of households who live in Mangolpuri. All specifications include locality fixed effects, home state fixed effects, and control for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. Robust standard errors are reported below the coefficients.

Table 3.9
Correlation between property titles and consumption and assets

Panel A. For the full sample

	Log expenditure	Asset index (normalised)	Asset index (Anderson)
Property title	0.007 (0.034)	-0.125 (0.099)	-0.097 (0.114)
Outcome Mean	8.558	-0.003	0.091
Outcome SD	0.464	0.970	1.087
Locality fixed effects	Yes	Yes	Yes
Home state fixed effects	Yes	Yes	Yes
Household characteristics	Yes	Yes	Yes
N	1376	1376	1376

Panel B. For the Mangolpuri sub-sample

	Log expenditure	Asset index (normalised)	Asset index (Anderson)
Property title	-0.013 (0.044)	-0.141 (0.126)	-0.113 (0.149)
Outcome Mean	8.722	0.561	0.635
Outcome SD	0.436	1.188	1.369
Locality fixed effects	Yes	Yes	Yes
Home state fixed effects	Yes	Yes	Yes
Household characteristics	Yes	Yes	Yes
N	421	421	421

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

Notes: This table presents OLS regression results obtained by regressing three measures of household financial capability: (1) log of monthly household expenditure; (2) an asset index generated by aggregating the normalised values of different household assets; and (3) an asset index generated by aggregating different household assets using the procedure developed by Anderson (2008). The different household assets are bicycle, motorcycle, car, television, washing machine, livestock, and jewellery. Panel A contains results for the full sample of households, Panel B contains results for the sub-sample of households who live in Mangolpuri. All specifications include locality fixed effects, home state fixed effects, and control for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. Robust standard errors are reported below the coefficients.

Table 3.10
Balance table for kernel matching using propensity scores

Panel A. For the full sample

Controls	With exact matching	Without exact matching
HH size at move	0.118 (0.125)	0.222 (0.218)
HH head age	2.394 (1.844)	0.394 (1.784)
HH head age squared	217.316 (184.720)	11.768 (191.253)
HH head male	-0.106 (0.036)***	-0.063 (0.038)
Number of matched observations	1205	1295

Panel B. For the Mangolpuri sub-sample

Controls	With exact matching	Without exact matching
HH size at move	-0.052 (0.057)	0.043 (0.079)
HH head age	-0.003 (0.545)	0.070 (0.737)
HH head age squared	9.801 (56.864)	5.405 (77.921)
HH head male	0.009 (0.020)	0.010 (0.020)
Number of matched observations	401	395

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

Notes: This table tests whether the matching covariates are balanced between the matched treatment and control groups. I use the following matching covariates: (1) size of household at the time of moving to the current location; (2) age of the household head; (3) age squared of the household head; and (4) gender of the household head. Column 2 contains results obtained from forcing the matching procedure to match exactly on households' home state and locality. Column 3 contains results obtained from not forcing an exact match on households' home state and locality. Standard errors are obtained using bootstrapping with 50 replications. All estimations account for regression adjustment with respect to the matching variables after matching (equivalent to the bias-correction proposed by Abadie and Imbens (2011)).

Table 3.11
Differences between raw and matched samples using propensity score matching

<i>Panel A. For the full sample</i>						
	Raw (full sample) (N=1376)			Matched sample (N=1205)		
Variables	Title	No Title	St Diff	Title	No Title	St Diff
HH size at move	4.323	3.399	0.493	4.160	4.041	0.063
HH head age	46.519	48.977	-0.203	47.053	44.659	0.198
HH head age squared	2305.79	2548.2	-0.195	2359.95	2142.64	0.175
HH head male	0.820	0.862	-0.115	0.819	0.925	-0.289

<i>Panel B. For the Mangolpuri sub-sample</i>						
	Raw (full sample) (N=421)			Matched sample (N=401)		
Variables	Title	No Title	St Diff	Title	No Title	St Diff
HH size at move	2.585	3.146	-0.372	2.711	2.762	-0.034
HH head age	52.771	49.854	0.244	51.454	51.457	-0.0002
HH head age squared	2926.76	2627.18	0.235	2795.61	2785.81	0.008
HH head male	0.847	0.854	-0.018	0.847	0.838	0.025

Notes: This table presents the means of the matching covariates for both treatment and control households, for the raw data (i.e. the full sample) and the matched sample. It also shows the standardised difference (St Diff) between the treatment and control means within the full sample, and within the matched sample. The matching procedure does not impose an exact match within locality and home-state cells.

Table 3.12
Results from kernel matching using propensity scores

Panel A. For the full sample

Outcomes	With exact matching	Without exact matching
Decision to invest	-0.007 (0.075)	-0.045 (0.948)
Housing investment (IHS)	0.087 (0.861)	-0.425 (0.948)
HQI (normalised)	-0.404 (0.388)	-0.344 (0.470)
HQI (Anderson)	-0.098 (0.114)	-0.094 (0.134)
Number of matched observations	1205	1295

Panel A. For the full sample

Outcomes	With exact matching	Without exact matching
Decision to invest	0.009 (0.033)	0.009 (0.043)
Housing investment (IHS)	-0.007 (0.395)	0.026 (0.437)
HQI (normalised)	0.147 (0.175)	0.084 (0.169)
HQI (Anderson)	0.055 (0.056)	0.040 (0.062)
Number of matched observations	401	395

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

Notes: This table presents the average treatment effect of property title status on four outcomes of interest: (1) whether the household has invested in improving housing quality in the preceding 5 years; (2) the inverse hyperbolic sine (IHS) transformation of the amount of housing investment in the preceding 5 years; (3) a housing quality index generated by aggregating the normalised values of 6 indicators (kitchen, solid roof, solid wall, solid floor, toilet, piped water); and (4) a housing quality index generated using the procedure developed in Anderson (2008). The matching is done using propensity scores. Column 2 contains results obtained from forcing the matching procedure to match exactly on households' home state and locality. Column 3 contains results obtained from not forcing an exact match on households' home state and locality. Standard errors are estimated using bootstrapping with 50 replications. All estimations account for regression adjustment with respect to the matching variables after matching (equivalent to the bias-correction proposed by Abadie and Imbens (2011)).

Table 3.13
Balance table for kernel matching using propensity scores, with
migration variables

Panel A. For the full sample

Controls	With exact matching	Without exact matching
HH size at move	0.019 (0.226)	0.145 (0.344)
HH head age	1.576 (1.231)	-0.459 (2.368)
HH head age squared	125.181 (125.069)	-94.037 (241.194)
HH head male	-0.093 (0.030)***	-0.090 (0.034)***
Years since migration at move	1.959 (2.205)	0.955 (2.774)
Number of matched observations	834	970

Panel B. For the Mangolpuri sub-sample

Controls	With exact matching	Without exact matching
HH size at move	-0.267 (0.093)***	-0.137 (0.110)
HH head age	-0.002 (0.953)	-0.034 (1.039)
HH head age squared	23.369 (105.805)	-2.796 (118.322)
HH head male	0.008 (0.032)	0.018 (0.032)
Years since migration at move	-0.156 (1.032)	0.235 (0.962)
Number of matched observations	245	263

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

Notes: This table tests whether the matching covariates are balanced between the matched treatment and control groups. For the set of migrant households, I use the following variables to match treatment households to control households: (1) size of household at the time of moving to the current location; (2) age of the household head; (3) age squared of the household head; (4) gender of the household head; and (5) number of years between their migration to Delhi and their move to the current house. Column 2 contains results obtained from forcing the matching procedure to match exactly on households' home state, locality, and reason for migrating to Delhi. Column 3 contains results obtained from not forcing an exact match on households' home state, locality, and reason for migrating to Delhi. Standard errors are estimated using bootstrapping with 50 replications. All estimations account for regression adjustment with respect to the matching variables after matching (equivalent to the bias-correction proposed by Abadie and Imbens (2011)).

Table 3.14
Differences between raw and matched samples for propensity score matching, with migration variables

Panel A. For the full sample

Variables	Raw (full sample) (N=1063)			Matched sample (N=834)		
	Title	No Title	St Diff	Title	No Title	St Diff
HH size at move	4.407	3.562	0.453	4.188	4.169	0.010
HH head age	46.202	49.301	-0.254	46.977	45.401	0.129
HH head age squared	2279.18	2583.1	-0.241	2358.06	2232.88	0.099
HH head male	0.819	0.904	-0.248	.836	.929	-0.272
Years since migration at move	17.977	13.292	0.401	17.173	15.215	0.168

Panel B. For the Mangolpuri sub-sample

Variables	Raw (full sample) (N=276)			Matched sample (N=245)		
	Title	No Title	St Diff	Title	No Title	St Diff
HH size at move	2.617	3.393	-0.487	2.776	3.044	-0.168
HH head age	54.338	50.369	0.330	53.547	53.549	-0.0002
HH head age squared	3097.38	2679.78	0.318	3023.41	3000.04	0.018
HH head male	0.883	0.893	-0.033	0.872	0.865	0.025
Years since migration at move	12.312	12.648	-0.029	12.205	12.360	-0.013

Notes: This table presents the means of the matching covariates for both treatment and control households, for the raw data (i.e. the full sample) and the matched sample. It also shows the standardised difference (St Diff) between the treatment and control means within the full sample, and within the matched sample. The matching procedure does not force an exact match on home state, locality, and reason for migrating cells.

Table 3.15
Results from kernel matching using propensity scores, with migration variables

<i>Panel A. For the full sample</i>		
Outcomes	With exact matching	Without exact matching
Decision to invest	-0.040 (0.085)	-0.018 (0.110)
Housing investment (IHS)	-0.331 (1.147)	-0.103 (1.243)
HQI (normalised)	-0.182 (0.678)	-1.012 (0.697)
HQI (Anderson)	-0.008 (0.143)	-0.240 (0.184)
Number of matched observations	834	970
<i>Panel B. For the Mangolpuri sub-sample</i>		
Outcomes	With exact matching	Without exact matching
Decision to invest	-0.052 (0.059)	-0.067 (0.063)
Housing investment (IHS)	-0.744 (0.629)	-0.928 (0.688)
HQI (normalised)	0.054 (0.271)	-0.053 (0.197)
HQI (Anderson)	0.070 (0.086)	0.031 (0.085)
Number of matched observations	245	263

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

Notes: This table presents the average treatment effect of property title status on four outcomes of interest: (1) whether the household has invested in improving housing quality in the preceding 5 years; (2) the inverse hyperbolic sine (IHS) transformation of the amount of housing investment in the preceding 5 years; (3) a housing quality index generated by aggregating the normalised values of 6 indicators (kitchen, solid roof, solid wall, solid floor, toilet, piped water); and (4) a housing quality index generated using the procedure developed in Anderson (2008). The matching is done using propensity scores. Column 2 contains results obtained from forcing the matching procedure to match exactly on households' home state, locality, and reason for migrating to Delhi. Column 3 contains results obtained from not forcing an exact match on households' home state, locality, and reason for migrating to Delhi. Standard errors are estimated using bootstrapping with 50 replications. All estimations account for regression adjustment with respect to the matching variables after matching (equivalent to the bias-correction proposed by Abadie and Imbens (2011)).

Table 3.16
Selection on unobserved variables

Value of δ	β_1	95% Confidence Interval	Completed replications
$\delta = 2$	1.419 (0.156) ^{***}	[1.113,1.725]	21
$\delta = 1.5$	1.730 (1.132)	[-0.488,3.948]	12
$\delta = 1$	-0.399 (0.076) ^{***}	[-0.547,-0.250]	15
$\delta = 0.5$	-0.149 (0.047) ^{***}	[-0.240, -0.057]	15
$\delta = 0.4$	-0.110 (0.039) ^{***}	[-0.186,-0.035]	12
$\delta = 0.3$	-0.075 (0.036) ^{**}	[-0.145,-0.004]	15
$\delta = 0.25$	-0.058 (0.047)	[-0.150,0.035]	17
$\delta = 0.2$	-0.041 (0.033)	[-0.105,0.023]	21
$\delta = 0.1$	-0.010 (0.027)	[-0.062,0.043]	12
$\delta = 0.05$	0.006 (0.033)	[-0.059,0.071]	10
N	1376		

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

Notes: This table presents estimates for the treatment effect of property titles on the households' decision to invest in housing improvements during the preceding five years, assuming different values δ , the degree of selection on unobserved variables relative to the degree of selection on observed variables. Column 1 lists the different values of δ . Column 2 lists the value of the treatment effect, β , assuming the particular value of δ . Bootstrapped standard errors are reported below the values of β . Column 3 lists the implied 95% confidence interval. To obtain the standard errors through bootstrap, 100 replications were run. The number of completed replications, where all parameters were estimated, is mentioned in Column 4.

Table 3.17
Minimum detectable effect sizes

Panel A. For the full sample

Outcome	Control group mean	Minimum detectable effect size	Implied treatment group mean
Decision to invest	0.225	0.102	0.327
IHS investment	2.588	1.200	3.788
HQI (normalised)	0.806	0.206	1.013
HQI (Anderson)	0.850	0.211	1.061

Panel B. For the Mangolpuri sub-sample

Outcome	Control group mean	Minimum detectable effect size	Implied treatment group mean
Decision to invest	0.162	0.103	0.265
IHS investment	1.891	1.181	3.072
HQI (normalised)	1.035	0.116	1.151
HQI (Anderson)	1.045	0.161	1.207

Notes: This table presents estimates for the size of minimum effects of property titles on different outcomes that can be detected with the given sample size, using confidence intervals of 95% and a power value of 0.8.

Table 3.18
Experience with formal financial sector

	Full sample	Mangolpuri	Narela	Savda Ghevra
Bank account	0.593 (0.491)	0.781 (0.414)	0.460 (0.499)	0.560 (0.497)
Life insurance policy	0.200 (0.400)	0.328 (0.470)	0.115 (0.319)	0.173 (0.378)
Taken bank debt	0.0196 (0.139)	0.0214 (0.145)	0.0104 (0.102)	0.0274 (0.163)
Taken microfinance debt	0.00799 (0.0891)	0.00238 (0.0487)	0.0208 (0.143)	0 (0)
Total savings (INR)	8,972.9 (26,691.8)	19,934.4 (44,287.6)	5,250 (10,908.5)	3,019.5 (7,227.7)
Total debt (INR)	4,566.9 (24,643.3)	5,156.8 (33,901.3)	3,039.6 (14,790.7)	5,587.4 (22,770.9)
Share of formal sector debt	0.244 (0.429)	0.312 (0.471)	0.278 (0.452)	0.179 (0.382)
Share of informal sector debt	0.756 (0.429)	0.688 (0.471)	0.722 (0.452)	0.821 (0.382)
	(N= 149)	(N=32)	(N=54)	(N=63)
N	1376	421	480	475

Notes: This table presents descriptive statistics for various indicators measuring households' participation in the formal financial sector across the three resettlement localities.

Table 3.19
Property titles and access to formal and informal credit sources

	Formal credit index	Informal credit index
Property title	-0.338 (0.087)***	-0.084 (0.092)
Outcome Mean	.07974	-.003201
Locality fixed effects	Yes	Yes
Home-state fixed effects	Yes	Yes
Household characteristics	Yes	Yes
N	1376	1376

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

Notes: This table presents OLS regression coefficients from estimating Equation 3.1 for indices measuring households' participation in the formal and informal financial sector. The formal credit index is constructed based on whether the household has a bank account, a life insurance policy, and whether it has taken debt from a bank, a chit fund or a microfinance company. The informal credit index is constructed based on whether the household has taken debt from moneylenders, family members, employers or in the form of goods on credit. Both indices are created using the method developed in (Anderson, 2008). All regressions include locality fixed effects, home-state fixed effects and control for exogenous household characteristics: household size at the time of moving to the current residence, age of household head, age squared of household head, and gender of household head. Robust standard errors are reported below the coefficients.

Table 3.20
Role of savings and debt in financing investment in housing improvements

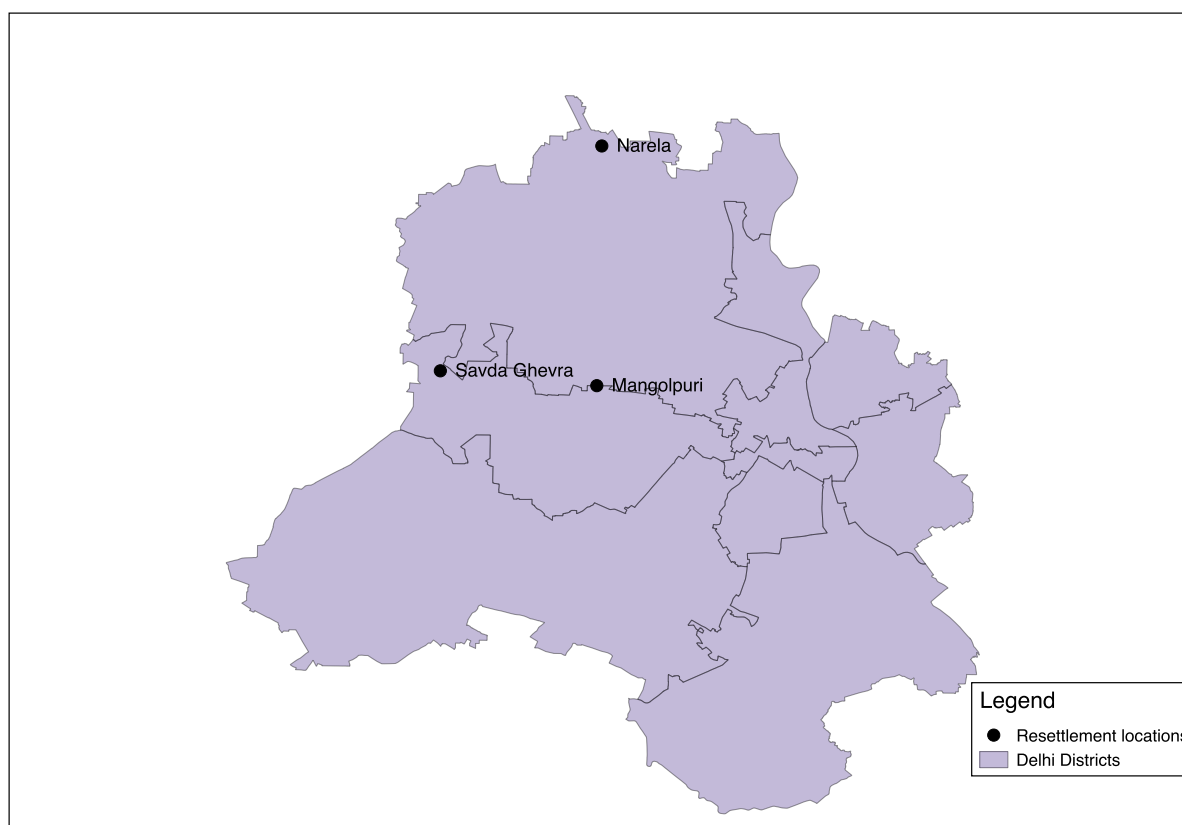
<i>Panel A. For the decision to invest in housing improvements</i>			
	(1)	(2)	(3)
IHS (Total savings)	0.005 (0.002)**		0.005 (0.002)**
IHS (Total debt)		0.004 (0.003)	0.004 (0.003)
Outcome Mean	0.412	0.412	0.412
Locality fixed effects	Yes	Yes	Yes
Home-state fixed effects	Yes	Yes	Yes
Household characteristics	Yes	Yes	Yes
N	1376	1376	1376

<i>Panel B. For the amount of investment in housing improvements</i>			
	(1)	(2)	(3)
IHS (Total savings)	0.149 (0.059)**		0.159 (0.059)***
IHS (Total debt)		0.113 (0.085)	0.134 (0.085)
Outcome Mean	4.817	4.817	4.817
Locality fixed effects	Yes	Yes	Yes
Home-state fixed effects	Yes	Yes	Yes
Household characteristics	Yes	Yes	Yes
N	1376	1376	1376

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

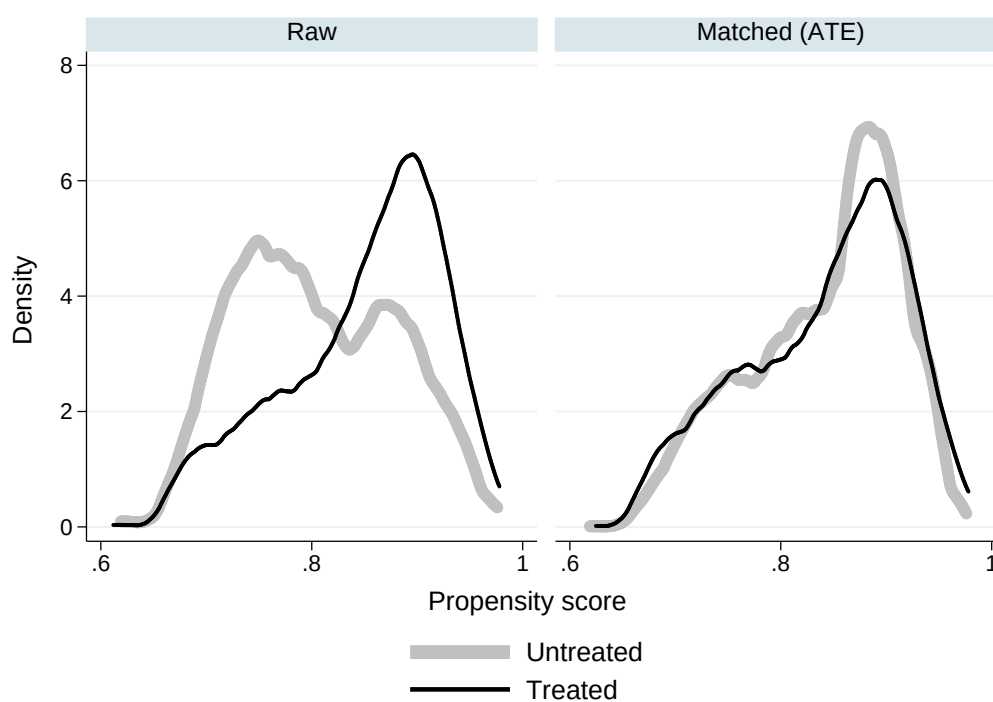
Notes: This table presents OLS regression coefficients from regressing households' investment behaviour on their stock of total savings and total debt. Panel A shows regression coefficients using the decision to invest in housing improvements as the outcome variable, and Panel B shows regression coefficients using the IHS value of the amount of investment in improving housing quality as the outcome variable. All regressions include locality fixed effects, home-state fixed effects, and control for the full set of exogenous housing characteristics. Robust standard errors are reported below the coefficients.

Figure 3.1
Location of the three resettlement localities



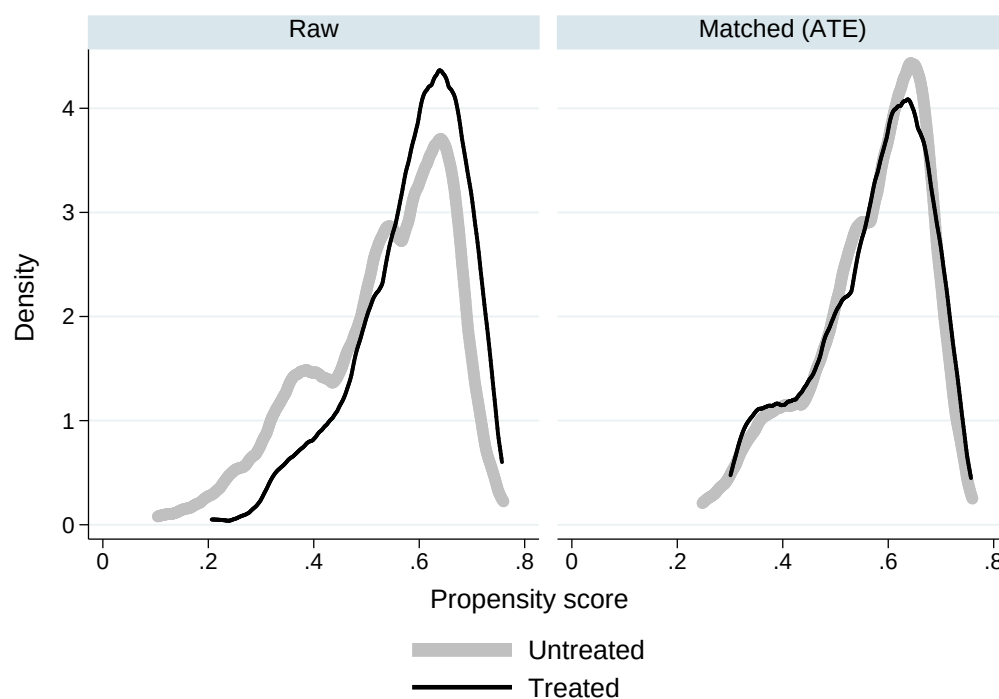
Notes: This map shows the location of Mangolpuri, Narela and Savda Ghevra in Delhi. Mangolpuri was established in 1976, while Narela and Savda Ghevra have been established more recently in the 2000s.

Figure 3.2
Density distributions for propensity score matching



Notes: This figure shows the density distributions of the propensity scores for the unmatched (left panel) and the matched (right panel) sample from all three resettlement localities. The black line represents the density distribution for the sample of treated households (those with formal property titles) and the grey line represents the density distribution for the sample of untreated households (those without formal property titles).

Figure 3.3
Density distributions for propensity score matching for Mangolpuri sub-sample



Notes: This figure shows the density distributions of the propensity scores for the unmatched (left panel) and the matched (right panel) sample from Mangolpuri. The black line represents the density distribution for the sample of treated households (those with formal property titles) and the grey line represents the density distribution for the sample of untreated households (those without formal property titles).

Appendix

3.A Restricting sample to households that have lived in their current house for five years or longer

Table 3.A1
OLS results for decision to invest in housing improvements

	(1)	(2)	(3)	(4)	(5)
Property title	0.014 (0.031)	0.005 (0.036)	0.004 (0.036)	0.010 (0.037)	-0.034 (0.048)
Outcome Mean	0.178	0.178	0.178	0.178	0.195
Locality fixed effects	No	Yes	Yes	Yes	Yes
Home state fixed effects	No	No	Yes	Yes	Yes
Household characteristics	No	No	No	Yes	Yes
Migration controls	No	No	No	No	Yes
N	864	864	864	864	650

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

Notes: This table presents OLS regression coefficients from estimating Equation 3.1 for whether or not the household has engaged in any housing investment in the preceding five years. I restrict the sample to households that have lived in their current house for five years or longer. Panel A contains results for the full sample of households. The locality fixed effects specification in Column 2 adds indicator variables for Mangolpuri, Narela and Savda Ghevra. The home-state fixed effects specification in Column 3 adds indicator variables for the households' home state. In Column 4, I add controls for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. In Column 5, I add controls for migration-related variables, which include reason for migrating to Delhi and the number of years the household had been in Delhi when it moved to the current residence. Column 5 has fewer observations because 19% of the sample are originally from Delhi and therefore no migration information is available for them. Panel B contains results for the sub-sample of households who live in Mangolpuri. The columns contain almost the same specifications as in Panel A, except for the locality fixed effects specification. Robust standard errors are reported below the coefficients.

Table 3.A2
OLS results for housing quality index

	(1)	(2)	(3)	(4)	(5)
Property title	-0.998 (0.060)***	0.047 (0.045)	0.038 (0.045)	0.035 (0.046)	0.011 (0.058)
Outcome Mean	0.156	0.156	0.156	0.156	0.050
Locality fixed effects	No	Yes	Yes	Yes	Yes
Home state fixed effects	No	No	Yes	Yes	Yes
Household characteristics	No	No	No	Yes	Yes
Migration controls	No	No	No	No	Yes
N	864	864	864	864	650

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

Notes: This table presents OLS regression coefficients from estimating Equation 3.1 for the housing quality index (HQI) generated by aggregating the normalised values of six indicators of housing quality (whether the house has a separate kitchen, whether the roof is solid, whether the walls are solid, whether the floor is solid, whether house has a toilet, and whether it receives piped water). I restrict the sample to households that have lived in their current house for five years or longer. Panel A contains results for the full sample of households. The locality fixed effects specification in Column 2 adds indicator variables for Mangolpuri, Narela and Savda Ghevra. The home-state fixed effects specification in Column 3 adds indicator variables for the households' home state. In Column 4, I add controls for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. In Column 5, I add controls for migration-related variables, which include reason for migrating to Delhi and the number of years the household had been in Delhi when it moved to the current residence. Column 5 has fewer observations because 19% of the sample are originally from Delhi and therefore no migration information is available for them. Panel B contains results for the sub-sample of households who live in Mangolpuri. The columns contain almost the same specifications as in Panel A, except for the locality fixed effects specification. Robust standard errors are reported below the coefficients.

Table 3.A3
OLS results for housing quality index following Anderson (2008)

	(1)	(2)	(3)	(4)	(5)
Property title	-0.876 (0.067)***	0.074 (0.059)	0.061 (0.059)	0.056 (0.060)	0.062 (0.073)
Outcome Mean	0.276	0.276	0.276	0.276	0.179
Locality fixed effects	No	Yes	Yes	Yes	Yes
Home state fixed effects	No	No	Yes	Yes	Yes
Household characteristics	No	No	No	Yes	Yes
Migration controls	No	No	No	No	Yes
N	864	864	864	864	650

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

Notes: This table presents OLS regression coefficients from estimating Equation 3.1 for the housing quality index (HQI) generated by aggregating six indicators of housing quality (whether the house has a separate kitchen, whether the roof is solid, whether the walls are solid, whether the floor is solid, whether house has a toilet, and whether it receives piped water) using the procedure developed in Anderson (2008). I restrict the sample to households that have lived in their current house for five years or longer. Panel A contains results for the full sample of households. The locality fixed effects specification in Column 2 adds indicator variables for Mangolpuri, Narela and Savda Ghevra. The home-state fixed effects specification in Column 3 adds indicator variables for the households' home state. In Column 4, I add controls for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. In Column 5, I add controls for migration-related variables, which include reason for migrating to Delhi and the number of years the household had been in Delhi when it moved to the current residence. Column 5 has fewer observations because 19% of the sample are originally from Delhi and therefore no migration information is available for them. Panel B contains results for the sub-sample of households who live in Mangolpuri. The columns contain almost the same specifications as in Panel A, except for the locality fixed effects specification. Robust standard errors are reported below the coefficients.

Table 3.A4
Tobit results for IHS of housing investment amount

	(1)	(2)	(3)	(4)	(5)
Property title	0.800 (2.082)	0.006 (2.255)	0.215 (2.198)	0.723 (2.195)	-1.771 (2.415)
Outcome Mean	2.041	2.041	2.041	2.041	2.222
Locality fixed effects	No	Yes	Yes	Yes	Yes
Home state fixed effects	No	No	Yes	Yes	Yes
Household characteristics	No	No	No	Yes	Yes
Migration controls	No	No	No	No	Yes
N	864	864	864	864	650

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

Notes: This table presents Tobit regression coefficients from estimating Equation 3.2 for the inverse hyperbolic sine (IHS) transformation of the amount of housing investment in the preceding five years. I restrict the sample to households that have lived in their current house for five years or longer. Panel A contains results for the full sample of households. The locality fixed effects specification in Column 2 adds indicator variables for Mangolpuri, Narela and Savda Ghevra. The home-state fixed effects specification in Column 3 adds indicator variables for the households' home state. In Column 4, I add controls for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. In Column 5, I add controls for migration-related variables, which include reason for migrating to Delhi and the number of years the household had been in Delhi when it moved to the current residence. Column 5 has fewer observations because 19% of the sample are originally from Delhi and therefore no migration information is available for them. Panel B contains results for the sub-sample of households who live in Mangolpuri. The columns contain almost the same specifications as in Panel A, except for the locality fixed effects specification. Robust standard errors are reported below the coefficients.

Table 3.A5
Correlation between property titles and consumption and assets

	Log expenditure	Asset index (normalised)	Asset index (Anderson)
Property title	-0.014 (0.040)	-0.164 (0.117)	-0.145 (0.137)
Outcome Mean	8.632	0.133	0.209
Locality fixed effects	Yes	Yes	Yes
Home state fixed effects	Yes	Yes	Yes
Household characteristics	Yes	Yes	Yes
N	864	864	864

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

Notes: This table presents OLS regression results obtained by regressing three measures of household financial capability: (1) log of monthly household expenditure; (2) an asset index generated by aggregating the normalised values of different household assets; and (3) an asset index generated by aggregating different household assets using the procedure developed by Anderson (2008). The different household assets are bicycle, motorcycle, car, television, washing machine, livestock, and jewellery. I restrict the sample to households that have lived in their current house for five years or longer. Panel A contains results for the full sample of households, Panel B contains results for the sub-sample of households who live in Mangolpuri. All specifications include locality fixed effects, home state fixed effects, and control for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. Robust standard errors are reported below the coefficients.

Table 3.A6
OLS regressions controlling for duration in the current house

	Decision to invest	Housing investment (IHS)	HQI (normalised)	HQI (Anderson)
Property title	0.009 (0.037)	0.612 (2.217)	0.050 (0.047)	0.072 (0.061)
Duration in current house	0.001 (0.002)	0.037 (0.140)	-0.005 (0.003)*	-0.005 (0.004)
Outcome Mean	0.178	2.041	0.156	0.276
Locality fixed effects	Yes	Yes	Yes	Yes
Home state fixed effects	Yes	Yes	Yes	Yes
Household characteristics	Yes	Yes	Yes	Yes
N	864	864	864	864

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

Notes: This table presents OLS regression coefficients obtained from estimating Equation 3.1 and Equation 3.2 for the key outcomes of interest, also controlling for the households' duration of stay in the current house. I restrict the sample to households that have lived in their current house for five years or longer. Panel A contains results for the full sample of households, Panel B contains results for the sub-sample of households who live in Mangolpuri. All specifications include locality fixed effects, home-state fixed effects, and control for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. Robust standard errors are reported below the coefficients.

Table 3.A7
Results from kernel matching using propensity scores

Outcomes	With exact matching	Without exact matching
Decision to invest	-0.135 (0.113)	-0.099 (0.106)
Housing investment (IHS)	-1.621 (1.306)	-1.208 (1.179)
HQI (normalised)	-0.070 (0.244)	-0.189 (0.169)
HQI (Anderson)	-0.051 (0.249)	-0.185 (0.197)
Number of matched observations	750	819

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

Notes: This table presents the average treatment effect of property title status on four outcomes of interest: (1) whether the household has invested in improving housing quality in the preceding 5 years; (2) the inverse hyperbolic sine (IHS) transformation of the amount of housing investment in the preceding 5 years; (3) a housing quality index generated by aggregating the normalised values of 6 indicators (kitchen, solid roof, solid wall, solid floor, toilet, piped water); and (4) a housing quality index generated using the procedure developed in Anderson (2008). I restrict the sample to households that have lived in their current house for five years or longer. The matching is done using propensity scores. Column 2 contains results obtained from forcing the matching procedure to match exactly on households' home state and locality. Column 3 contains results obtained from not forcing an exact match on households' home state and locality. Standard errors are estimated using bootstrapping with 50 replications. All estimations account for regression adjustment with respect to the matching variables after matching (equivalent to the bias-correction proposed by Abadie and Imbens (2011)).

Appendix

3.B Robustness to using log investment

Table 3.B1
Tobit results for the log of housing investment amount

Panel A. For the full sample

	(1)	(2)	(3)	(4)	(5)
Property title	6.861 (1.128)***	0.447 (1.009)	0.409 (1.001)	0.619 (0.998)	-0.378 (1.205)
Outcome Mean	4.531	4.531	4.531	4.531	4.661
Locality fixed effects	No	Yes	Yes	Yes	Yes
Home state fixed effects	No	No	Yes	Yes	Yes
Household characteristics	No	No	No	Yes	Yes
Migration controls	No	No	No	No	Yes
N	1376	1376	1376	1376	1063

Panel B. For the Mangolpuri sub-sample

	(1)	(2)	(3)	(4)
Property title	0.308 (2.373)	0.033 (2.331)	0.774 (2.353)	-2.087 (2.705)
Outcome Mean	1.777	1.777	1.777	2.033
Home state fixed effects	No	Yes	Yes	Yes
Household characteristics	No	No	Yes	Yes
Migration controls	No	No	No	Yes
N	421	421	421	276

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

Notes: This table presents Tobit regression coefficients from estimating Equation 3.2 for the log of housing investment in the preceding five years. Panel A contains results for the full sample of households. The locality fixed effects specification in Column 2 adds indicator variables for Mangolpuri, Narela and Savda Ghevra. The home-state fixed effects specification in Column 3 adds indicator variables for the households' home state. In Column 4, I add controls for household characteristics, which include household size at the time of moving to the current residence, as well as age, age squared and gender of household head. In Column 5, I add controls for migration-related variables, which include reason for migrating to Delhi and the number of years the household had been in Delhi when it moved to the current residence. Column 5 has fewer observations because 19% of the sample are originally from Delhi and therefore no migration information is available for them. Panel B contains results for the sub-sample of households who live in Mangolpuri. The columns contain almost the same specifications as in Panel A, except for the locality fixed effects specification. Robust standard errors are reported below the coefficients.

CONCLUDING REMARKS

DISCUSSION AND AVENUES FOR FUTURE RESEARCH

This thesis has examined the role of the state in shaping economic outcomes in a rapidly urbanising country, India. The chapters that constitute this thesis investigate both the relationship between urban and rural areas across the country, as well as the effects of specific policy tools within particular cities. In doing so, these chapters contribute new results to several strands of the economics literature, such as structural transformation, the sources of spatial inequality, urbanisation, decentralisation, governance, the effects of property titling, and the importance of transport infrastructure in developing countries.

In Chapter 1, with my co-authors, I have shown the importance of geography in constraining state effectiveness in rural India. Using rich village-level data from the universe of Indian villages, and a spatial regression discontinuity design, we show that villages that are more distant from their administrative headquarters receive fewer public goods and services, and have worse economic outcomes, as compared to villages that are located in close geographic proximity, but are closer to their administrative headquarters. We further show that, at least for paved feeder roads, these differences in access to public goods are driven not by the cost of constructing the infrastructure, but by the cost of monitoring its quality once it has been constructed.

These results raise two important questions that should be the subject of future research. First, when faced with poor delivery of important public goods, what actions do citizens in administratively remote villages take to demand public goods from the state? Unfortunately, there are no systematic data that record citizens' actions in re-

sponse to the under-provision of public goods. Developing and analysing such data is an important area of future research. Second, what policies should the state itself follow to reduce the costs of administrative remoteness? One option could be to create smaller administrative units that reduce the physical distance between the citizens and the state. But this can impose high fiscal costs. Another option could be to change the incentives of frontline bureaucrats so that they build new infrastructure, and regularly monitor its quality, in administratively remote villages. Providing high-quality and frequent data on public service delivery to the district administration can also be beneficial. Governments and researchers will have to experiment with some of these policy options to understand how to mitigate the long-run costs of administrative remoteness in India and in other developing countries.

In addition to serving as administrative headquarters, Indian towns and cities are also important labour markets. Turning to this function, Chapter 2 investigates the rural effects of urban employment growth. In this chapter, my co-authors and I estimate the change in rural employment growth when there is an exogenous increase in manufacturing employment in nearby towns and cities. The exogenous shocks to urban employment are constructed using a Bartik (or shift-share) instrument for the urban manufacturing sector.

We find that an exogenous increase in the employment growth rate in the manufacturing sector in the nearest town *reduces* the growth rate of manufacturing employment in the village. But it has no effect on the growth rate of rural service sector employment, or on the growth rate of rural non-farm employment in general. We argue that these results favour the hypothesis that high returns to agglomeration in towns cause manufacturing activity to relocate away from villages, driving labour exit from the rural manufacturing sector.

We further find that the negative effects of urban manufacturing employment growth are more pronounced for villages located close to the larger towns and cities in our sample, which is precisely where we expect the agglomeration forces to be the strongest.

The effects are also more pronounced for villages that are ill-suited for manufacturing activities in the first place, such as villages without paved roads and villages that are located farthest from their nearest town or city.

This finding raises the bigger question of what happens to the rural economy when neighbouring towns start growing faster. Does the rural economy suffer from the exodus of manufacturing workers, or do the workers continue to live in the village and increase local demand for non-farm goods and services? The data indicate that urban employment growth has no statistically significant effect on the village population growth rate.¹ This suggests that manufacturing workers do not migrate out of the village permanently, but start commuting to the nearest town for work. This should allow some gains from urban growth to be absorbed into the rural economy. To what extent this happens should be the subject of future research.

Finally, in Chapter 3, I turn my attention to the effect of more secure property titles in Delhi. Using data from three resettlement localities, I show that more secure property titles are not associated with higher rates of investment in improving housing quality, with the amount of investment in improving housing quality, or with housing quality itself. I show that the results are robust to controlling for a number of exogenous households characteristics, and to using propensity score matching to address selection on observables. Considering the possibility of selection on unobserved variables, I find that the degree of selection on unobserved variables, relative to the degree of selection on observed variables, has to be very high for my observed null effect to mask a true non-zero effect of property titles on housing investment outcomes. I argue that property titles do not seem to matter in this context because all households have fairly low levels of access to the formal financial sector, because the perceived risk of eviction for titled and non-titled households are likely to be similar, and because the titles do not confer any transfer rights to the titled households.

While these results should temper the enthusiasm of governments and economists

¹However, this finding is constrained by the absence of contemporaneous data on employment and population in India.

for rolling out large property titling programs, the results are constrained by the observational nature of the study. Without random assignment of property titles, it is difficult to conclusively establish the effectiveness of property titles in driving households' investment behaviour. If governments in India and elsewhere are sincerely interested in using property titles to improve the welfare of the urban poor, they should invest in more experimental work to investigate the effects of property titles on relevant household outcomes. Such experimental research should also be used to infer the best set of complementary inputs that can maximise the benefits from urban property titles.

BIBLIOGRAPHY

- Abadie, Alberto and Guido W Imbens**, “Bias-corrected matching estimators for average treatment effects,” *Journal of Business & Economic Statistics*, 2011, 29 (1), 1–11.
- , **Susan Athey, Guido Imbens, and Jeffrey Wooldridge**, “When Should You Adjust Standard Errors for Clustering?,” *arXiv preprint arXiv:1710.02926*, 2017.
- Adamopoulos, Tasso**, “Transportation Costs, Agricultural Productivity, And Cross-Country Income Differences,” *International Economic Review*, 2011, 52 (2), 489–521.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg**, “Fractionalization,” *Journal of Economic Growth*, 2003, 8 (2), 155–194.
- Allen, Treb**, “Information frictions in trade,” *Econometrica*, 2014, 82 (6), 2041–2083.
- Alpa, Shah, Jens Lerche, Richard Axelby, Delel Benbabaali, Brendan Donegan, Jayaseelan Raj, and Vikramaditya Thakur**, *Ground Down by Growth. Tribe, Caste, Class, and Inequality in Twenty-First Century India*, Pluto Press and Oxford University Press (India), 2018.
- Alston, Lee J, Gary D Libecap, and Bernardo Mueller**, *Titles, Conflict, and Land Use: The development of property rights and land reform on the Brazilian Amazon frontier*, University of Michigan Press, 1999.
- Altonji, Joseph G, Todd E Elder, and Christopher R Taber**, “Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools,” *Journal of political economy*, 2005, 113 (1), 151–184.
- Anderson, Michael L**, “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*, 2008, 103 (484), 1481–1495.
- Anderson, Siwan, Patrick Francois, and Ashok Kotwal**, “Clientelism in Indian villages,” *The American Economic Review*, 2015, 105 (6), 1780–1816.
- Asher, Sam and Paul Novosad**, “Rural Roads and Local Economic Development,” 2018.
- Ashraf, Nava, Oriana Bandiera, and B Kelsey Jack**, “No margin, no mission? A field experiment on incentives for public service delivery,” *Journal of Public Economics*, 2014, 120, 1–17.

- Atkin, David and Dave Donaldson**, "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," Working Paper, National Bureau of Economic Research 2015.
- Baharoglu, Deniz**, "World Bank experience in land management and the debate on tenure security," *World Bank Housing Research Background–Land Management Paper*, 2008.
- Banerjee, A, R Hanna, and S Mullainathan**, "Corruption, the Handbook of Organizational Economics," 2013.
- Banerjee, Abhijit and Lakshmi Iyer**, "History, institutions, and economic performance: the legacy of colonial land tenure systems in India," *The American Economic Review*, 2005, 95 (4), 1190–1213.
- , —, and **Rohini Somanathan**, "History, social divisions, and public goods in rural India," *Journal of the European Economic Association*, 2005, 3 (2-3), 639–647.
- , —, and —, "Public action for public goods," *Handbook of Development Economics*, 2007, 4, 3117–3154.
- Barankay, Iwan and Ben Lockwood**, "Decentralization and the productive efficiency of government: Evidence from Swiss cantons," *Journal of Public Economics*, 2007, 91 (5), 1197–1218.
- Bardhan, Pranab**, "The Political Economy of Development in India," 1984.
- , "Decentralization of governance and development," *The Journal of Economic Perspectives*, 2002, 16 (4), 185–205.
- Barnhardt, Sharon, Erica Field, and Rohini Pande**, "Moving to opportunity or isolation? network effects of a randomized housing lottery in urban india," *American Economic Journal: Applied Economics*, 2017, 9 (1), 1–32.
- Bartik, Timothy J**, "Who Benefits from State and Local Economic Development Policies?," 1991.
- Bellemare, Marc F, Christopher B Barrett, and David R Just**, "The welfare impacts of commodity price volatility: evidence from rural Ethiopia," *American Journal of Agricultural Economics*, 2013, 95 (4), 877–899.
- Besley, Timothy**, "Property rights and investment incentives: Theory and evidence from Ghana," *Journal of Political Economy*, 1995, 103 (5), 903–937.
- and **Maitreesh Ghatak**, "Property rights and economic development," in "Handbook of Development Economics," Vol. 5, Elsevier, 2010, pp. 4525–4595.
- Bhan, Gautam**, "“This is no longer the city I once knew.” Evictions, the urban poor and the right to the city in millennial Delhi," *Environment and Urbanization*, 2009, 21 (1), 127–142.

-
- and **Swathi Shivanand**, “Settling the city. Analysing displacement in Delhi from 1990 to 2007,” *Economic and Political Weekly*, 2013, 48 (13), 54–61.
- Blanchard, Olivier Jean, Lawrence F Katz, Robert E Hall, and Barry Eichengreen**, “Regional evolutions,” *Brookings Papers on Economic Activity*, 1992, 1992 (1), 1–75.
- Brueckner, Jan K and Somik V Lall**, “Cities in developing countries: fueled by rural–urban migration, lacking in tenure security, and short of affordable housing,” in “Handbook of regional and urban economics,” Vol. 5, Elsevier, 2015, pp. 1399–1455.
- Burgess, Robin, Francisco Costa, and Benjamin Olken**, “The Power of the State: National Borders and the Deforestation of the Amazon,” 2016.
- , **Remi Jedwab, Edward Miguel, Ameet Morjaria et al.**, “The value of democracy: evidence from road building in Kenya,” *The American Economic Review*, 2015, 105 (6), 1817–1851.
- Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli**, “Agricultural productivity and structural transformation: Evidence from Brazil,” *American Economic Review*, 2016, 106 (6), 1320–65.
- Cali, Massimiliano and Carlo Menon**, “Does urbanization affect rural poverty? Evidence from Indian districts,” *The World Bank Economic Review*, 2012, p. lhs019.
- Callen, Michael Joseph, Saad Gulzar, Syed Ali Hasanain, and Muhammad Yasir Khan**, “The political economy of public employee absence: Experimental evidence from Pakistan,” 2013.
- Caselli, Francesco**, “Accounting for cross-country income differences,” *Handbook of Economic Growth*, 2005, 1, 679–741.
- Chattopadhyay, Raghavendra and Esther Duflo**, “Women as policy makers: Evidence from a randomized policy experiment in India,” *Econometrica*, 2004, 72 (5), 1409–1443.
- Collin, Matthew, Justin Sandefur, and Andrew Zeitlin**, “Falling off the map: The impact of formalizing (some) informal settlements in Tanzania,” Technical Report, Centre for the Study of African Economies, University of Oxford 2015.
- de Soto, Hernando**, “The mystery of capital: Why capitalism triumphs in the West and fails elsewhere,” 2000.
- Deininger, Klaus and Hans Binswanger**, “The evolution of the World Bank’s land policy: principles, experience, and future challenges,” *The World Bank Research Observer*, 1999, 14 (2), 247–276.
- Dell, Melissa**, “The persistent effects of Peru’s mining mita,” *Econometrica*, 2010, 78 (6), 1863–1903.

-
- and **Benjamin A Olken**, “The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java,” 2017.
- , **Nathaniel Lane**, and **Pablo Querubin**, “The Historical State, Local Collective Action, and Economic Development in Vietnam,” Technical Report, National Bureau of Economic Research 2017.
- Diamond, Rebecca**, “The determinants and welfare implications of US workers’ diverging location choices by skill: 1980–2000,” *The American Economic Review*, 2016, 106 (3), 479–524.
- Duflo, Esther**, “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment,” *The American Economic Review*, 2001, 91 (4), 795–813.
- Dupont, Véronique**, “Slum demolitions in Delhi since the 1990s: an appraisal,” *Economic and Political Weekly*, 2008, pp. 79–87.
- Durand-Lasserve, Alain**, **Edesio Fernandes**, **Geoffrey Payne**, and **Carole Rakodi**, “Social and economic impacts of land titling programmes in urban and peri-urban areas: a review of the literature,” in “World Bank Urban Research Symposium, Washington DC, May” 2007.
- Fafchamps, Marcel** and **Forhad Shilpi**, “The spatial division of labour in Nepal,” *The Journal of Development Studies*, 2003, 39 (6), 23–66.
- and ———, “Cities and specialisation: evidence from South Asia,” *The Economic Journal*, 2005, 115 (503), 477–504.
- and **Jackline Wahba**, “Child labor, urban proximity, and household composition,” *Journal of Development Economics*, 2006, 79 (2), 374–397.
- , **Michael Koelle**, and **Forhad Shilpi**, “Gold mining and proto-urbanization: recent evidence from Ghana,” *Journal of Economic Geography*, 2016, p. lbw015.
- Faguet, Jean-Paul**, “Does decentralization increase government responsiveness to local needs?: Evidence from Bolivia,” *Journal of Public Economics*, 2004, 88 (3), 867–893.
- Feyrer, James**, “Distance, trade, and income—the 1967 to 1975 closing of the Suez Canal as a natural experiment,” Technical Report, National Bureau of Economic Research 2009.
- Field, Erica**, “Fertility responses to urban land titling programs: The roles of ownership security and the distribution of household assets,” Technical Report, Mimeo, Harvard University 2003.
- , “Property rights and investment in urban slums,” *Journal of the European Economic Association*, 2005, 3 (2-3), 279–290.

-
- , “Entitled to work: Urban property rights and labor supply in Peru,” *The Quarterly Journal of Economics*, 2007, 122 (4), 1561–1602.
- **and Maximo Torero**, “Do property titles increase credit access among the urban poor? Evidence from a nationwide titling program,” *Department of Economics, Harvard University, Cambridge, MA*, 2006.
- Finan, Frederico, Benjamin A Olken, and Rohini Pande**, “The personnel economics of the developing state,” in “Handbook of Economic Field Experiments,” Vol. 2, Elsevier, 2017, pp. 467–514.
- Foster, Andrew D and Mark R Rosenzweig**, “Economic development and the decline of agricultural employment,” *Handbook of Development Economics*, 2007, 4, 3051–3083.
- Galiani, Sebastian and Ernesto Schargrodsky**, “Property rights for the poor: Effects of land titling,” *Journal of Public Economics*, 2010, 94 (9), 700–729.
- **and** ———, “The deregularization of land titles,” *Man and the Economy*, 2016, 3 (2), 169–188.
- Gelman, Andrew and Guido Imbens**, “Why high-order polynomials should not be used in regression discontinuity designs,” *Journal of Business & Economic Statistics*, 2017.
- Glaeser, Edward, Abha Joshi-Ghani et al.**, “The Urban Imperative,” *World Bank Publications*, 2015.
- **and J Vernon Henderson**, “Urban economics for the developing World: An introduction,” *Journal of Urban Economics*, 2017, 98, 1–5.
- Glaeser, Edward L**, “A world of cities: the causes and consequences of urbanization in poorer countries,” *Journal of the European Economic Association*, 2014, 12 (5), 1154–1199.
- Gollin, Douglas and Richard Rogerson**, “Productivity, transport costs and subsistence agriculture,” *Journal of Development Economics*, 2014, 107, 38–48.
- , **David Lagakos, and Michael E Waugh**, “Agricultural productivity differences across countries,” *The American Economic Review*, 2014, 104 (5), 165–170.
- , **Remi Jedwab, and Dietrich Vollrath**, “Urbanization with and without Industrialization,” *Journal of Economic Growth*, 2016, 21 (1), 35–70.
- Government of National Capital Territory of Delhi**, “Delhi Human Development Report,” Technical Report 2006.
- Green, Elliott**, “Patronage, district creation, and reform in Uganda,” *Studies in comparative international development*, 2010, 45 (1), 83–103.

- Grossman, Guy and Janet I Lewis**, "Administrative unit proliferation," *American Political Science Review*, 2014, 108 (1), 196–217.
- , **Jan H Pierskalla**, and **Emma Boswell Dean**, "Government fragmentation and public goods provision," *The Journal of Politics*, 2017, 79 (3), 823–840.
- Gunning, Jan Willem, Pramila Krishnan, Andualem Telaye Mengistu et al.**, "Fading Choice: Transport Costs and Variety in Consumer Goods," Technical Report, CEPR Discussion Papers 2018.
- Gutierrez, Italo A and Oswaldo Molina**, "Reverting to Informality: Unregistered Property Transactions and the Erosion of the Titling Reform in Peru," 2016.
- Hanna, Rema and Shing-Yi Wang**, "Dishonesty and selection into public service: Evidence from India," *American Economic Journal: Economic Policy*, 2017, 9 (3), 262–290.
- Hassan, Mai**, "A state of change: district creation in Kenya after the beginning of multi-party elections," *Political Research Quarterly*, 2016, 69 (3), 510–521.
- Henderson, J Vernon, Adam Storeygard, and Uwe Deichmann**, "Has climate change driven urbanization in Africa?," *Journal of Development Economics*, 2017, 124, 60–82.
- Iyer, Lakshmi**, "Direct versus indirect colonial rule in India: Long-term consequences," *The Review of Economics and Statistics*, 2010, 92 (4), 693–713.
- Jann, Ben**, "KMATC: Stata module for multivariate-distance and propensity-score matching," *Statistical Software Components, Boston College Department of Economics*, 2017.
- Jedwab, Remi, Luc Christiaensen, and Marina Gindelsky**, "Demography, urbanization and development: Rural push, urban pull and urban push?," *Journal of Urban Economics*, 2017, 98, 6–16.
- Kis-Katos, Krisztina and Bambang Suharnoko Sjahrir**, "The impact of fiscal and political decentralization on local public investment in Indonesia," *Journal of Comparative Economics*, 2017, 45 (2), 344–365.
- Kochar, Anjini**, "Urban influences on rural schooling in India," *Journal of Development Economics*, 2004, 74 (1), 113–136.
- Kone, Zovanga, Maggie Liu, Aaditya Mattoo, Caglar Ozden, and Siddharth Sharma**, "Internal Borders and Migration in India," 2017.
- Kothari, Rajni**, *Politics in India*, Orient Blackswan, 1970.
- Krishna, Anirudh and Gregory Schober**, "The gradient of governance: distance and disengagement in Indian villages," *Journal of Development Studies*, 2014, 50 (6), 820–838.

- Kuznets, Simon**, "Quantitative aspects of the economic growth of nations: II. industrial distribution of national product and labor force," *Economic development and cultural change*, 1957, 5 (S4), 1–111.
- Lewis, W Arthur**, "Economic development with unlimited supplies of labour," *The Manchester School*, 1954, 22 (2), 139–191.
- Mansuri, Ghazala and Vijayendra Rao**, *Localizing development: does participation work?*, World Bank Publications, 2013.
- Marx, Benjamin, Thomas Stoker, and Tavneet Suri**, "The economics of slums in the developing world," *The Journal of Economic Perspectives*, 2013, 27 (4), 187–210.
- Mehra, Rakhi, Marco Ferrario, Mukta Naik, and Aden Van Noppen**, "Enabling Self-Construction: Safe and Affordable Housing in India," Technical Report 2012.
- Michaels, Guy, Ferdinand Rauch, and Stephen J Redding**, "Urbanization and structural transformation," *The Quarterly Journal of Economics*, 2012, 127 (2), 535–586.
- Molina, Oswaldo, Mans Söderbom et al.**, "Heterogeneous Effects of Property Rights on Housing Investment in Urban Peru," Technical Report 2018.
- Mookherjee, Dilip**, "Political decentralization," *Annual Review of Economics*, 2015, 7 (1), 231–249.
- Moretti, Enrico and Per Thulin**, "Local multipliers and human capital in the United States and Sweden," *Industrial and Corporate Change*, 2013, 22 (1), 339–362.
- Munshi, Kaivan and Mark Rosenzweig**, "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *The American Economic Review*, 2016, 106 (1), 46–98.
- Muralidharan, Karthik and Venkatesh Sundararaman**, "Teacher performance pay: Experimental evidence from India," *Journal of Political Economy*, 2011, 119 (1), 39–77.
- North, Douglas C**, "Structure and Change in Economic History," *Institutions, Institutional Change, and Economic Performance*, 1981.
- Nunn, Nathan and Diego Puga**, "Ruggedness: The blessing of bad geography in Africa," *Review of Economics and Statistics*, 2012, 94 (1), 20–36.
- and **Nancy Qian**, "The potato's contribution to population and urbanization: evidence from a historical experiment," *The Quarterly Journal of Economics*, 2011, 126 (2), 593–650.
- Oates, Wallace E**, "Fiscal Federalism," *Harcourt Brace Jovanovich*, 1972.
- Olken, Benjamin A**, "Monitoring corruption: evidence from a field experiment in Indonesia," *Journal of Political Economy*, 2007, 115 (2), 200–249.

- Oster, Emily**, "Psacalc: Stata module to calculate treatment effects or bounds under proportional selection of observables and unobservables," *Statistical Software Components*, Boston College Department of Economics, 2013.
- , "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 2017, pp. 1–18.
- Pierskalla, Jan H**, "Splitting the difference? The politics of district creation in Indonesia," *Comparative Politics*, 2016, 48 (2), 249–268.
- Porta, Rafael La and Andrei Shleifer**, "Informality and development," *Journal of Economic Perspectives*, 2014, 28 (3), 109–26.
- PRIndex Analytical Report**, "PRIndex Analytical Report," Technical Report 2017.
- Rasul, Imran and Daniel Rogger**, "Management of bureaucrats and public service delivery: Evidence from the Nigerian civil service," *The Economic Journal*, forthcoming.
- Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu**, "Agriculture and aggregate productivity: A quantitative cross-country analysis," *Journal of Monetary Economics*, 2008, 55 (2), 234–250.
- Richards, Peter and Leah VanWey**, "Where deforestation leads to urbanization: How resource extraction is leading to urban growth in the Brazilian Amazon," *Annals of the Association of American Geographers*, 2015, 105 (4), 806–823.
- Rosenbaum, Paul R and Donald B Rubin**, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 1983, 70 (1), 41–55.
- Ryan, Nicholas Nicholas James**, "Empirical essays on firm behavior in India." PhD dissertation, Massachusetts Institute of Technology 2012.
- Sebastian, Gertler Paul Galiani and Ernesto Schargrodsky**, "School decentralization: Helping the good get better, but leaving the poor behind," *Journal of Public Economics*, 2008, 92 (10), 2106–2120.
- Second Administrative Reforms Commission**, "Fifteenth report: State and District Administration," Technical Report, Second Administrative Reforms Commission 2009.
- Smith, Adam**, "An Inquiry into the Nature and Causes of the Wealth of Nations," in "in," London: printed for W. Strahan and T. Cadell, 1776, 1776.
- Storeygard, Adam**, "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa," *The Review of Economic Studies*, 2016, p. rdw020.
- Tella, Rafael Di, Sebastian Galiant, and Ernesto Schargrodsky**, "The formation of beliefs: evidence from the allocation of land titles to squatters," *The Quarterly Journal of Economics*, 2007, 122 (1), 209–241.

The Hindu, "What is the special category status?," August 2016.

The World Bank, "The Rural Access Index," Technical Report, World Bank 2015.

Tobin, James, "Estimation of relationships for limited dependent variables," *Econometrica*, 1958, pp. 24–36.

UN Habitat, *State of the world's cities 2012/2013: Prosperity of cities*, Routledge, 2013.