Roads and Loans

Sumit AGARWAL, Abhiroop MUKHERJEE and S. Lakshmi NAARAAYANAN

HKUST IEMS Working Paper No. 2019-61

May 2019

Revised January 2021

Roads and Loans

Sumit Agarwal, Abhiroop Mukherjee and S. Lakshmi Naaraayanan

HKUST IEMS Working Paper No. 2019-61

Abstract

Does financing respond to changes in productive opportunities, even for the world's poor? We answer this question by examining the response of private bank financing to a rural infrastructure program in India, which brought road access to unconnected villages. The program prioritized roads for villages above specific population thresholds, allowing us to exploit resulting discontinuities for identification. We find large financing responses to roads using detailed data from a large bank – 75% more villagers get loans, and the average amount lent to them is about 30-35% higher– for villages just above the threshold compared to those below.

JEL classification: G21, O16

Keywords: Finance and growth, Infrastructure development and finance, Rural lending

Author's contact information

Sumit Agarwal
National University of Singapore
E: bizagarw@nus.edu.sg

Abhiroop Mukherjee
Institute for Emerging Market Studies, The Hong Kong University of Science and Technology
E: amukherjee@ust.hk

S Lakshmi Naaraayanan
London Business School
E: sln@connect.ust.hk

We are grateful to Nicholas Barberis, Utpal Bhattacharya, Chiman Cheung, Vidhi Chhaochharia, Darwin Choi, James Choi, Lauren Cohen, Sudipto Dasgupta, Pengjie Gao, Pulak Ghosh, Radha Gopalan, John Griffin, Rawley Heimer, Rustom Irani, Yan Ji, Yatang Lin, Hanno Lustig, Christopher Malloy, Kasper Nielsen, Wenlan Qian, Shivaram Rajgopal, Rik Sen, Amit Seru, Manpreet Singh, Philip Strahan, Paula Suh, Mingzhu Tai, Prasanna Tantri, Sheridan Titman, Vikrant Vig, Sujata Visaria, Baolian Wang, Daniel Wolfenzon, and conference/seminar participants at the ABFER, Bocconi-RFS New Frontiers in Banking Conference, China International Conference in Finance, Deakin University, ISB Summer Research Conference, Hong Kong Baptist University, Hong Kong University of Science & Technology, Hong Kong Polytechnic University, University of New South Wales, and SFS Cavalcade Asia for helpful comments. Agarwal and Mukherjee gratefully acknowledge financial support from the General Research Fund of the Research Grants Council of Hong Kong (Project Number: 16505617). Naaraayanan thanks Columbia University for hosting him for a part of the time during which this research was conducted.

1 Introduction

One of the first principles of finance is that capital should flow to its most productive uses. So when productive opportunities improve, financing should flow to those who see these gains, allowing them to fully realize potential benefits. Naturally, then, a very large literature is devoted to understanding the response of financing to productivity changes (starting with Schumpeter, 1911; more recently King and Levine, 1993a, 1993b); and this literature has indeed been helpful in influencing many important economic policies of our time. But does private, profit-motivated financing really respond to productivity changes – not just for large corporations or rich households in countries with developed markets, but also in the day-to-day lives of the rural poor?

This is a very important issue because about half of world population – nearly 3.5 billion people – still live in rural areas, often characterized by poverty. As Levine (2008) points out,

> *"...the operation of the formal financial system is profoundly important for the poor. It influences how many people are hungry, homeless and in pain. It shapes the gap between the rich and the poor. It arbitrates who can start a business and who cannot, who can pay for education and who cannot, who can attempt to realize one's dreams and who cannot."*

But whether banks will respond to and facilitate productivity improvements for the rural poor is far from obvious. Private profit-motivated banks have typically only begun lending in rural areas recently, and face various social, political and economic impediments in this new setting, leading to much skepticism about their efficacy (e.g., Basu, 2006).

We shed light on this issue by examining a shock to productive opportunities arising from a large rural road-building initiative in India. Program rules allow us to exploit populationbased discontinuities in road construction to identify lending effects using a novel, proprietary loan-level dataset. We find that the odds of villager getting a loan is nearly twice as high, and the average amount lent to them is about 30–35% higher – for villages right above the thresholds used for road construction, compared to those just below.

The road-building program we study is among many such infrastructure projects being undertaken in various parts of the world – projects that are thought to be key to unlocking productivity increases among growing populations of surplus rural labor. Hundreds of thousands of miles of such roads have been built in Asia, Africa, South America, and Eastern Europe in the past two decades – India alone built 1.96 million kilometers of rural roads between 2000-2016. But the type of productive opportunities policy makers often talk about as examples of trickle down benefits – for example, opening or expansion of village grocery shops, or changing crop patterns from subsistence cereal farming to more profitable marketbased crops – very often require availability of financing (e.g., King and Levine, 1993b, Aghion and Bolton, 1997, Levine, 1997). It is typically assumed in much of the policy discourse that such financing to households will automatically follow once roads are built. But this assumption sits in stark contrast to a substantial literature pointing out inefficiencies in rural financial markets (e.g., see Conning and Udry, 2005, or more recently, Agarwal et al. 2017, for further references).

This literature points out, for example, that state-subsidized financial institutions – main lending sources in rural areas, where present – have had significant difficulties in terms of both outreach and profitability. Some of these problems are endemic – e.g., a lack of political will to allow independent operation of rural financial institutions – leading many to question their capacity to effectively meet rural credit demand (e.g., Coffey, 1998, Satish, 2004).[1] Moreover, governments financing some of these infrastructure projects face tight budget constraints, and are often under heavy debt; this is especially true in poorer countries. Given the typical loss-making nature of state lending in rural areas, these governments often cannot afford to simultaneously finance infrastructure, as well as provide loan financing through state-owned banks to realize its productivity benefits.

But is there a way for profit-motivated private banks to lend a helping hand? There has been increasing interest among private banks across the developing world in rural banking recently.[2] Rural banking divisions of private banks are, however, typically small, and mostly these banks are just starting to operate in these greenfield markets. This motivates our main question: could these private sector financiers respond to changing productive opportunities in rural areas in the way policy makers expect them to?

[1] As an example of the acuteness of problems, state-owned banks in many countries – that have had rural operations for decades – are known to have hired management consulting firms to advise them on how to do rural banking better, as recently as in 2010 (e.g., see "SBI pays Rs 62.8 cr to McKinsey to learn rural banking").

[2] See, e.g., media and industry articles on private banking interest in rural areas. In India: "New private banks upbeat on rural expansion", livemint.com, 2013; China: *Foreign Banks in China*, PriceWaterhouse Coopers, 2010 (e.g., page 5); Tanzania: "JPM promises to support banks' rural expansion", 2018, etc.

Moreover, even if financing does follow infrastructure improvements, does it disproportionately benefit the relatively rich villagers who had assets prior to the infrastructure being built, and were therefore in a better position to exploit the resultant opportunities? Or does it benefit the poorer parts of society more – people who were excluded from formal finance before, but can now find a way in (Beck, Demirguc-Kunt, and Levine, 2007)?

One reason why these questions have not already been answered is the difficulty researchers face in accurately identifying the causal impact of new infrastructure. This difficulty arises because it is hard to identify an appropriate counterfactual or comparison group. Although one can observe what happens before and after new infrastructure is constructed in 'treated' areas, it is hard to attribute the change exclusively to the project and not to any other environmental or policy factors that may also have been changing at the same time. If infrastructure were located randomly, a natural comparison group would be locations that did not (randomly) receive infrastructure, allowing us to assess program impact. Infrastructure, of course, is not placed randomly in practice, making comparisons with untreated areas problematic.

We find a way to progress by exploiting a policy directive surrounding a major public road construction program in India. The objective of this nationwide program – called the *Pradhan Mantri Gram Sadak Yojna*, (henceforth PMGSY) – was to provide all-weather road connectivity to hitherto unconnected villages. The roads program we study created a nearly random comparison group for policy evaluation, by explicitly focusing on building new roads to connect all villages above explicit population thresholds. By doing so, program rules created discontinuities in the probability of treatment at these village population thresholds, which we exploit to identify our effects. For example, villages with populations just above a round figure, say 500, were to be prioritized under the program. Under the assumption that villages with populations just below the threshold are very similar to those above, especially if they are located in close geographic proximity, the resultant variation in roads is quasirandom. Asher and Novosad (2018) have shown that these thresholds indeed predict actual road construction using data from six Indian states; we verify that this is also true in our sample, which comes from the states of *Odisha* in the east and *Uttarakhand* in the north of the country. Above-threshold villages are 55% more likely to have received a road in our sample, relative to those below.

Our empirical analysis is made possible by our access to a unique, proprietary loan-level dataset from one of India's largest private lenders. We begin our analysis by examining the effect of population thresholds on the external margin of lending. The odds that our bank lends in a village is twice as high if that village has population above the cutoff, relative to one below. Net loan disbursement as a proportion of income also shows a significant jump of 31–36% at the cutoff, even after controlling for the entire set of borrower characteristics that the bank cares about and collects information on. Other loan characteristics, however, do not vary at the cutoff: loans in connected villages are similar to those in unconnected ones in terms of default probability, maturity, and interest rates. These results are robust; they also do not show up in a placebo test looking at villages around the same cutoffs, but those that were all connected more than a decade ago under a different program that did not use population-based cutoffs.

While our setting buys us an advantage in terms of identification, we face two difficulties. First, to keep our treatment (above-cutoff) and control (below-cutoff) sample comparable, we need to restrict our bank lending data to villages with populations close to the cutoffs. Once we do this, we are left with 48 (58) villages with a cut-off of 200 (250) in the bank lending sample, which does not allow us to use a full-fledged village-level regression discontinuity design involving higher order polynomials etc. Instead, it is best to think of our research design in terms of a treatment-control setup, wherein the 'new road' treatment is administered to a few villages at random – chosen depending on which side of the cutoff they were at – and the rest are controls. We do, however, have detailed within-village data, which we exploit by doing our analysis in detail at the individual borrower-level – similar to papers examining differences in individual firm-level outcomes across two groups that are exogenously subjected to different policies/regulations.

Second, the reduced form nature of our analysis makes it difficult for us to quantify the magnitudes of demand shifts (roads increase marginal productivity of capital, villagers demand more loans) versus supply-side shifts (the bank actively seeks out lending opportunities to easier-to-reach newly connected villagers or finds it easier to monitor lending in connected villages). While we are open to both explanations, we make some progress on distinguishing between them in the data through a few further tests. First, while the supply side story can explain our extensive margin results, loan *amounts* conditional on the borrower having been reached being higher above cutoff is not an obvious prediction in the ease-of-reaching-out story. Second, we uncover evidence that almost all our results come from *productive* loans, which are loans taken out for crops, micro enterprises, etc. On the other hand, loan amounts granted for consumption uses are actually *lower* in villages with populations above thresholds. This is consistent with some reallocation of credit from consumption uses to productive uses when road connectivity improves. This, again, is not consistent with a supply-side reach-the-borrower story. Third, we present a test based on the *variation* in loan contract terms (similar to Fisman, Paravisini and Vig, 2017), which also shows no support for a soft information/ monitoring-based supply side explanation. The weight of our evidence from these tests, therefore, tilts more towards a demand-based explanation. Of course, demand and supply effects are not mutually exclusive -– both could be at play here. Ultimately, whether or not *equilibrium* financing responds to new rural roads is important for the lives and livelihoods of millions of poor villagers, and hence, for policy – even if we cannot conclusively pin down the relative sizes of shifts in demand and supply curves.

Next, we examine the distributional consequences of connectivity from our lending sample. Our data allows us to focus on individual-level differences. This is a critical step in understanding the trickle-down effects of development, as well as for the financial inclusion and inequality literature (e.g., Aghion and Bolton, 1997, Beck and Demirguc-Kunt, 2008, Demirguc-Kunt and Levine, 2009, Beck, 2012). We find that villagers with less assets benefit more. This is consistent with the view that productivity shocks release collateral constraints, and improve financial inclusion for those with a lack of traditional collateralizable assets (Agarwal et al., 2017).

In the last section of the paper, we address the macro implications of our findings using data from beyond our bank loan sample. Unfortunately, we lack detailed village level lending data in this broader sample, so we cannot use population threshold-based cutoffs here. Instead, we use Reserve Bank of India (RBI) data on overall private lending activity by sector (e.g., rural, urban, etc.) aggregated at the district level for 19 Indian states. Our evidence suggests that higher lending and deposits follow rural road-building well beyond our baseline bank-loan sample. These findings are robust to controlling for many political and economic variables that might simultaneously affect financial development and economic growth, fixed effects at the district-level, as well as state-year fixed effects. While increases in rural lending follow rural road-building in a district, there is little impact on urban lending within the same district, as one might expect.

Finally, we find a significant association between rural road-building and output growth, but only in regions with better rural credit markets. In these regions, rural roads are followed by higher district-level GDP growth rates, particularly in the agricultural sector. Growth effects are not statistically distinguishable from zero in areas with less developed rural financial markets. Roads and loans, therefore, seem to be complements in the growth process.

Our paper contributes to the growing literature on the role of financing in economic development and poverty alleviation (King and Levine, 1993, Black and Strahan, 2002, Burgess and Pande, 2004, Levine, 2005, Beck, Demirguc-Kunt, and Levine, 2007, Beck, DemirgucKunt and Peria, 2007, Demirguc-Kunt and Levine, 2008a, 2008b, Visaria, 2009, Allen et al., 2011, Beck, 2012, Demirguc-Kunt, Feyen, and Levine, 2013, Vig, 2013, Beck, Lu and Yang, 2014, among others). Many of these papers have established important results on the effects of financing on economic growth and development. In the context of financing in a rural setting in India, our paper is related to Burgess and Pande (2004) and Agarwal et al. (2017), who both examine the effect of government-led expansion of credit and savings facilities. Our paper, different from these papers, does not examine the causal effect of financing policy on other outcomes; instead, we look at whether formal, private-sector financing *responds* to improved connectivity in poor rural areas.

This focus on poor households is the key distinction between this paper and a contemporaneous study by Das et al. (2017), who examine aggregate industrial financing by district around a different road upgradation program in India. The other distinction, of course, lies in that we focus on examining the effects of new road connectivity, rather than upgradation, by focusing on a discontinuity design based on PMGSY. The PMGSY program has also been used by Asher and Novosad (2017, 2018), who show that new roads led to a reallocation of village labor from agriculture to wage labor, and by Mukherjee (2011) and Adukia et al (2018) to examine schooling decisions. Shamdasani (2016) and Aggarwal (2018) have also examined the effects of this program on rural households, and find evidence of improvements in productivity for affected villages through the use of yield-improving fertilizer and hybrid seeds on farms, as well as transitions from subsistence to market-oriented farming. Our paper is different from these studies on at least two counts: first, our outcome of interest is very different – we examine financing responses to productivity shocks, which is an important issue in finance, but missing from these other papers. Second, our unique individual-loan-level dataset allows us to study who benefits from such financing

– which is an independently important question from an inequality and development policy point of view. Third, our results qualify those in Asher and Novosad (2018): while they show that in an average village – which typically lacks access to credit in rural India – roads have little impact on economic outcomes, we complement macro data from the government of India with lending data from the RBI to show that in regions with relatively better access to finance, it might be possible to see some benefits of connectivity, especially in the agriculture sector.

Our evidence also adds to extant literature estimating the effects of public infrastructure in low- and middle-income countries. This literature generally finds economically meaningful effects of such projects on a wide range of outcomes. Specifically, transportation infrastructure has been shown to raise the value of agricultural land (Donaldson and Hornbeck, 2015), increase agricultural trade and income (Donaldson, 2016), reduce the risk of famine (Burgess and Donaldson, 2012), increase migration (Morten and Oliveira, 2014) and accelerate urban decentralization (Baum-Snow et al., 2015). In addition, there is mixed evidence that transportation costs can increase (Ghani et al., 2016, 2017; Khanna, 2014; Storeygard, 2014), decrease (Faber, 2014) or leave unchanged (Banerjee et al., 2012) growth rates in local economic activity. Relative to these papers, our bank dataset allows us to focus on detailed rural financing outcomes.

Finally, the empirical literature has often found mixed evidence on the effects of infrastructure on inequality. In a recent survey, Calderon and Serven (2004) note that crosscountry empirical studies often find weak and suggestive evidence that infrastructure reduces inequality. Within-country studies, however, offer mixed evidence. For example, Artadi and Sala-i-Martin (2004) find that infrastructure spending may have contributed to income inequality in Africa, whereas Khandker, Bakht, and Koolwal (2009), find that the poorest households benefitted the most from road improvement projects in Bangladesh. Given these mixed results, there is a clear need for more work on identifying the impact of road construction on local inequality. Section 5 in our paper takes a modest step towards this goal.

2 Data

Our main data source is a proprietary, rural, bank-account level dataset that we obtained from one of India's largest private banks. One of the main obstacles limiting research of questions like ours is the lack of availability of granular private financing data at the individual level, particularly in the case of small villages. Our data comes from the coastal district of *Ganjam* in the eastern state of *Odisha*, and from the mountainous districts of *Tehri Garhwal*, *Uttarkashi*, *Chamoli* and *Garhwal* in the northern state of *Uttarakhand*. Note that while most of our villages within our bandwidth are in *Uttarakhand* (45 out of 58), the density of our bank's presence (in terms of total amount lent or number of borrowers) is substantially greater in *Odisha*. For example, out of 1084 villagers with whom the bank has ever had a lending relationship, 792 are from *Odisha*.[3] Our dataset contains information on individual accounts and transactions in loans over the period 2009-2014. The data also contains relatively detailed demographic information, such as the borrower's gender, education, assets, and income, as provided at the time of the bank account opening. The bank further provides asset values and the breakdown of assets on the number of dwellings owned, the type of dwelling (brick or mud), number of livestock, etc. However, the values of these sub-categories are not available.

We obtain data on road construction in India from the website of *Pradhan Mantri Gram Sadak Yojna* (PMGSY), the road-building program we study. The data includes detailed information on road sanction and completion dates, which we scrape. The PMGSY data is structured to consist of information both at the habitation-level and at the road-level. We conduct our analysis at the village-level and perform a time-intensive manual match to villages in our bank data.

To do so, we first perform a hand-match for each village from our bank lending sample to the habitations and villages receiving rural roads under the PMGSY program, and finally to the Population Census 2001. Our match finds that the smallest unit of analysis happens to be a village in 54 out of 58 cases, making village-level and habitation-level analyses identical for most of our sample. However, we find that four of the villages (all in the state of *Uttarakhand*) have habitations associated with them. For these four villages, we consider a village to be treated under PMGSY if at least one habitation in the village – which was previously

[3] Resulting power considerations limit our ability to do tests contrasting the two states.

unconnected to the paved “all-weather” road network – received a (completed) road during our sample period. As a robustness test, in panel A of Table IA8, we drop these four villages with habitations and find very similar results.

We successfully match over 85% of habitations listed on the PMGSY website to their corresponding census villages. Further, the hand-match of the administrative road data to our proprietary dataset at the village-level yields a match of 270 villages spread over 15 blocks and two states across both the datasets. We also use data on demographics and village-level amenities (such as electricity, distance to nearest town, schools, etc.) from the 2001 Population Census and the previously listed PMGSY webpage. Finally, we look at all unconnected villages in 2009 (which is the year our bank started lending in this area), which ensures that we are indeed capturing the effect of *newly constructed* rural roads.

We supplement this dataset with district-level data on GDP from Indicus Analytics, aggregate district level lending data from the Reserve Bank of India (RBI), and various district-level time-varying economic and political variables from the *International Crops Research Institute for the Semi-Arid Tropics* (ICRISAT) and the Election Commission of India (ECI).

3 PMGSY and empirical strategy

3.1 The PMGSY program

The main challenge in identifying the impact of infrastructure investments on financing – even if one had the data required to measure outcomes – is the endogenous placement of such infrastructure. Factors such as political favoritism or local economic conditions could be correlated directly with both road placement as well as the outcomes of interest, which can render OLS estimates biased (Beck, 2008). In this section we describe the empirical strategy we use to make some progress in identification.

Our identification strategy is based on guidelines set forth by a national road building program, called *Pradhan Mantri Gram Sadak Yojana* (PMGSY). This program was launched by the central government in December of 2000 to provide access to “all-weather” roads to all 74% of India’s population that still lived in villages. PMGSY proved to be one the largest rural road programs the world has ever seen, with 480,000 kilometers of rural roads built under it by 2016, doubling the size of India's rural road network.

The program mainly focused on hitherto unconnected villages, defined as those without any pre-existing all-weather road within 500 meters of its boundaries, and its aim was to construct roads to connect these villages to the closest town or market center. Program guidelines prioritized villages to receive new roads based on population. At the time most of these roads were constructed, the last nationwide official population record was from the 2001 census. The instructions required state officials to target villages in the following order: (i) villages with population greater than 1000; (ii) villages with populations greater than 500; and (iii) villages with populations greater than 250.

Our identifying assumption is therefore that even if selection into road connectivity could be determined by many factors in general, these factors are not likely to change discontinuously at these population thresholds. Hence, if these rules were followed by the officials in charge – which we can test – we can estimate the effect of road connectivity on financing outcomes using a discontinuity design. Note that throughout the paper, we will use thresholds of 500 and 1000, but not 250, because there are no villages below the 250 population threshold where our bank lends.

Papers before us, for example, Asher and Novosad (2018) have used PMGSY-based discontinuity before, and have shown its validity/strength as an instrument. However, their results were for six states of India, not just for *Odisha* or *Uttarakhand*. We show that instrument strength/validity extends to *Odisha* and *Uttarakhand*, as well as to our banklending sample in the next sections.

3.2 Empirical strategy

Our bank data comes from villages in *Odisha* and *Uttarakhand*. We first test for threshold manipulation under the PMGSY program. This is important to understand whether, for example, a powerful politician was getting local officials to systematically classify some villages with populations below the threshold as being above it, so that these villages get roads. This can be problematic for identification, since then we will not know whether any lending effect we identify in these villages that get roads is indeed attributable to the road connectivity, or to the same politician's simultaneous influence on bank lending. To make sure that our estimates are not confounded by such issues, we use population figures from 2001 census – which was conducted mostly before PMGSY policy cutoffs were finalized. While this may produce noise in estimates if the road-building authorities used more updated figures, it ensures validity.[4] Still, we check for any indication of manipulation using tests for discontinuities in the density of our running variable, population (McCrary, 2008).

In Figure 1, where we plot the histogram of villages in *Odisha* by population, we can see that there are no discrete jumps in population around the PMGSY thresholds of 500 and 1000, indicating no manipulation for these thresholds. In Figure 2, we provide a formal test of discontinuity following McCrary (2008). To provide one summary test, we first combine our thresholds of 500 and 1000 into one above-cutoff variable. We do so through a normalized measure of village population, which we create by subtracting the closest threshold from each village's population; our above-cutoff variable takes the value of 1 for villages with normalized populations just above zero.[5] The point estimate for the discontinuity at the cutoff is 0.08, with a standard error of 0.058; so we fail to reject the null hypothesis of no discontinuity in the running variable.

Second, in Figure 3, we examine the geographic clustering of above- versus below-cutoff villages in our two states, and find that the above- and below-cutoff villages come from geographically proximate areas within each state.

Further, since our identifying assumption is that crossing the population threshold discontinuously affects the probability of receiving a road under PMGSY – but not other things at the village level – there should be no jumps in other village characteristics (baseline covariates) at the population thresholds (Imbens and Lemieux, 2008). In Figure 4, we examine a simple scatter plot of means of various village characteristics by different population bins (each of size 25) around the threshold, to check for discontinuities of baseline covariates, and find no such evidence. The characteristics we examine include the presence of schools, health centers, electricity, presence of a telegraph office, distance from the nearest town, percentage share of scheduled castes/tribes in population, and land

[4] As we show in Section 4, our estimates still retain enough power to ensure that our cutoffs are not weak instruments.

[5] As explained below, we only have limited observations in the bank lending sample. As a result, we do not have enough power to examine the two cutoffs separately in most of our following tests, and make use of this combined above-cutoff variable. Hence, to be consistent, we present results with normalized population here as well.

irrigated. Table IA1 (Panel A) in the internet appendix shows that none of these characteristics are statistically different across cutoffs, even in a regression setting.

In Figure 5, however, when we examine a simple scatter plot of the proportion of villagers in each population bin with access to a road, we find clear indication of a significant jump in the probability of receiving a road just above the population cutoff relative to just below. To check for the statistical stability of this jump, we estimate the following regression specification as a first step, where we examine the effect of population cutoffs on actual PMGSY road construction:

$$Road_{s,v} = \beta_0 + \beta_1 1[Population\_above\_threshold] + \beta_2 1[500 - h \leq pop_{s,v} < 500 + h] * (pop_{s,v} - 500) + \beta_3 1[1000 - h \leq pop_{s,v} < 1000 + h] + \beta_4 1[1000 - h \leq pop_{s,v} < 1000 + h] * (pop_{s,v} - 1000) + \sigma_s + \epsilon_{s,v} \tag{1}$$

where $Road_{s,v}$ measures whether the village v – unconnected as of 2009 – received a PMGSY road by the end of 2014, and $pop_{s,v}$ is the baseline village population as recorded in 2001 census, and σ_s are state fixed effects. We restrict our sample to villages with population within a certain bandwidth around the threshold, such that $pop_{s,v} \in$ [c-h, c+h], where h is the population bandwidth around threshold c.

Our main specifications allow for piece-wise linearity, that is, we allow outcome variables to be related to population differently in villages with populations around 500 and 1000, with different slopes and different intercepts. In Table IA8, we show that our results are robust to alternative functional forms, such as restricting slopes and intercepts to be the same, etc.).

Note that our identification comes from threshold effects based on population. As a matter of deliberate choice *we do not exploit the differences in timing* of road construction within these threshold groups. There is indeed variation in *when* individual villages receive roads, but this time-variation in largely *endogenous*. While the government rules specify that villages with population 1080, say, should get roads before a village with population 920, it does not specify whether a village with populations 1080 should get a road before or after one with population 1170. Therefore, we do not use any time series information in our formal tests; instead, we take a snapshot of our cross-sectional data at the last available year-end, and exploit the discontinuity based on where our borrowers live.[6]

Next, we come to our bank lending sample to test our key hypotheses.

Since our bank loan data contains few small villages to start with (not surprisingly, our bank – like other private banks – finds it more fruitful to lend to larger habitations with existing roads), we choose 200 and 250 as our bandwidths for estimation purposes. Unfortunately, the number of villages falls rapidly if we restrict bandwidth further, and the resultant decline in statistical power makes our estimates lose significance.

Even so, we are only left with a total of 48 villages in our bank loan sample within a bandwidth of 200, and 58 within 250; so, again, to retain enough statistical power, we cannot estimate the thresholds of 500 and 1000 separately. In addition, given such a limited sample, we cannot employ a full-fledged regression discontinuity design with higher order polynomials etc. (for example, we do not have enough villages to construct something similar to figure 5 within the bank-lending sample). It might therefore be better to think of our test design in the bank loan sample as a treatment-control setup, where some villages are given the treatment (new roads) in a quasi-random way depending on 2001 population.

We check carefully to ensure the validity of this identification assumption – that the assignment is indeed likely to have been random – by showing that above- and below-cutoff villages are very similar. They are similar in terms of size (by design, since our sample is restricted to a bandwidth of 200/250 from the cuotffs), access-to-banking (our bank was the only lender operating in this precise area at that time in both above- and below-cutoff villages). Table IA1, Panel B shows that above- and below-cutoff villages – even within our bank lending sample – are also very similar in terms of village-level characteristics like electricity connections, irrigated land, percentage of SC/STs, primary health care centers, primary schools and distance from nearest town. Further, we also ensure that there was no other government program at the time with a population-based cutoff rule (that could have differentially affected our treatment and control sample).

[6] We take the last available year in our dataset as it allows us to measure financing responses which might take time to show up in our data, i.e., to ensure that we can measure effects even if financing takes time to respond to road connectivity.

In our bank-lending sample, we estimate the effect of roads on our financing variables by running a reduced form specification, using population-based cutoffs, as below.

$$Y_{s,v} = \gamma_0 + \gamma_1 1[Population\_above\_threshold]$$
$$+ \gamma_2 1[500 - h \leq pop_{s,v} < 500 + h] * (pop_{s,v} - 500)$$
$$+\gamma_3 1[1000 - h \leq pop_{s,v} < 1000 + h] + \gamma_4 1[1000 - h \leq pop_{s,v} < 1000 + h] * (pop_{s,v} - 1000) + + \sigma_s + \epsilon_{s,v} \tag{2}$$

where $Y_{s,v}$ are the borrower level outcomes of interest. All other variables are as described above. Standard errors are bootstrapped throughout (the Internet Appendix shows that the results with robust standard errors or stratified bootstrapping, which accounts for correlation within villages are very similar, if not stronger).

4 Results

We describe and discuss our main findings in this section.

4.1 Do population cutoffs predict road construction?

Table 1 formalizes the visual evidence in Figure 5 by presenting first stage estimates from Equation 1 using the census sample of all unconnected villages in our states of *Odisha* and *Uttarakhand*.[7] Here our unit of observation is a village. The estimates imply a 6.6–6.8 percentage point increase in the probability of treatment around the cutoff. The unconditional probability of getting a road is about 11–12%, this is about a 55% jump. This jump is highly statistically significant, with F-statistics of 39.2 and 44.9 for our two bandwidths, implying that we are not subject to a weak instrument problem. Note that although we present results for the bandwidths of 200 and 250 here to be consistent with the rest of the paper, these results are robust to other bandwidths. We show evidence for bandwidths of 100 and 150 in Table IA2 in the internet appendix.

[7] This is the set of village without paved roads according to census 2001, leaving out those that received a PMGSY road between 2001 and 2009, the year our bank started lending.

Next, we concentrate on our bank loan sample. In that sample, when we examine a bandwidth of 200 (250), we have 19 (21) villages below cutoff and 29 (37) above. Eleven of the above-cutoff villages get connected to the road network by the end of 2014, but only 10 (11) of the below-cutoff ones received roads at the same time. So the jump at cutoffs is also present in our bank-lending sample, even though the base rate for villages getting a road in this sample is much higher than that for the entire sample. This is because our bank sample villages (both above- and below-cutoff) are less remote compared to the average unconnected village – our bank was experimenting with rural banking in villages not too far (about 30 kms, on average) from district towns, that is, mainly around *Behrampur* in *Odisha*) and *New Tehri* in *Uttarakhand*.

Overall, our results confirm that there is a significant increase in the probability of treatment around the population threshold. This shows that Asher and Novosad's (2017, 2018) results on the validity of PMGSY also hold if we only focus on *Odisha* and *Uttarakhand*, and are consistent with patterns even within our bank-lending sample.

4.2 Summary statistics for our bank lending data

First, we outline the geography of our sample within *Odisha* and *Uttarakhand* in Figure 6. In this figure, we plot the locations of our 58 sample villages within bandwidth. As mentioned, above- and below-cutoff villages are geographically very close to each other within each state.

In other words, as a group above-cutoff villages are likely to be very similar to those below, e.g., in terms of topography or climate.

Table 2 shows the summary statistics for our bank dataset. The bank data we use contains cash-flow information on each loan granted, and the sample consists of all individuals who had some kind of record with our bank by the end of the calendar year 2014. We present means and standard deviations for our main variables of interest.

Panel A presents main loan characteristics observed in our dataset. Columns 1 and 2 (3 and 4) present means for borrowers residing in villages with populations within 200 (250) of the population thresholds (e.g., 300-700 and 800-1200 for 200). *Net Disbursement* is the logarithm of the net loan outstanding per villager (log of the (net) loan amount disbursed, calculated as the loan amount outstanding for each borrower at the end of the calendar year

2014 net of all repayments on that loan). *Loan maturity* is the average loan maturity for each borrower while *Interest Rate %* is the average interest rate across loans for each borrower. Interest rate information is not directly reported in our data, but we are able to back it out using information on loan amounts, type, installment payments and maturity. In order to measure loan performance, we create a variable *Overdue amount %* which captures the fraction of loan amount disbursed that was overdue at our time of measurement.

Panel B presents borrower characteristics for our main sample. *Age* is in years and *Female* is an indicator variable equal to one if the borrower is a female. To measure education level of the borrower, we create an indicator, *School education*, which takes the value of one if the borrower attended school, and zero otherwise. We also use information on borrower incomes.
All these are reported to the bank at the time of opening the account.

We find that average net loan amount disbursed in our overall sample is around Rs. 28,000 (about USD 480, at 1 USD = Rs.58, the exchange rate at that time). Loan maturity is about 3 years on average, while the average interest rate is 14.7%. Defaults are very rare in our sample, with only 0.12% of loans granted being overdue at an average point in time. Bank officials indicate to us that these low defaults are a feature of borrowers being desperate to maintain a good record with the bank for future borrowing possibilities, as their only other source of credit in these villages are the local moneylenders (who charge usurious interest rates).

The average monthly income for individual borrowers is about Rs. 11,000 (about USD 200), translating to a little over Rs. 131,000 per year.[8] Defaults, maturity and interest rates are very similar across these states. The average borrower in our sample is 37 years old. Men account for about 75% of all loans and over 87% of those who receive a loan have school education. This is a higher education level than in the underlying population, which had an average literacy rate of 63% in the 2001 census.

[8] We account for potential state-level differences in lending parameters by using state fixed effects in all our specifications, and state-year fixed effects in our panel regressions in Tables 11 and 12.

4.3 The extensive margin: village-level results

In this section, we examine whether the bank is more likely to have lent to villagers in above-cutoff villages. We do so in two steps: first, we examine extensive margins at the village level; that is, whether the bank is more likely to have given credit in an above-cutoff village, relative to a village below-cutoff. Then we examine whether the bank lent to a higher number of villagers in above-cutoff villages relative to those below.

In order to examine this evidence, we first augment our sample villages in which our bank made loans with an equal number of villages with the highest propensity scores for bank lending. The objective here is to come up with a list of very similar villages that the bank could have potentially lent in, so that we can create measures of the likelihood of our bank lending in above- and below-cutoff villages. These propensity score-matched villages come from the same state, district and block as our sample villages, and are further matched to their nearest neighbors on village population, primary school presence, the balance of lower castes (an indicator for both level of development and political balance in rural India), and distance to the nearest town. These variables are taken from the 2001 Population Census.[9] Table IA3 presents covariate balance for our matching variables, and shows that our propensity-matched villages are very similar to those that the bank actually lent in.

Table 3 presents our population cutoff-based discontinuity estimates of the impact of new roads on the extensive margin for our bank loan sample. The dependent variable in Panel A, *Extensive Margin Village*, is an indicator variable that takes on the value one for a particular village if at least one individual from that village received a loan from our bank (zero otherwise), so we run a logistic regression specification here. Column 1 presents discontinuity estimates for villages with populations within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold); and column 2 uses a threshold of 250. We find that the odds of our bank lending to someone from a given village is twice as high for villages above-cutoff as compared to those below.

The dependent variable in Panel B, *Number of Village Borrowers*, is the logarithm of (one plus) the number of villagers from each village who received a loan from our bank. Columns 1 and 2 are analogous to Panel A in terms of bandwidths. Here, again, we find that a new road

[9] We measure the proportion of scheduled caste (SC), and scheduled tribe (ST) villagers in each village to get the balance of lower castes.

is associated with a significant increase in number of villagers who receive financing. Relative to the mean of our dependent variable in below-cutoff villages, about twice as many villagers receives a loan from our bank in a village above the cutoff.

Note that villagers from surrounding villages come to take loans at the bank's branches, which are typically located in much larger villages or sub-divisional towns. Bank officials indicate to us that there is no official policy of actively going out to different villages to seek customers. In this setting, the jump in the number of customers we see around the cutoff is consistent with a demand-side story, where villagers who lacked profitable investment opportunities before but recently gained access to the road network seek out these loans. However, we cannot rule completely out the supply side story here that bank employees find it easier to provide information on the bank's loan products to connected villagers, and hence these villagers are more likely to be served. We return to this issue in Section 4.9.

4.4 Loan quantities

In Table 4 we focus on the loan amounts granted at the intensive margin, that is, within the sample of borrowers registered with the bank.

Our main dependent variable is $\frac{NetDisbursement}{AnnualIncome}$ (Panel A).[10] The coefficient of interest in these regressions give us a sense of how outstanding loans to the average villager on the bank's balance sheet differs in above- vs. below-cutoff villages.

Columns 1–2 presents coefficient estimates for villages with populations within 200 of the population threshold, while Columns 3–4 presents estimates expanding the sample to include villages within 250. Columns 1 and 3 present results with state and threshold fixed effects but no other control variables. One advantage of getting this data from the bank itself is that we have access to – and can therefore control for – borrower-level characteristics that the bank looks at in its lending decisions. So, in columns 2 and 4 we control for age, value of assets, education, gender, and a dummy variable indicating land ownership.

Our coefficients on the *Above cutoff* variable are similar across specifications, and suggests that the expansion of lending activity extends to the intensive margin. Not only do

[10] We scale by annual household income, instead of annual individual income, since the bank looks at the former variable in its decisions.

more villagers living in villages above cutoffs get loans, they also get significantly *larger* loans. In terms of economic magnitudes, the net amount lent to an average villager above the population cutoff is 30–35% higher than that to an average villager below the cutoff.

Looking at the coefficient estimates on other borrower characteristics, we find that villagers with more collateralizable assets are likely to get higher loan amounts from our bank. Younger and more educated villagers also seem to get higher loan amounts, although the latter results are not statistically significant. Women get lower loan amounts than men. One possible explanation for this latter result could be that these agrarian societies are genderbiased, and the bias shows up even in bank lending decisions; another explanation could be that the bias is in the demand side – when a family decides to take a loan, it is the male member under whose name the loan is applied for. Further, in Table IA4 in the internet appendix, we show that our results are not driven by the denominator, that is, income; lending is higher in above-cutoff villages even for our unscaled measure (log of net disbursement).

Overall, our evidence suggests that the lack of productive opportunities and infrastructure may be one reason behind lower banking penetration levels in developing economies (Agarwal et al., 2017), both on the extensive and on the intensive margins.

4.5 Loan maturities and performance

In this section we examine the maturity structure of loans granted, and their performance. If the flow of increased financing to areas with recently improved infrastructure indeed reflects improvements in productive lending opportunities, we expect loan performance, i.e., default behavior, not to be worse than in unconnected villages. Performance could either remain unchanged or improve. Given that we should measure maturity and performance only on similar loans, we add loanpurpose fixed effects in our regressions.

Table 5 presents coefficient estimates from Equation 2 on the effect of the population threshold-based discontinuity on the maturity and quality of loans disbursed. Columns 1– 3 presents coefficient estimates for villages with populations within 200 of the population threshold, while Columns 4–6 presents estimates expanding the sample to include villages within 250.

When we examine loan structure, we generally find that loans made out to villagers in above- and below-cutoff villages are of very similar maturity. This is particularly evident when the economic magnitude of the coefficients on maturity are put into perspective by benchmarking against the control group mean, that is, the average (log) maturity in belowthreshold villages, which is 1.11 (corresponds to an average maturity of 3 years, as in Table 2).

In order to measure loan performance, we create two measures: (1)Total loan amount that was overdue (*Overdue amount*), and (2)% Overdue amount (*% Overdue*), which captures overdue amount as a fraction of total loan disbursed. The evidence from the table generally suggests that individuals in villages above the threshold had slightly better repayment behaviour than those in villages below, although our estimates are not precise. Note that one reason behind the lack of significance is that the control group means themselves indicate a very low level of default, e.g., overdue amount is on average 0.12% even in below-cutoff villages. Default is very rare in our entire sample, as mentioned in Section 4.2.

Overall, both maturity and performance seem largely similar for loans made to villagers more likely to have received a new road.

4.6 Loan interest rates

In this section we present discontinuity estimates for interest rates (Table 6) on loans. The dependent variable is *AvgInterestRate*, the average interest rate across loans for each borrower (most borrowers have only one loan, a few have two loans, typically of the same type, e.g., both crop loans). Again, to ensure that we compare Interest rates only on similar loans, we add loanpurpose fixed effects in this table.

We find that interest rates on loans made out to villagers living in villages above the threshold were very similar to those on loans in below-threshold villages. These differences are not only statistically insignificant, they are also small in terms of economic magnitude – relative to an average interest rate of 15%, average interest rates in above-cutoff villages are between 14.5–14.8%.

So overall, controlling for loan type, the loans made to newly-connected villagers looked very similar to the ones made out to those likely to lack connectivity: the connected villagers were just getting more of these loans, and were paying them back at similar (if not slightly higher) rates.

Note that in our setting, the bank's aggregate lending to the rural sector was tiny relative to the size of its overall balance sheet. So supplying more capital to more profitable rural sector projects was not subject to any binding balance-sheet constraint – the situation with rural banking was probably not far from a highly elastic supply curve. Hence, it is difficult to use these interest rate (loan price) results to tease out relative shifts of demand vs. supply.

4.7 Robustness

In the Internet Appendix (IA), we present various tests to assess the robustness of our main results. First, we examine robustness with respect to various alternative standard error structures. First, in Tables IA5 - IA6 we show results for heteroskedasticity and autocorrelationrobust standard errors (results become statistically stronger, if anything). Then, in Table IA7, we present standard errors from a stratified (at the village-level) bootstrap procedure. This can account for any potential correlation across different observations coming from the same village. Again, if anything, our results get stronger statistically.

Next, in Table IA8, we present further robustness results for our baseline intensive margin specification on lending amounts. We find that our results are robust to dropping the four villages with habitations in *Uttarakhand* (Panel A), various different specifications where we explore alternative types of piecewise linearity (same slopes around the two cutoffs, same slope and intercepts, Panels B and C) and the choice of data winsorization (Panel D).

Finally, we examine the issue of 'evergreening'. 'Evergreening' in the Indian banking context refers to banks' unwillingness to recognize bad loans on their books by giving backto-back follow up loans to be used by the borrower just to pay off the previous bad loan. In the last panel of Table IA8 we rule out our effects being driven by 'evergreening' in abovethreshold villages, by showing that our results remain similar even if we only look at first time borrowers, or borrowers for whom the loans granted are not back-to-back, i.e., current loan issue date is at least one year after the last instalment pay date of his or her previous loan.

One due caveat however, is that we do not have data to distinguish whether the increase in formal financing that we document was a replacement for informal financing (e.g., village moneylenders). Note that if there were a general tendency to replace such informal borrowing with formal finance, it would affect both below- and above-cutoff villages; so our caveat here is relevant if newly connected villagers somehow had a greater tendency to replace informal with formal finance. While this is possible, even if this were the case, any replacement of informal with formal finance might still be considered a positive development given the prevalence of usurious interest rates and brutal enforcement associated with village moneylenders.[11]

4.8 A falsification test

In this subsection we conduct a placebo test to explore the possibility that some factor other than the road treatment associated with population-cutoffs may be spuriously driving our results.

In our placebo exercise, we run our baseline specification for the set of villages that had populations similar to those in our earlier tables, but were *already connected* to the road network in 2001. Importantly, all of these villages were connected to the road network under a *different program that had nothing to do with population based cutoffs*. For this sample, therefore, there is no discontinuous increase in probability of road treatment at the population threshold, although our estimation methodology remains identical. Estimates from this exercise are reported in Table 7.

We find no evidence of any effect on loan outcomes for the placebo sample, both in terms of economic magnitude and statistical significance, indicating that our results are unlikely to be driven by other discontinuous differences in villages around the cutoffs whose effect we spuriously attribute to new roads.

[11] For one of many unfortunate horror stories involving village moneylenders and their ways in India: https://www.theweek.in/news/biz-tech/2018/03/30/inside-the-bloody-world-of-india-mafia-loan-sharks.html Also, village moneylenders typically lend money for short periods (months, or even days, while the average maturity for our bank loans is about 3 years. Such longer-term loans at lower rates might allow villagers to invest in longer duration productive activities, such as replacing subsistence crops with cash crops.

4.9 Channels: Demand vs. supply

We have found that lending activity responds to new road connectivity in the previous sections. But can we say something more about the underlying mechanism? Is this driven by a shift in the demand or the supply curve – or both?

The demand side explanation here is that greater productive opportunities result in a higher demand for loans, for example by farmers who need funding to move from subsistence cereal cultivation to cash crops. On the supply side, they are two possible explanations: first, the bank might find it easier to sell loan products and reach villagers in connected villages; second, bank employees might find it easier to screen/monitor borrowers in connected villages.

While we have found that the quantity of loans responds, and the price (interest rate) typically does not, this by itself cannot be taken as evidence of a similar shift in both curves. This is because loan supply at the level of these small villages is such a tiny part of the bank's overall balance sheet that the supply curve could be highly elastic at this level. Note also that while the supply side story can explain our extensive margin results, under that view it is unclear a-priori why loan *amounts* should be higher conditional on the borrower having been reached. Still, we are open to both explanations, and in this section we focus on what we can learn from our data that can potentially distinguish between them.

4.9.1 Evidence from loan uses

First, we examine what uses the increase in financing in connected villages was being put to. For this, we partition the loan sample based on whether the financing was provided for productive uses versus other non-productive uses.

Table 8 presents our estimates of the impact of new roads on loan use. We partition the loan sample based on whether the financing was provided for productive uses (*Productive Loans*) such as crop and micro-enterprise loans, and loans for business expansion, asset acquisition, and working capital needs or other uses such as consumption needs, marriage and festival expenses (*Non-productive Loans*). Columns 1–2 presents results for Productive loans while columns 3–4 present results for Non-productive loans.

We find that the higher financing was mostly for productive purposes within above-cutoff villages. Interestingly, we find lower financing for consumption in these same villages– the bank provided less non-productive loans in above cutoff villages. These findings suggest that our main results in Table 4 are not being driven purely by wealth effects. Had it been so, we should also have seen increases in consumption loans being made out to the newly-connected villages. Also, these results are again consistent with greater demand – rather than ease of loan supply – driving our effects: there is no obvious reason why a bank that finds it easier to reach a connected village will increase supply of productive loans, but reduce consumption loans to its residents.

Finally, note that it is possible that some of the loans reported by borrowers – and classified by the bank – as productive could indeed have been used for consumption purposes. In addition, it may be hard to distinguish between these two loan purposes as there may not really be two different set of accounts in many household enterprises. These would make our classification noisy. We do not, however, have reason to believe that such misclassification would have shown a discontinuous jump at our population cutoffs. Similarly, Indian government directives urge banks to lend to certain sectors on a priority basis, and crop loans and micro enterprise loans fall under this purview. But priority lending policy again is not discontinuous at population threshold-based cutoffs. Therefore, such policy directives are unlikely to be driving our results.

4.9.2 Variability in contract terms

Another possibility for a supply-side explanation is that it is the ease of collecting soft information/monitoring, conditional on having reached the borrower, that leads banks to be more willing to lend to connected villagers. Specifically, banks might have more/better soft information on borrowers in villages with roads (where information is easier to collect), so they are willing to supply more loans at the margin. Note that our results on default rates being similar across the cutoff does not provide support for this hypothesis: if the bank did have better information on connected villagers, default rates should have been lower for them. However, default rates are so low in our sample in both below- and above-cutoff villages that this explanation perhaps warrants further attention.

Information differences are difficult to observe directly, especially if the bank's advantage in connected villages is on soft information (recall that hard information that the bank collects on borrowers is already controlled for in our tests). However, an information-based theory that can potentially explain our results here would make one other testable prediction: if the bank really had more information on borrowers in connected villages, then it should be able to better screen borrowers with the same observables. That is, loan contracts will look different for two borrowers who may look identical to an outsider based on hard information, but on whom the bank has soft information to differentiate. For example, Cornell and Welch (1996) suggest that proximity may reduce information asymmetry in a lending transaction by improving the precision of the signal that the officer obtains about a borrower. Their model predicts that proximity should increase the variance of loan sizes, as the officer's distribution of prior beliefs of borrower quality widens with the more precise signal.

This yields an empirically rejectable hypothesis: if our results are driven by soft information/better screening, loan contract terms will be more variable within groups of borrowers – who are similar on observables – in connected villages relative to unconnected ones. We test this prediction here. Note that our test here is similar in spirit to that conducted by Fisman, Paravisini and Vig (2017).[12]

To test this hypothesis, we divide borrowers into groups based on observable characteristics such as gender, education, household assets, and age. We generate two groups each based on gender and school education, and within each of these groups, we create three further groups based on household assets and borrower age. Within each of these groups – now we are looking at observationally similar borrowers – we compute the coefficient of variation of loan contract terms: loan amounts, Interest rates, maturity. We then test whether the *coefficient of variation* of loan contract terms are different in above-cutoff villages relative to below-cutoff villages. The prediction from the soft information story will be that the coefficient of variation will be higher for above-cutoff villages.

We find in Table 9 that the variability of loan contract terms in above- vs below-cutoff villages is very similar; their differences are small, and statistically indistinguishable from zero. In sum, we do not find support for a soft information-based supply-side channel.

[12] In their paper, they show that loans made by officers from the same religion/caste group in India have a substantially larger size dispersion, relative to those made by out-group officers. They suggest that this evidence is consistent with information advantages in within-group transactions, indicative of 'soft' information.

Finally, conversations with bank officials also corroborate the demand view. The bank operates on a nodal-branch network: bank branches are set up where villagers from surrounding villages come to take loans. Given that the bank interest rates are far lower than the local moneylenders', and that these are severely under-banked areas with no competition from alternative formal lenders (our bank is the only formal/institutional lender in all of our sample villages, for example); there is ample demand for its services that the bank does not need to actively go out to these villages and market itself.

Overall, then, the weight of our evidence seems to favor the demand- rather than the supply-side being dominant in our context, although we cannot definitively rule out that both demand and supply curves might have shifted outward in response to connectivity.

5 Distributional consequences of connectivity: Evidence from loan financing

We have thus far established the causal impact of rural roads on lending flows. In this section, we examine the heterogeneity of the treatment effects based on baseline borrower characteristics. Under the assumption (from e.g., our aggregate evidence) that financing indeed flows to those who see largest productivity changes, these estimates can also be interpreted as being useful to understand who benefits more from transportation infrastructure. Also, in our discussion of these results below, we will focus on loan amounts, as there is no meaningful difference in default behavior or interest rates of note across different types of borrowers (although these numbers are also reported in the table).

Here we use individual-level data to examine the distribution of treatment effects across subgroups with different household assets and income. We exploit the data on demographic information, such as the gender of the borrower, and importantly, information on the borrower's assets, income and education at the time of the bank account opening. We present these results in Table 10. Columns 1–3 (4–6) presents discontinuity estimates for villages with populations within 200 (250) of the population thresholds.

Results from Table 10 suggest that new roads seem to alleviate collateral constraints among borrowers. Low asset households seem to benefit disproportionately more from connectivity. The coefficient on *LowAssets*, a dummy variable that takes the value of one if the assets is below the sample median, suggests that loan sizes for this group is higher by about 25% in above-cutoff villages. This is interesting because this evidence rules out connectivity driving our effects through increases in land value. Had the higher loan amounts to the newly connected villagers reflected roads increasing the value of their collateral, e.g., existing land, and therefore borrowing capacity, our effects would have been stronger among those with higher assets. This is important, because giving access to financial markets to landless peasants – some of the poorest sections of village society in India – has long been a focus of policy for governments.

In terms of other important issues in lending in rural India, we find that on average women get smaller loans, and are charged about half a percent higher interest rates (relative to a sample mean of 15%). While we find that these loan amount differences persist in above-cutoff villages, the interest rate charged to women seem to converge to those for men in these villages. However, the difference in interest rates between above- vs below-cutoff villages is not statistically significant in one of our specifications, so this evidence has to be interpreted with caution. We also examine differences between lending to SC/ST/OBCs (who are often poorer and have less opportunities) and others in our data, and find no significant differences. Finally, we observe that loan disbursements are higher in newly connected villages for villagers with some education (which we measure using an indicator on whether the borrower attended a village school). These effects are however not statistically significant. Still, these results on the educated are directionally consistent with Mukherjee (2011) and Adukia et al. (2018), who respectively show that PMGSY increases school enrollment, and that children stay in school longer in connected villages. If villagers saw benefits of the road accrue more visibly to the more educated, this would encourage them to invest more in their children's education.

Overall, we find that the financing response to new roads seems to disproportionately benefit villagers who lack collateralizable assets.

6 External validity and macroeconomic effects

One concern with discontinuity designs like our is that the identification comes at the cost of external validity of findings. Unfortunately our proprietary data does not allow us to examine the causal impact of new roads on lending across a more general geography; neither do we have any other data at the same level of granularity which allows similar analysis. However, the Reserve Bank of India (RBI) does provide macro data, aggregated by districts, on overall lending activity by sector (rural, urban, etc.). We use this data to examine the macro associations between roads and lending, and examine whether these effects are qualitatively consistent with our earlier results.

The main cost that we incur to translate things to the macro level is that we lose tightness of identification. The RBI data is not at the village-level, so we cannot identify effects based on the program discontinuity; still we do our best to best to account for many time-varying control variables, as well as for state-level macroeconomic trends through the use of highdimensional fixed effects. Our main explanatory variable here is the length of road built under the same PMGSY program in the past three years at the district level. This allows our effects to show up even if they take some time to manifest. All our empirical specifications include district fixed effects to control for district-level time-invariant characteristics. We augment this by adding state-by-year fixed effects to remove time-varying local economic confounds (e.g., regional macroeconomic shocks).

We first regress aggregate changes in annual lending, the number of bank branches, and the total deposits across all private banks in each district on the aggregate length of PMGSY road constructed in the past three years. Table 11 presents results. Panel A presents results without any control variables, but with district and state-year fixed effects. In columns (1)-(3), we examine private bank activity in *rural* areas. Here, we find an increase in rural lending and deposits following the construction of new rural roads. The number of bank branches does not seem to respond. The coefficient on credit indicates that every onestandard deviation increase in the length of new rural roads is followed by about 10.52% increase in rural lending (coefficient of 0.466, multiplied by standard deviation of 2, relative to the mean of the dependent variable being 8.86). Similarly, higher deposits also follow – here a one-standard-deviation increase in roads is followed by a 17.98% increase (coefficient of 0.705, multiplied by standard deviation of 2, relative to the mean of the dependent variable being 7.84).

Next, in column (4)-(6) we look at *urban* lending in the same districts using the same specification. There is no response in urban lending, both in terms of economic magnitudes and statistical significance. This is consistent with our effects being a likely response to *rural* roads – which affects rural areas disproportionately– and not some overall macroeconomic or political change in these districts.

In Panel B, our specification accounts for a number of economic and political variables which might affect bank activity and road construction at the same time. As control variables, we incorporate various economic and political indicators that might simultaneously affect financial development and economic growth. First, we rely on the *Village Dynamics in South Asia* (VDSA) dataset, maintained by the *International Crops Research Institute for the Semi-Arid Tropics* (ICRISAT). We use this data to construct district-level time-varying control variables for the total geographical area under land use (reflecting, for example, better irrigation facilities, which could create higher agricultural growth as well as change lending by reducing the risk of crop failures). We also control for the field wage for males, which is a key indicator of growth frequently used by policy makers in the rural Indian context. Next, we add the fraction of population that is literate; literacy is often thought to be an impediment to both rural growth and to villagers accessing formal finance (which requires them to fill out many forms, for example).

Next, we add information on political balance and competition, collected from the *Election Commission of India* (ECI). In particular, we compute the district-level vote margins for the two leading political groups in our states, as well as the difference between the winning and the runner-up candidate in the general elections. The latter variable is a measure of political competition. These political controls account for the possibility that constituencies aligned with the party in power at the state/central level, and/or closely contested areas, could see more resources devoted to them, which might simultaneously affect infrastructure/financial development and economic growth. We find very similar results to those in Panel A.

In our final test, presented in Table 12, we come back to the motivation we started with. Should we care about the availability of finance in relation to infrastructure-building? This issue assumes importance especially in light of Asher and Novosad (2018), who find little or no effects of road-building activity on village income or output-related outcomes. Analogous to Table 11, Panel A presents results with district and state-year fixed effects, while Panel B additionally controls for same district-level time-varying controls described above.

Here we look at changes in district-level GDP growth rates following the construction of new rural roads during the previous three years, depending on the depth of the rural credit market in each district. Our findings suggest that while roads are not always followed by higher output on average (Asher and Novosad, 2018), such higher output can follow *when roads are constructed in areas with better credit markets*.[13]

A one-standard-deviation increase in rural road length in the previous three years is associated with a 40.6 basis points (0.203 times standard deviation 2) higher district GDP growth rate (on a mean GDP growth rate of 7.32% during this period) in districts with an above-average rural credit to GDP ratio. There is no discernable effect of roads on districts with below-average depth of rural credit markets.

In columns 2 and 3, we look at GDP growth – still at the district level – but broken down by sectors. Our effects come only from agriculture, and not from industries. A one-standard deviation increase in new the length of new rural roads is followed by an approximately 1% increase in district-level agrarian GDP growth rate, relative to a base of 6.25%. The analogous number for industrial GDP is statistically, as well as economic magnitude-wise, close to zero. This, again, is what one would expect if the effects occur through village roads – these small villages are predominantly agricultural. Just as in Table 11, adding time-varying district-level control variables in Panel B does not alter these results meaningfully.

Overall, our macro evidence suggests that financing is important to realize the benefits of connectivity, and that such financing indeed follows rural road-building well beyond our baseline bank-loan sample. Our conclusions, however, need to tempered by the fact that access to credit in rural areas remains very low in India and other developing countries, and branch expansion does not seem to respond to roads. Much, therefore, remains to be done.

7 Conclusion

Increasing infrastructure investments are a key component of growth strategy in many countries, and a particular focus of policy now (e.g., China's massive "Belt and Road Initiative"). Although it is typically assumed that financing to households will follow once roads are built, allowing them to make the best use of new productive opportunities, little is known about whether this really happens, especially in poor countries. Moreover, even if financing does follow infrastructure improvements, does it disproportionately benefit the

[13] Districts with better credit markets are defined based on whether the total (private plus state-owned banks) rural credit per capita in a given year is above (below) the median district.

rich who had assets prior to the infrastructure being built, and were in a better position to exploit the resultant opportunities? Or does it benefit the poor too who were excluded from formal finance before, but can now find a way in?

We use a population-based discontinuity setting around a large rural road construction program in India to answer these questions. We find that private financing does indeed respond to changes in productive opportunities resulting from connectivity. Financing flows disproportionately to villagers who lack collateralizable assets, and have traditionally been amongst the most disadvantaged. Our results have important implications for understanding trickle down benefits of infrastructure-building and its distributional consequences.

8 References

Adukia, A., Asher, S. and Novosad, P., 2018, Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction, American Economic Journal: Applied Economics, forthcoming.

Agarwal, S., Alok, S., Ghosh, P., Ghosh, S., Piskorski, T. and Seru, A., 2017, Banking the Unbanked: What Do 255 Million New Bank Accounts Reveal about Financial Access?, Columbia Business School Research Paper No. 17-12.

Aggarwal, Shilpa, 2018, Do Rural Roads Create Pathways out of Poverty? Evidence from India, Journal of Development Economics 133, 375-395.

Aghion, P. and Bolton, P., 1997, A Theory of Trickle-Down Growth and Development, Review of Economic Studies 64(2), 151-72.

Allen, F., Elena Carletti, Robert Cull, Jun “QJ” Qian, Lemma Senbet, and Patricio Valenzuela, 2011, Improving Access to Banking: Evidence from Kenya, CEPR Discussion Paper No. DP9840.

Artadi, E., Sala-I-Martin, X., 2004, The Economic Tragedy of the Twentieth Century: Growth in Africa, The Africa Competitiveness Report 2004.

Asher, S., and Paul Novosad, 2017, Market Access and Structural Transformation: Evidence from Rural Roads in India, Working Paper.

Asher, S., and Paul Novosad, 2018, Rural Roads and Local Economic Development, American Economic Review, conditionally accepted.

Aterido, R., Thorsten Beck, and Leonardo Iacovone, 2013, Gender and Finance in SubSaharan Africa: Are Women Disadvantaged?, World Development 47, 102-120.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, 2012, On the Road: Access to Transportation Infrastructure and Economic Growth in China, Working Paper.

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, and Matthew A Turner, 2015, Roads, Railroads and Decentralization of Chinese Cities, Working Paper.

Beck, T., 2008, The Econometrics of Finance and Growth, Palgrave Handbook of Econometrics, Vol. 2.

Beck, T., 2012, The Role of Finance in Economic Development – Benefits, Risks, and Politics, in: Dennis Mu¨ller (Ed.): Oxford Handbook of Capitalism.

Beck, T. and A. Demirguc-Kunt, 2008, Access to Finance - An Unfinished Agenda, World Bank Economic Review, 22, 383-396.

Beck, T. and A. Demirguc-Kunt, and Levine, R., 2007, Finance, Inequality, and the Poor, Journal of Economic Growth, Vol. 12, pp. 27–49.

Beck, T., A. Demirgu¨¸c-Kunt and M. Martinez Peria, 2007, Reaching Out: Access to and Use of Banking Services across Countries, Journal of Financial Economics 85, 234-66.

Beck, T., Liping Lu and Rudai Yang, 2014, Finance and Growth for Microenterprises: Evidence from Rural China, World Development 67, 38-56.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, 2004, How much should we trust differences-in-differences estimates? The Quarterly journal of economics 119 (1), 249-275.

Black, Sandra E., and Strahan E. Philip, 2002, Entrepreneurship and Bank Credit Availability, Journal of Finance 57, no. 6, 2807 – 2833

Burgess, Robin and Dave Donaldson, 2012, Railroads and the Demise of Famine in Colonial India, Working Paper.

Burgess, Robin, and Pande Rohini, 2004, Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment, American Economic Review 95, no. 3, 780 – 795.

Caldero´n, C´esar, and Luis Serv´en, 2004, The effects of infrastructure development on growth and income distribution, No. 270., World Bank Publications.

Coffey, E., 1998, Agricultural finance: Getting the policies right, Agricultural Finance Revisited: vol. 2, FAO and GTZ.

Conning, J. and Udry, C., 2005, Rural Financial Markets in Developing Countries, in *The Handbook of Agricultural Economics Vol. 3, Agricultural Development: Farmers, Farm Production and Farm Markets*, edited by Evenson, R.E., P. Pingali, and T. P. Schultz.

Das, Abhiman, Ejaz Ghani, Arti Grover, William R. Kerr, and Ramana Nanda, 2019, Infrastructure and Finance : Evidence from India's GQ Highway Network. Policy Research Working Paper No. 8885. World Bank.

Demirguc-Kunt, A., and Ross Levine, 2008a, Finance, financial sector policies, and longrun growth, World Bank Publications Vol. 4469.

Demirguc-Kunt, A. and Ross Levine, 2008b, Finance and Economic Opportunity, Proceedings of the Annual World Bank Bank Conference on Development Economics, Washington, DC.

Demirguc-Kunt, A., and Ross Levine, 2009, Finance and Inequality: Theory and Evidence, Annual Review of Financial Economics 1, 287-318.

Demirguc-Kunt, A., Erik Feijen, and Ross Levine, 2013, The Evolving Importance of Banks and Markets in Economic Development, World Bank Economic Review 27(3): 476-
490.

District Plan Report(2010-2011), GANJAM District, ORISSA.

Donaldson, Dave, 2017, Railroads of the Raj: Estimating the Impact of Transportation Infrastructure, American Economic Review, forthcoming.

Donaldson, Dave and Richard Hornbeck, 2017, Railroads and American Economic Growth: A "Market Access" Approach, Quarterly Journal of Economics, forthcoming.

Faber, Benjamin, 2014, Trade Integration, Market Size, And Industrialization: Evidence from China's National Trunk Highway System, Review of Economic Studies 81 (June).

Ghani, Ejaz, Arti Grover Goswami, and William R. Kerr, 2016, Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing, Economic Journal, 317-357.

Ghani, Ejaz, Arti Grover Goswami, and William R. Kerr, 2017, Highways and Spatial Location within Cities: Evidence from India, World Bank Economic Review, S97-S108.

Imbens, Guido W. and Thomas Lemieux, 2008, Regression discontinuity designs: A guide to practice, Journal of Econometrics 142 (2), 615–635.

Khandker, Shahidur R., Zaid Bakht, and Gayatri B. Koolwal, 2009, The poverty impact of rural roads: evidence from Bangladesh, Economic Development and Cultural Change 57(4), 685-722.

King, R.G. and Levine, R., 1993a, Finance, entrepreneurship and growth: theory and evidence, Journal of Monetary Economics, 32, 513-542.

King, R.G. and Levine, R., 1993b, Finance and growth: Schumpeter might be right, The Quarterly Journal of Economics, 108(3), 717-737.

Levine, R., 1997, Financial Development and Economic Growth: Views and Agenda, Journal of Economic Literature 35: 688–726.

Levine, R., 2005, Finance and growth: Theory and evidence, in Handbook of Economic Growth, Eds:Philippe Aghion and Steven Durlauf, The Netherlands: Elsevier Science.

Levine, R., 2008, Finance, Growth, and the Poor, The Financial Development Report 2008, World Economic Forum.

McCrary, Justin, 2008, Manipulation of the running variable in the regression discontinuity design: A density test, Journal of econometrics 142.2, 698-714.

Morten, Melanie and Jaqueline Oliveira, 2014, Migration, roads and labor market integration: Evidence from a planned capital city, Working Paper.

Mukherjee, Mukta, 2011, Do Better Roads Increase School Enrollment? Evidence from a Unique Road Policy in India, Working paper.

Satish, P., 2004, Rural Finance: Role of State and State-Owned Institutions, Economic and Political Weekly Vol. 39, No. 12.

Schumpeter, J., 1911, The Theory of Economic Development, Harvard University Press, Cambridge, MA.

Shamdasani, Y., 2017, Rural Road Infrastructure and Agricultural Production: Evidence from India, Working Paper.

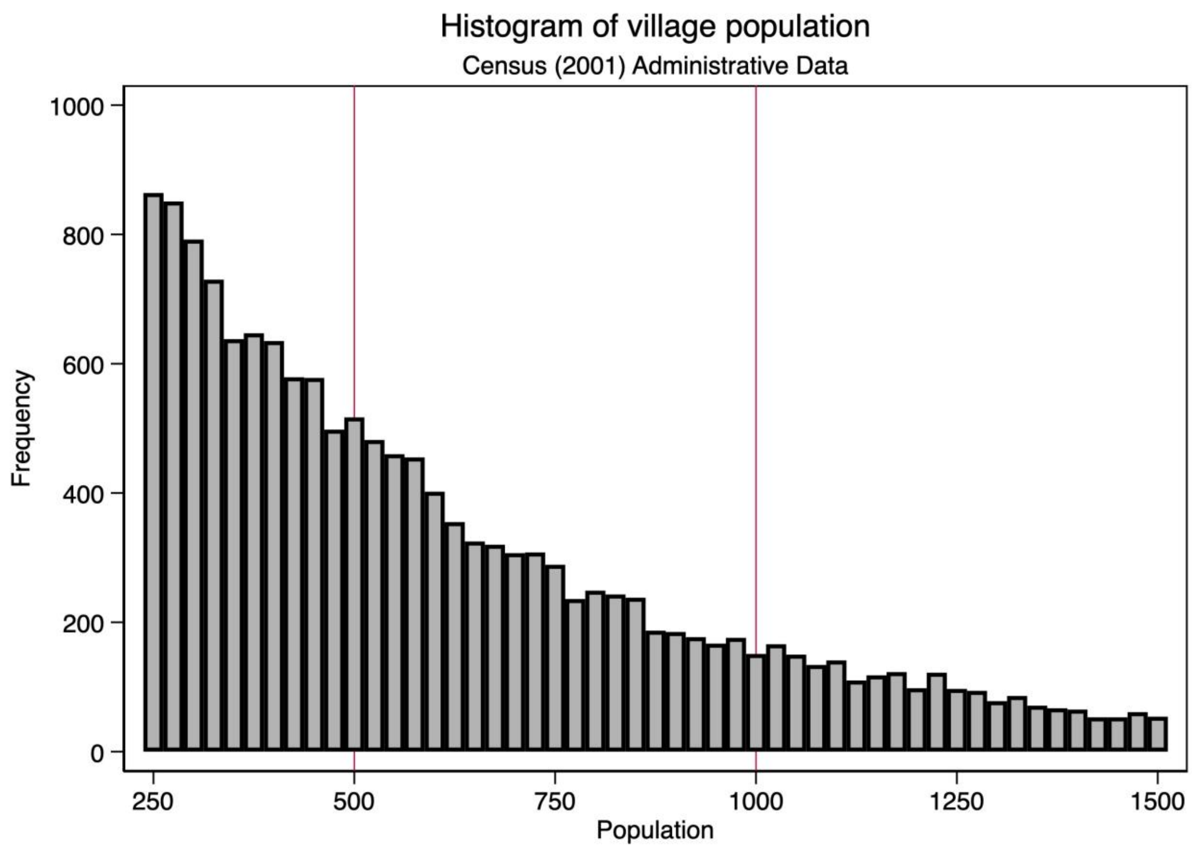
Storeygard, Adam, 2014, Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa, Working Paper.

Vig, V., 2013, Access to Collateral and Corporate Debt Structure: Evidence from a Natural Experiment, Journal of Finance LXVIII, No. 3, 881-928.

Visaria, S., 2009, Legal reform and loan repayment: The microeconomic impact of debt recovery tribunals in India, American Economic Journal: Applied Economics 1, 59–81.

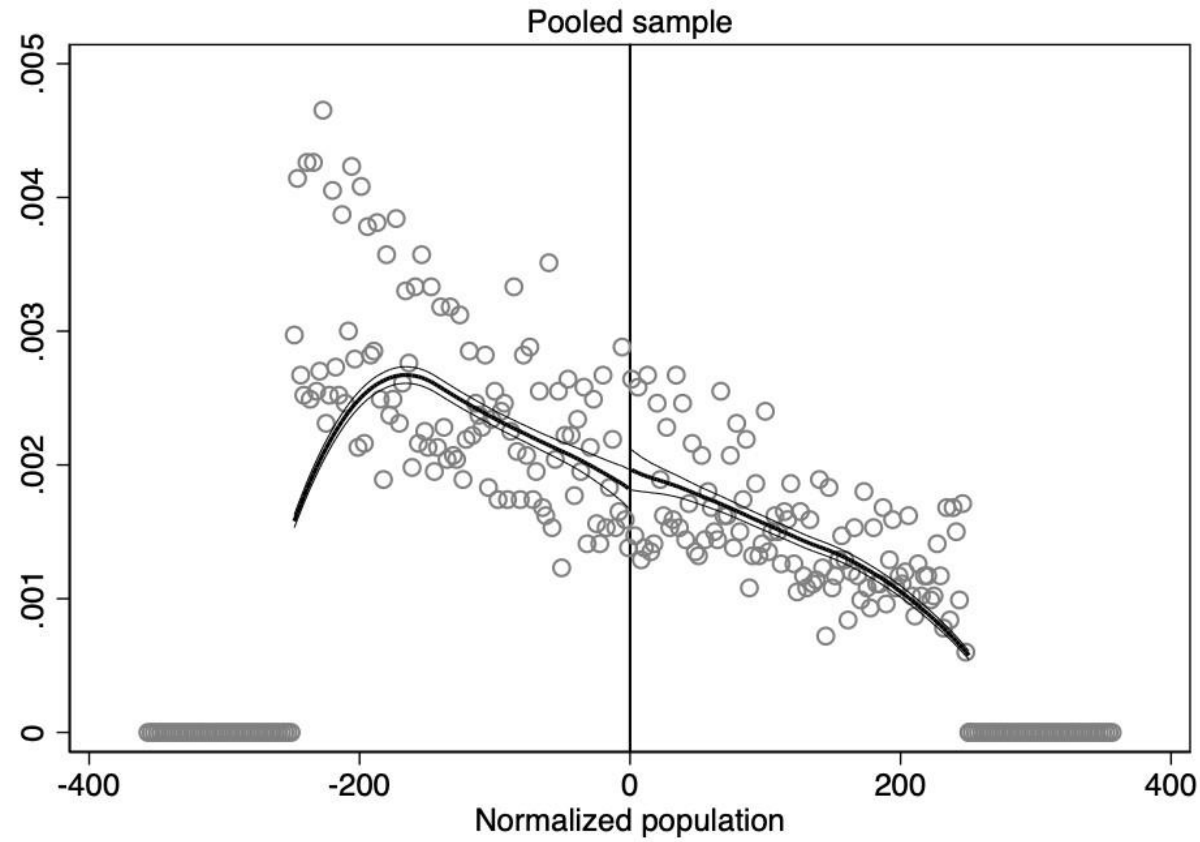
World Bank, 2007, A Decade of Action in Transport: An Evaluation of World Bank Assistance to the Transport Sector, 1995–2005. Independent Evaluation Group, World Bank, Washington, D.C.

Figure 1: Distribution of running variable



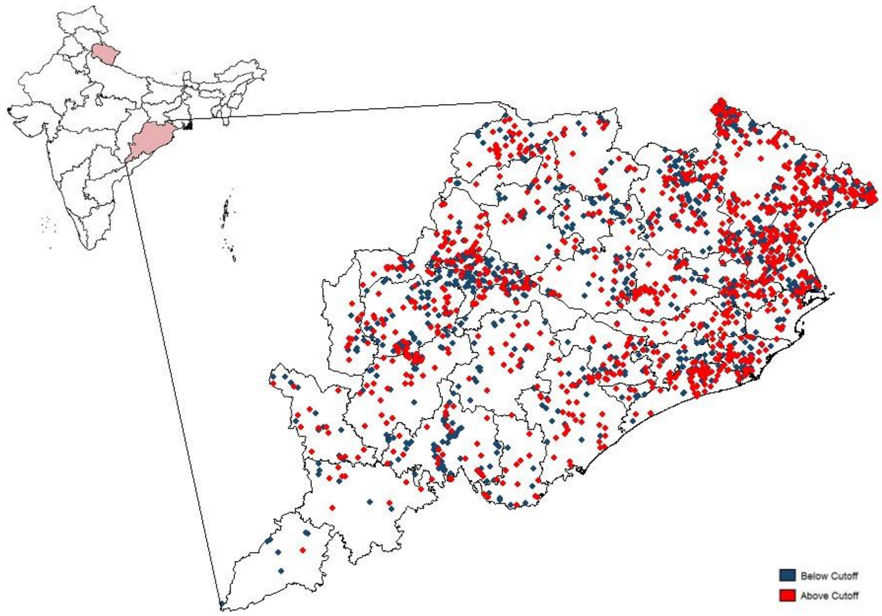
Notes: The figure shows the distribution of village population around the different population thresholds as outlined under PMGSY guidelines. We present the histogram of village population as recorded in the 2001 Population Census. The vertical lines depict the program eligibility cutoffs as defined in PMGSY at 500 and 1000. The sample consists of villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample.

Figure 2: McCrary Test for discontinuity in the running variable

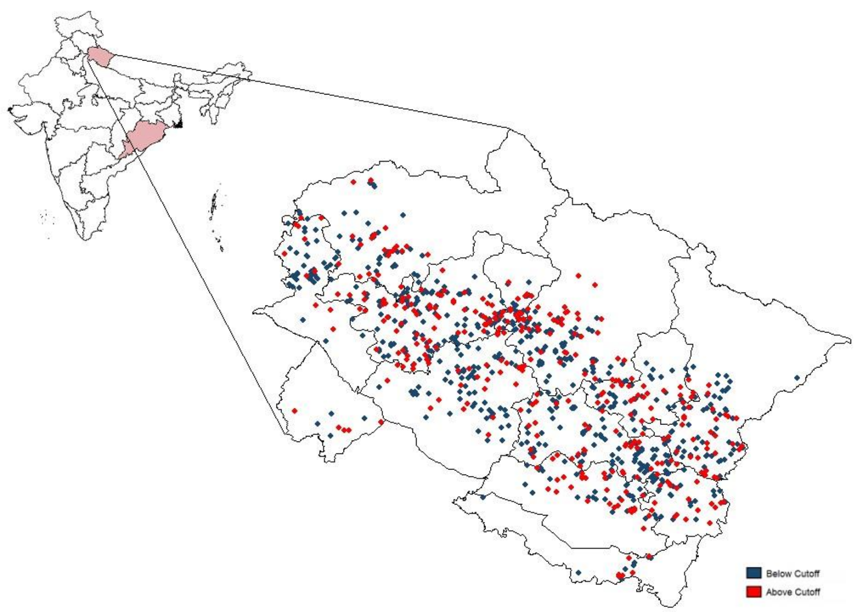


Notes: These figures plot non-parametric regressions of the distribution following McCrary (2008), testing for a discontinuity at zero. The village population is normalized by subtracting the population threshold, either 500 or 1000. The sample consists of villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample. The point estimate for the discontinuity is 0.080, with a standard error of 0.058.

Figure 3: Distribution of unconnected villages based on cutoff



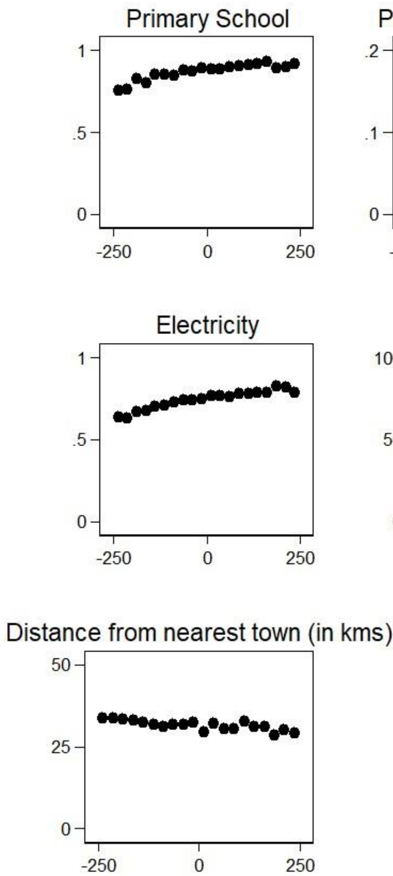
(a) Odisha



(b) Uttarakhand

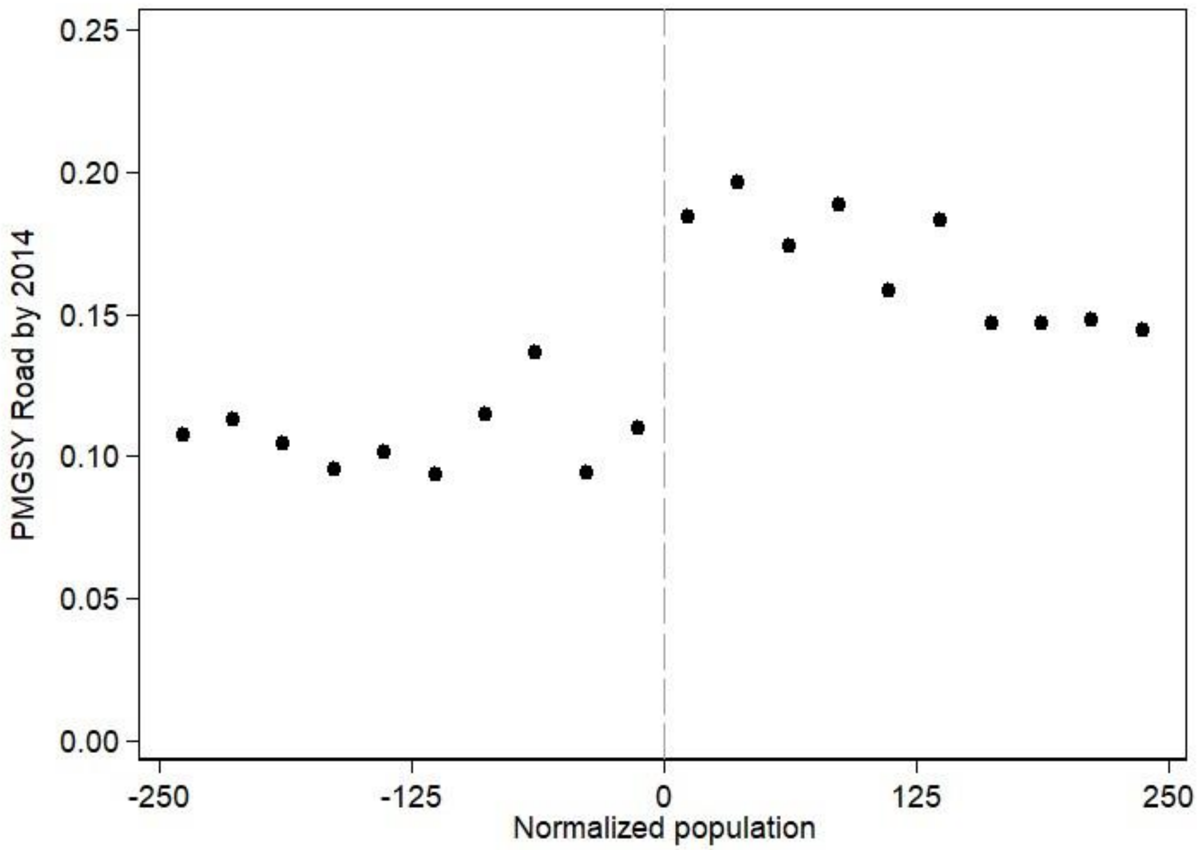
Notes: The figure shows the distribution of unconnected villages in all districts of Odisha (panel a) and Uttarakhand (panel b). The sample consists of villages that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. Blue shaded regions represent villages right below the population cutoff while red shaded regions display villages right above the population cutoff.

Figure 4: Balance of baseline village characteristics



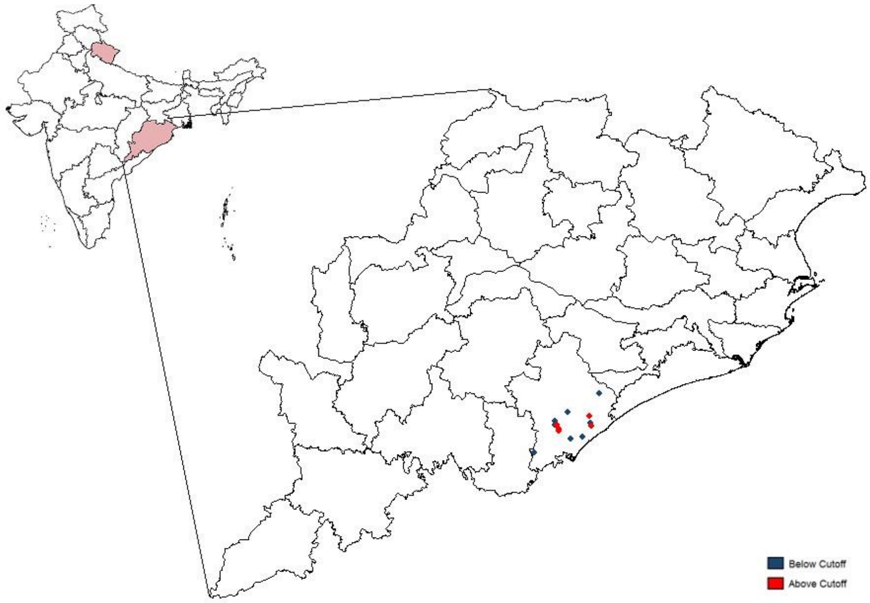
Notes: The figure plots means of baseline village characteristics over normalized popuation. Points to the right of zero are above treatment thresholds, while points to the left of zero are below treatment thresholds and the bin width is 25 on either side of the threshold. The sample consists of villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample.

Figure 5: First stage: effect of road prioritization on probability of PMGSY road by 2014

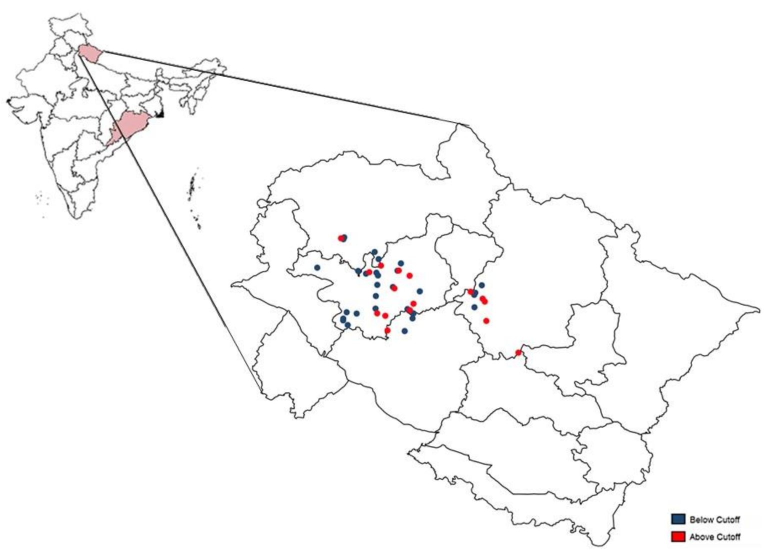


Notes: The figure plots the probability of receiving a road under PMGSY by 2014 over village population as recorded in the 2001 Population Census. The village population is normalized by subtracting the population threshold, either 500 or 1000. The bin width is 25 on either side of the threshold. The sample consists of villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample.

Figure 6: Geographic dispersion in banking lending sample



(a) Odisha



(b) Uttarakhand

Notes: The figure displays the geographic dispersion of the villages based on the cutoff in our bank lending sample. The sample consists of villages in the Ganjam district of Odisha (panel a) and the districts of Chamoli, Garhwal, Tehri Garhwal, and Uttarkashi of Uttarakhand (panel b).

Table 1: First stage effect of road priority on PMGSY road treatment: (2009 - 2014)

The table presents first stage estimates from Equation 1 of the effect of being above the population threshold on a village's probability of receiving a road under PMGSY by 2014. The dependent variable is an indicator variable that takes on the value one if a village has received a PMGSY road before 2014. Column 1 presents results for villages within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while column 2 expands the sample to include villages within 250 of the population threshold. The regression specification includes state and threshold fixed effects. The sample consists of all the villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. We report bootstrapped standard errors below point estimates.

	(1)	(2)
Bandwidth	±200	±250
Above cutoff	0.068***	0.066***
	(0.011)	(0.010)
Control group mean	0.12	0.11
F-statistic	39.22	44.94
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	11,136	14,205

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Summary statistics: Bank loan sample

The table presents means and standard deviations for our primary variables of interest. Our sample is a proprietary rural bank-account level dataset from one of India's largest publicly traded banks. Panel A presents main loan characteristics observed in our dataset while Panel B presents borrower characteristics for our main sample. Columns 1 and 2 present means for borrowers residing in villages with populations within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while columns 3 and 4 presents means and standard deviations expanding the sample to include villages within 250 of the population thresholds. *Net Disbursement* is the net loan amount disbursed. For each borrower, we compute the net loan amount disbursed as loan amount disbursed minus any repayment made by the end of the calendar year 2014. *Loan maturity* is the maturity in years for each borrower while *Interest rate* is the average interest rate across loans for each borrower. *Overdue amount* captures the fraction of loan amount disbursed that was overdue by the end of 2014. *Age* is in years and *Female* is an indicator variable equal to one if the borrower is a female. We create an indicator measure, *Schooling*, which takes the value of one if the borrower has attended a school, and zero otherwise. *Annual income* is the individual income of the borrower at the time of opening an account with the bank. The bank loan sample consists of individuals from 58 villages in Odisha and Uttarakhand in which the bank lent and who had a loan with the bank by the end of the calendar year 2014.

Bandwidth	0		±250	
		(2)	(3)	(4)
	Mean	Std. dev.	Mean	Std. dev.
Panel A: Loan characteristics				
Net disbursement (Rs.)	28,091	19,106	28,064	19,589
Loan maturity (years)	3.06	0.15	3.06	0.15
Interest rate (%)	14.7	3.7	14.7	3.8
Overdue amount (%)	0.12	2.25	0.12	2.19
Panel B: Borrower characteristics				
Age (years)	37	9	37	9
Female (%)	25	43	25	43
Schooling (%)	87	33	87	34
Annual income (Rs.)	131,892	88,718	132,286	89,368

Table 3: Impact of new roads on lending: Extensive margin

The table presents reduced form estimates of the effect of new rural roads on the propensity of the bank to enter a village in our sample. Column 1 presents reduced form estimates for villages within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while column 2 presents reduced form estimates expanding the sample to include villages within 250 of the population threshold. The dependent variable in panel A, *ExtMargin*, is an indicator variable that takes on the value one if an individual in the village received a loan from the bank while the dependent variable in panel B, *LogCustomers*, is the natural logarithm of one plus the number of customers served by the bank in each village.

We construct the control group villages using propensity score matching. Specifically, we require the control group villages to be in the same block and match them on the following village-level covariates as recorded in the 2001 Population Census: fraction of SC/ST population, village population, presence of primary school, and distance from the nearest town. Internet Appendix Table IA 2 presents the covariate balance. All specifications include state and threshold fixed effects. Panel A reports the odds ratio which are estimated using a Logit specification while coefficients in panel B are estimated using an Ordinary Least Squares (OLS) specification. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. We report bootstrapped standard errors below point estimates.

	PanelA:Bankentry	
Bandwidth	± 200	± 250
	(1)	(2)
Above cutoff	2.003*	1.733**
	(1.059)	(0.858)
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	93	116

	PanelB:Numberofcustomers	
Bandwidth	± 200	± 250
	(1)	(2)
Above cutoff	0.938**	0.652*
	(0.409)	(0.374)
Control group mean	1.012	0.867
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	93	116

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Impact of new roads on the lending quantities

The table presents reduced form estimates from Equation 2 of the effect of new rural roads on lending activity within the villages. Column 1 presents reduced form estimates for villages within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while column 2 presents reduced form estimates expanding the sample to include villages within 250 of the population threshold. The dependent variable, *NetDisburse/Inc*, is the net loan amount disbursed divided by household income of each borrower. For each borrower, we compute the net loan amount disbursed as loan amount disbursed minus any repayment made by the end of the calendar year 2014. Our bank loan sample consists of individuals who had a loan with the bank by the end of the calendar year 2014. We include villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. The specification also includes baseline borrower-level controls for age, land ownership, household assets, education, gender, and household income. All specifications include state and threshold fixed effects. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. We report bootstrapped standard errors below point estimates.

Bandwidth	±200		±250	
	(1)	(2)	(3)	(4)
Above cutoff	0.025**	0.025**	0.025**	0.030***
	(0.012)	(0.012)	(0.011)	(0.011)
Age (years)		-0.001**		-0.000*
		(0.000)		(0.000)
Land		0.006		0.003
		(0.007)		(0.007)
Log (1+assets)		0.002*		0.002**
		(0.001)		(0.001)
School education		0.009		0.009
		(0.007)		(0.006)
Female		-0.051***		-0.049***
		(0.006)		(0.006)
Control group mean	0.083	0.083	0.085	0.085
State fixed effects	Yes	Yes	Yes	Yes
Threshold fixed effects	Yes	Yes	Yes	Yes
Observations	1,032	1,032	1,084	1,084

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Impact of new roads on loan maturity and quality

The table presents reduced form estimates from Equation 2 of the effect of new rural roads on quality of loan disbursed. Columns 1 through 3 present reduced form estimates for villages within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while columns 4 through 6 present reduced form estimates expanding the sample to include villages within 250 of the population threshold. To measure loan performance, we create two measures: (1) *% OD Amount*, captures the fraction of loan amount disbursed that was overdue (2) *ODAmount*, is the total loan amount that was overdue. The dependent variable in columns 1 and 4 is natural logarithm of loan maturity. The dependent variable in columns 2 and 5 is Total Overdue amount while in columns 3 and 6 it is % Overdue amount. Our sample consists of individuals who had a loan with the bank by the end of the calendar year 2014. We include villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All specifications include loan purpose, state, and threshold fixed effects and baseline borrower-level controls for age, land ownership, household assets, education, and gender. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. We report bootstrapped standard errors below point estimates.

Bandwidth	±200			±250		
	(1)	(2)	(3)	(4)	(5)	(6)
	Ln(Maturity)	ODAmount	%OD Amount	Ln(Maturity)	ODAmount	%OD Amount
Above cutoff	-0.009	-164.872	0.057	-0.028	-184.306	-0.066
	(0.020)	(211.679)	(0.355)	(0.019)	(190.637)	(0.361)

Control group mean	1.11	104.7	0.12	1.11	100.6	0.12
Loanpurpose fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Threshold fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	630	630	630	665	665	665

Standard errors in parentheses
[*] $p < 0.10$, [**] $p < 0.05$, [***] $p < 0.01$

Table 6: Impact of new roads on interest rates

The table presents the effect of new rural roads on Interest rates of loan disbursed. Column 1 presents reduced form estimates for villages with populations within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while column 2 presents reduced form estimates expanding the sample to include villages within 250 of the population thresholds. The dependent variable is the average interest rate across loans for each borrower. Our sample consists of individuals who had a loan with the bank by the end of the calendar year 2014. We include villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All specifications include loan purpose, state, and threshold fixed effects and baseline borrower-level controls for age, land ownership, household assets, education, and gender. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. We report bootstrapped standard errors below point estimates.

Bandwidth	± 200	± 250
	(1)	(2)
Above cutoff	-0.002	-0.005
	(0.006)	(0.005)
Control group mean	0.15	0.15
Loanpurpose fixed effects	Yes	Yes
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	630	665

Standard errors in parentheses
[*] $p < 0.10$, [**] $p < 0.05$, [***] $p < 0.01$

Table 7: Placebo on connected villages

The table presents reduced form estimates of the effect of new rural roads on loan disbursed on a placebo sample of villages within Odisha and Uttarakhand. Specifically, we include villages that were already connected at baseline and hence the PMGSY thresholds were not applicable to them. Further, we restrict the villages to be within the same block and having similar amenities as recorded in the 2001 Population Census. Column 1 presents estimates for villages within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while column 2 presents estimates expanding the sample to include villages within 250 of the population threshold. The dependent variable, *NetDisburse/Inc*, is the net loan amount disbursed divided by household income of each borrower. For each borrower, we compute the net loan amount disbursed as loan amount disbursed minus any repayment made by the end of the calendar year 2014. All specifications include state and threshold fixed effects and baseline borrower-level controls for age, land ownership, household assets, education, and gender. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. We report bootstrapped standard errors below point estimates.

Bandwidth	±200	±250
	(1)	(2)
Above cutoff	0.004	-0.005
	(0.009)	(0.008)
Control group mean	0.064	0.072
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	2,256	2,675

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Impact of new roads on loan types

This table presents reduced form estimates from Equation 2 quantifying the effect of new rural roads on lending activity based on the type of loan disbursed. Columns 1 and 3 presents reduced form estimates for villages within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while columns 2 and 4 presents reduced form estimates expanding the sample to include villages within 250 of the population threshold. Columns 1 and 2 present results for *Productive loans* while columns 3 and 4 present results for *Non-Productive loans*. We partition the loan sample based on whether the financing is provided for productive uses (*Productive Loans*) such as business expansion, asset acquisition, and working capital needs or other purposes such as consumption needs, marriage and festival expenses (*Nonproductive Loans*). The dependent variable, *NetDisburse/Inc*, is the natural logarithm of one plus total net productive(non-productive) loan amount disbursed divided by household income of each borrower. For each borrower, we compute net productive (non-productive) loan amount disbursed as the total productive(nonproductive) loan amount disbursed minus any repayment made by the end of the calendar year 2014. We include villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All specifications include state and threshold fixed effects and baseline borrower-level controls for age, land ownership, household assets, education, and gender. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. We report bootstrapped standard errors below point estimates.

	ProductiveLoans		Non-ProductiveLoans	
	(1)	(2)	(3)	(4)
Bandwidth	±200	±250	±200	±250
Above cutoff	0.043***	0.045***	-0.044***	-0.038***
	(0.011)	(0.010)	(0.011)	(0.009)
Control group mean	0.047	0.047	0.066	0.067
State fixed effects	Yes	Yes	Yes	Yes
Threshold fixed effects	Yes	Yes	Yes	Yes
Observations	1,032	1,084	1,032	1,084

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Testing for the variability of loan contract terms

The table presents results that test the variability of loan contract terms between similar groups of borrowers in above- vs. below-cutoff villages. We divide our sample into groups of borrowers within above- and belowcutoff villages. For each state in our sample, we divide groups based on observable characteristics such as gender, whether the borrower is educated, household asset size, and age. We generate four groups based on gender and school education, and within each of these groups, we further create three groups based on household asset size and borrower age. Within each of these groups, we compute the coefficient of variation of loan contract terms: loan amounts, Interest rates, maturity. We then test for the difference in the coefficient of variation of loan contract terms between above and below-cutoff villages. Columns 1 and 2 report the coefficient of variation within below and above-cutoff villages respectively. Next two columns report difference in means and p-value on tests for equality of means respectively. Panel A reports values for villages within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while panel B expands the sample to include villages within 250 of the population threshold.

Panel A: Bandwidth, ±200

	Below Threshold	Above Threshold	Difference of means	p-value on difference

	(1)	(2)	(1) - (2)	
Scaled loan amount	0.619	0.635	-0.016	0.894
Interest rate	0.273	0.217	0.056	0.704
Loan maturity	0.103	0.095	0.008	0.869

Panel B: Bandwidth, ±250

	Below Threshold	Above Threshold	Difference of means	p-value on difference
	(1)	(2)	(1) - (2)	
Scaled loan amount	0.605	0.682	-0.078	0.512
Interest rate	0.249	0.249	0.001	0.997
Loan maturity	0.092	0.098	-0.006	0.895

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

Table 10: Impact of new roads, by borrower characteristics

The table presents reduced form estimates of the heterogeneous effects of new rural roads by borrower characteristics for the sample of villages. Columns 1 through 3 present reduced form estimates for villages within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while columns 4 through 6 present reduced form estimates expanding the sample to include villages within 250 of the population threshold. The dependent variable in column 1 and 4 is the net loan amount disbursed divided by the household income for each borrower. The dependent variable in columns 2 and 5 is the fraction of the loan amount disbursed that was overdue while in columns 3 and 6 it is the average interest rate across loans for each borrower. For each borrower, we compute the net loan amount disbursed as the loan amount disbursed minus any repayment made by the end of the calendar year 2014. We interact Above cutoff with the following characteristics: *Age (years)* a continuous variable that captures the age of the borrower in years at the time of opening the bank account, *Low assets* a dummy variable that takes the value of one if the borrower's household assets at the time of opening the bank account is below the sample median or zero otherwise, *School education* a dummy variable that takes the value of one if the borrower has School education at the time of opening a bank account or zero otherwise, *SC/ST/OBC* an indicator for the whether the borrower belongs to any of the minority sub-groups (Scheduled caste, Scheduled Tribe, or Other Backward Castes), and *Female* an indicator for whether the gender of the borrower is female. Our sample consists of individuals from the sample of villages in Odisha and Uttarakhand who had a loan with the bank by the end of the calendar year 2014. All specifications include loan purpose, state, and threshold fixed effects. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. We report bootstrapped standard errors below point estimates.

	± 200			± 250		
	(1)	(2)	(3)	(4)	(5)	(6)
	Loan Amount	%ODAmount	AvgIntRate	Loan Amount	%ODAmount	AvgIntRate
Above cutoff	-0.020	0.664	0.012	-0.013	0.555	0.007
	(0.026)	(1.344)	(0.016)	(0.025)	(1.240)	(0.016)
Age (years)	-0.001**	-0.000	-0.000	-0.000	0.002	-0.000
	(0.000)	(0.004)	(0.000)	(0.000)	(0.004)	(0.000)
Low assets	0.004	0.244	-0.000	0.004	0.228	-0.000
	(0.007)	(0.243)	(0.003)	(0.007)	(0.237)	(0.003)
School education	-0.006	-0.921	0.000	-0.004	-0.811	-0.001
	(0.007)	(0.798)	(0.004)	(0.006)	(0.726)	(0.004)
SC/ST/OBC	-0.002	0.166	0.000	-0.002	0.146	0.000
	(0.008)	(0.186)	(0.003)	(0.008)	(0.170)	(0.003)
Female	-0.015***	0.282	0.006*	-0.016***	0.248	0.005*
	(0.005)	(0.280)	(0.003)	(0.005)	(0.262)	(0.003)
Above cutoff x Age (years)	0.001	-0.019	-0.000	0.000	-0.020	-0.000
	(0.000)	(0.021)	(0.000)	(0.000)	(0.020)	(0.000)
Above cutoff x Low assets	0.021**	-0.206	-0.005	0.020**	-0.209	-0.004
	(0.010)	(0.233)	(0.004)	(0.010)	(0.213)	(0.004)
Above cutoff x School education	0.015	0.902	0.007	0.012	0.750	0.008
	(0.011)	(0.788)	(0.009)	(0.010)	(0.697)	(0.008)

Above cutoff x SC/ST/OBC	-0.011	-0.563	-0.006	-0.008	-0.508	-0.007
	(0.011)	(0.393)	(0.005)	(0.010)	(0.362)	(0.005)
Above cutoff x Female	0.001	-0.464	-0.009*	0.002	-0.439	-0.008
	(0.010)	(0.345)	(0.005)	(0.010)	(0.317)	(0.005)
Control group mean	0.083	0.12	0.15	0.085	0.12	0.15
Loanpurpose fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Threshold fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,032	630	630	1,084	665	665

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Macro evidence: New rural roads and bank credit

The table examines the relationship between new rural roads on private sector bank credit extended for the period 2004 to 2012. Our sample consists of districts from the 19 states for which we have non-missing control variables. Panel A presents estimates from a specification that excludes district-level time varying control variables while panel B presents estimates from a specification which includes district-level time varying covariates. Across both panels, the dependent variable in columns 1 (and 4), Δ *Log(Credit)*, is the annual difference in the natural logarithm of one plus total rural(urban) bank credit extended by private sector banks within a district over periods t+1 and t. In columns 2 and 5, the dependent variable, Δ *Log(Branches)*, is the annual difference in the natural logarithm of one plus the total number of rural(urban) private sector bank branches within a district over periods t+1 and t. In columns 3 and 6, the dependent variable, Δ *Log(Deposits)*, is the annual difference in the natural logarithm of one plus total rural (urban) private bank deposits within a district over periods t+1 and t. For each state, we aggregate the total kilometers of road constructed under PMGSY at the district-level. *Log(Sum road* $_{t-3,t-1}$*)*, is the natural logarithm of one plus sum of the length of new roads (in kilometers) constructed under PMGSY within a district over periods t-1, t-2 and t-3. The control variables include the total geographical area under land use, field wages for males, the literate population fraction, the vote margins for candidates from the two main political parties (i.e., the Bhartiya Janata Party (BJP) and the Congress Party), and the average vote margin difference between the candidates. All specifications include district and state × year fixed effects. All estimates are multiplied by 100 for ease of interpretation. Standard errors are clustered at the district-level and are reported below point estimates.

	PanelA:Withoutcontrols					
	ΔLog(1+rural...)			ΔLog(1+urban...)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Credit	Branches	Deposit	Credit	Branches	Deposit
Log(Sum road $_{t-3,t-1}$)	0.466*	-0.071	0.705*	0.192	-0.058	0.105
		(0.223)	(0.330)	(0.561)	(0.242)	(0.45[illegible])
		6.56	7.84	12.7	11.2	11.8
Controls	Yes	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,524	2,524	2,524	2,524	2,524	2,524
	PanelB:Withcontrols					
Dependentvariable	ΔLog(1+rural...)			ΔLog(1+urban...)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Credit	Branches	Deposit	Credit	Branches	Deposit
Log(Sum road $_{t-3,t-1}$)	0.465*	-0.088	0.684*	0.207	-0.059	0.206
		(0.224)	(0.331)	(0.570)	(0.242)	(0.46[illegible])
		6.56	7.84	12.7	11.2	11.8
Controls	Yes	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

State × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,524	2,524	2,524	2,524	2,524	2,524

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Macro evidence: New rural roads, bank credit & economic growth

The table examines the relationship between new rural roads on economic growth of the districts for the period 2004 to 2012. Our sample consists of districts from the 19 states for which we have non-missing control variables. Panel A presents estimates from a specification that excludes district-level time varying control variables while panel B presents estimates from a specification which includes district-level time varying covariates. Across both panels, The dependent variable in columns 1 through 3, Δ *Log(GDP)*, is the annual difference in the natural logarithm of gross domestic product (GDP) for each district over periods t+1 and t. We present the estimates for Overall GDP, GDP for Agriculture, and GDP for Industry & Services. For each state, we aggregate the total kilometers of road constructed under PMGSY at the district-level. *Log(Sum road*$_{t-3,t-1}$*)*, is the natural logarithm of one plus sum of length of new roads (in kilometers) constructed under PMGSY within a district over periods t-1, t-2 and t-3. *High (Low) rural credit* is defined based on whether the rural credit per capita in a given year is above (below) the median rural credit per capita. The control variables include the total geographical area under land use, field wages for males, the literate population fraction, the vote margins for candidates from the two main political parties (i.e., the Bhartiya Janata Party (BJP) and the Congress Party), and the average vote margin difference between the candidates. All specifications include district and state × year fixed effects. All estimates are multiplied by 100 for ease of interpretation. Standard errors are clustered at the district-level and are reported below point estimates.

	PanelA:Withoutcontrols		
Dependentvariable	ΔLog(GDP...)		
	(1)	(2)	(3)
	Overall	Agriculture	Industry & Services
Log(Sum road $_{t-3,t-1}$) × High rural credit	0.203**	0.519**	0.023
	(0.086)	(0.248)	(0.104)
Log(Sum road $_{t-3,t-1}$) × Low rural credit	-0.052	0.280	-0.072
	(0.122)	(0.327)	(0.112)
Mean of dep. var.	7.32	6.25	6.54
Controls	No	No	No
District fixed effects	Yes	Yes	Yes
State x year fixed effects	Yes	Yes	Yes
Observations	2,766	2,766	2,766

	PanelB:Withcontrols		
Dependentvariable	ΔLog(GDP...)		
	(1)	(2)	(3)
	Overall	Agriculture	Industry & Services

Log(Sum road $_{t-3,t-1}$) × High rural credit	0.189**	0.496*	0.016
	(0.087)	(0.253)	(0.106)
Log(Sum road $_{t-3,t-1}$) × Low rural credit	-0.069	0.237	-0.078
	(0.121)	(0.332)	(0.114)
Mean of dep. var.	7.32	6.25	6.54
Controls	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes
State x year fixed effects	Yes	Yes	Yes
Observations	2,766	2,766	2,766

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Internet Appendix: Additional tables

Table IA1: Covariate balance

The table presents mean values for baseline village characteristics, as recorded in 2001 Population Census. Panel A reports balance for all unconnected villages in Odisha and Uttarakhand while panel B reports balance for villages in the bank loan sample. In both panels, columns 1 and 2 present the unconditional means for villages below the treatment threshold, and villages above the treatment threshold, respectively. Column 3 presents the difference in means between villages below the treatment threshold and villages above the treatment threshold. Additionally, in panel A, column 4 shows the regression discontinuity estimate, following the main estimating equation, of the effect of being above the treatment threshold on the baseline variable and column 5 is the p-value for this estimate, using heteroskedasticity robust standard errors.

	PanelA:UnconnectedvillagesinOdishaandUttarakhand				
	Below	Above	Difference	RD estimate	p-value on estimate
	(1)	(2)	(3)	(4)	(5)
Primary school	0.85	0.90	-0.05	-0.01	0.34
Primary health centres	0.07	0.11	-0.04	-0.01	0.18
Telegraph office	0.00	0.00	-0.00	-0.00	0.30
Electricity	0.71	0.78	-0.07	-0.01	0.50
Scheduled caste share	49.43	45.39	4.04	-1.31	0.28
Irrigated land	0.28	0.34	-0.06	-0.03	0.41
Distance from nearest town (in kms)	32.30	30.75	1.55	-0.38	0.67
Observations	6,719	4,417			

	PanelB:Bankloansample			
	Below	Above	Difference	p-value on difference
	(1)	(2)	(3)	(4)
Primary school	0.95	1.00	-0.05	0.16
Primary health centres	0.05	0.00	0.05	0.16
Telegraph office	0.03	0.00	0.03	0.32
Electricity	0.92	0.90	0.01	0.86
Scheduled caste share	23.53	17.09	6.43	0.16
Irrigated land	1.92	2.07	-0.15	0.77
Distance from nearest town (in kms)	27.57	29.10	-1.53	0.77
Observations	37	21		

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA2: First stage effect of road priority on PMGSY road treatment: (2009 - 2014)

The table presents first stage estimates from Equation 1 of the effect of being above the population threshold on a village's probability of receiving a road under PMGSY by 2014. The dependent variable is an indicator variable that takes on the value one if a village has received a PMGSY road before 2014. Column 1 presents results for villages within 100 of the population threshold (400-600 for the 500 threshold and 900-1100 for the 1000 threshold) while column 2 expands the sample to include villages within 150 of the population threshold. The regression specification includes state and threshold fixed effects. The sample consists of all the villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. We report bootstrapped standard errors below point estimates.

Bandwidth	(1) ±100	(2) ±150
Above cutoff	0.081***	0.065***
	(0.016)	(0.013)
Control group mean	0.13	0.12
F-statistic	25.13	24.32
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	5,537	8,246

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA3: Covariate balance for matching variables: Extensive margin

The table presents mean values for baseline village characteristics used in propensity score matching for villages in the bank loan sample and their counterfactuals constructed using propensity score matching. Specifically, we require the control group villages to be in the same block and match them on the following village-level covariates as recorded in the 2001 Population Census: the presence of a primary school, village population, the fraction of SC/ST population, and distance from the nearest town. Columns 1 and 2 present the unconditional means for villages where the bank never entered and villages where the bank enter during the sample period, respectively. Column 3 presents the difference in means, while column 5 reports the p-value on the difference.

	No Bank	Bank	Difference	p-value
	(1)	(2)	(3)	(4)
Primary school	0.97	0.91	0.06	0.25
Population	530.40	562.71	-32.31	0.49
Scheduled caste share	20.14	21.15	-1.01	0.78
Log (1+distance to nearest town)	3.08	3.10	0.02	0.91
Observations	58	58		

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA4: Impact of new roads on the Lending quantities: Unscaled dependent variable

The table presents reduced form estimates from Equation 2 of the effect of new rural roads on lending activity within the villages. Column 1 presents reduced form estimates for villages within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while column 2 presents reduced form estimates expanding the sample to include villages within 250 of the population threshold. The dependent variable, *NetDisburse*, is the natural logarithm of one plus total net loan amount disbursed. For each borrower, we compute the net loan amount disbursed as loan amount disbursed minus any repayment made by the end of the calendar year 2014. Our bank loan sample consists of individuals who had a loan with the bank by the end of the calendar year 2014. We include villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. The specification also includes baseline borrower-level controls for age, land ownership, level of household assets, education, gender, and household income. All specifications include state and threshold fixed effects. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. We report bootstrapped standard errors below point estimates.

Bandwidth	± 200		± 250	
	(1)	(2)	(3)	(4)
Above cutoff	1.530**	1.150*	1.159*	0.939*
	(0.666)	(0.601)	(0.621)	(0.557)
Age (years)		-0.032**		-0.032**
		(0.013)		(0.012)
Land		0.902***		0.916**
		(0.347)		(0.357)

Log (1+assets)		0.117*		0.108*
		(0.060)		(0.057)
School education		0.683**		0.671**
		(0.342)		(0.332)
Female		-2.772***		-2.647***
		(0.321)		(0.303)
Log (1+income)		1.358***		1.376***
		(0.238)		(0.232)
Control group mean	6.33	6.33	6.44	6.44
State fixed effects	Yes	Yes	Yes	Yes
Threshold fixed effects	Yes	Yes	Yes	Yes
Observations	1,032	1,032	1,084	1,084

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA5: First stage effect of road priority on PMGSY road treatment, robust standard errors

The table presents first stage estimates from Equation 1 of the effect of being above the population threshold on a village's probability of receiving a road under PMGSY by 2014. The dependent variable is an indicator variable that takes on the value one if a village has received a PMGSY road before 2014. Column 1 presents results for villages within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while column 2 expands the sample to include villages within 250 of the population threshold. The regression specification includes state and threshold fixed effects. The sample consists of all the villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. We report heteroscedasticity standard errors below point estimates.

	(1)	(2)
Bandwidth	±200	±250
Above cutoff	0.068***	0.066***
	(0.011)	(0.010)
Control group mean	0.12	0.11
F-statistic	39.08	45.82
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	11,136	14,205

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA6: Impact of new roads on lending activity, robust standard errors

The table presents estimates from reduced form estimates of the effect of new rural roads on lending activity within villages. Panels A presents odds ratio from logit estimation as in panel A of Table 3 while panel B present estimates for Table 4. Panel C present estimates of Table 5 while panels D presents estimates of Table 6. Panels E presents estimates of Table 8. The dependent variable in panel A, *ExtMargin*, is an indicator variable that takes on the value one if an individual in the village received a loan from the bank. We construct the control group villages using propensity score matching. Specifically, we require the control group villages to be in the same block and match them on the following village-level covariates as recorded in the 2001 Population Census: fraction of SC/ST population, village population, presence of primary school, and distance from the nearest town. Internet Appendix Table IA 2 presents the covariate balance. The dependent variable for panels B and E, *NetDisburse/Inc*, is the net loan amount disbursed divided by household income of each borrower. For each borrower, we compute the net loan amount disbursed as loan amount disbursed minus any repayment made by the end of the calendar year 2014. We measure loan performance using two measures: (1) % Overdue amount captures the fraction of loan amount disbursed that was overdue (2) Total loan amount that was overdue. The dependent variable in columns 1 and 4 of panel C is natural logarithm of loan maturity. In columns 2 and 5, the dependent variable is Total Overdue amount while in columns 3 and 6 it is % Overdue amount. The dependent variable in panel D is the average interest rate across loans for each borrower. Our sample consists of individuals who had a loan with the bank by the end of the calendar year 2014. We include villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All specifications include state and threshold fixed effects and baseline borrower-level controls for age, land and asset ownership, education and gender. Panel A reports the odds ratio from a logit framework for estimation while all the remaining panels use Ordinary Least Squares (OLS) estimation and reports the coefficient estimates. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. We report heteroscedasticity standard errors below point estimates.

	PanelA:Extensivemargin	
Bandwidth	± 200	± 250
	(1)	(2)
Above cutoff	2.003**	1.733**
	(0.875)	(0.781)
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	93	116

	PanelB:Lendingactivity	
Bandwidth	±200	±250
	(1)	(2)
Above cutoff	0.025**	0.030***
	(0.012)	(0.011)
Control group mean	0.083	0.085
Controls	Yes	Yes
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	1,032	1,084

Continued...

	Panel C: Loan maturity and quality					
	±200			±250		
Bandwidth	(1)	(2)	(3)	(4)	(5)	(6)
	Ln(Maturity)	OD Amount	%OD Amount	Ln(Maturity)	OD Amount	%OD Amount
Above cutoff	-0.009	-164.872	0.057	-0.028	-184.306	-0.066
	(0.020)	(205.843)	(0.350)	(0.019)	(189.697)	(0.360)
Control group mean	1.11	104.7	0.12	1.11	100.6	0.12
Loanpurpose fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Threshold fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	630	630	630	665	665	665

	Panel D: Interest rates	
Bandwidth	± 200	± 250
	(1)	(2)
Above cutoff	-0.002	-0.005
	(0.006)	(0.005)
Control group mean	0.15	0.15
Loanpurpose fixed effects	Yes	Yes
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	630	665

	Panel E: Lending quantity by loan type			
	Productive Loans		Non-Productive Loans	
	(1)	(2)	(3)	(4)
	±200	±250	±200	±250
Above cutoff	0.043***	0.045***	-0.044***	-0.038***
	(0.011)	(0.010)	(0.011)	(0.009)
Control group mean	0.047	0.047	0.066	0.067
State fixed effects	Yes	Yes	Yes	Yes
Threshold fixed effects	Yes	Yes	Yes	Yes
Observations	1,032	1,084	1,032	1,084

Standard errors in parentheses

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

Table IA7: Impact of new roads on lending activity, stratified bootstrapping

The table presents estimates from reduced form estimates of the effect of new rural roads on lending activity within villages. Panels A presents odds ratio from logit estimation as in panel A of Table 3 while panel B present estimates for Table 4. Panel C present estimates of Table 5 while panels D presents estimates of Table 6. Panels E presents estimates of Table 7. The dependent variable in panel A, *ExtMargin*, is an indicator variable that takes on the value one if an individual in the village received a loan from the bank. We construct the control group villages using propensity score matching. Specifically, we require the control group villages to be in the same block and match them on the following village-level covariates as recorded in the 2001 Population Census: fraction of SC/ST population, village population, presence of primary school, and distance from the nearest town. Internet Appendix Table IA 2 presents the covariate balance. The dependent variable for panels B and E, *NetDisburse/Inc*, is the net loan amount disbursed divided by household income of each borrower. For each borrower, we compute the net loan amount disbursed as loan amount disbursed minus any repayment made by the end of the calendar year 2014. We measure loan performance using two measures: (1) % Overdue amount captures the fraction of loan amount disbursed that was overdue (2) Total loan amount that was overdue. The dependent variable in columns 1 and 4 of panel C is natural logarithm of loan maturity. In columns 2 and 5, the dependent variable is Total Overdue amount while in columns 3 and 6 it is % Overdue amount. The dependent variable in panel D is the average interest rate across loans for each borrower. Our sample consists of individuals who had a loan with the bank by the end of the calendar year 2014. We include villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All specifications include state and threshold fixed effects and baseline borrower-level controls for age, land and asset ownership, education and gender. Panel A reports the odds ratio from a logit framework for estimation while all the remaining panels use Ordinary Least Squares (OLS) estimation and reports the coefficient estimates. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. Bootstrap samples are taken independently within each village and bootstrapped standard errors are reported below point estimates.

	PanelA:Extensivemargin	
Bandwidth	± 200	± 250
	(1)	(2)
Above cutoff	2.003***	1.733***
	(0.000)	(0.000)
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	93	116

	PanelB:Lendingactivity	
Bandwidth	±200	±250
	(1)	(2)
Above cutoff	0.025***	0.030***
	(0.010)	(0.009)
Control group mean	0.083	0.085
Controls	Yes	Yes
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	1,032	1,084

Continued...

Bandwidth	PanelC:Loanmaturityandquality					
	±200			±250		
	(1)	(2)	(3)	(4)	(5)	(6)
	Ln(Maturity)	ODAmount	%OD Amount	Ln(Maturity)	ODAmount	%OD Amount
Above cutoff	-0.009	-164.872	0.057	-0.028*	-184.306	-0.066
	(0.017)	(184.396)	(0.349)	(0.016)	(175.312)	(0.347)
Control group mean	1.11	104.7	0.12	1.11	100.6	0.12
Loanpurpose fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Threshold fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	630	630	630	665	665	665

	PanelD:Interestrates	
Bandwidth	± 200	± 250
	(1)	(2)
Above cutoff	-0.002	-0.005
	(0.005)	(0.005)
Control group mean	0.15	0.15
Loanpurpose fixed effects	Yes	Yes
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	630	665

	PanelE:Lendingquantitybyloantype			
	ProductiveLoans		Non-ProductiveLoans	
	(1)	(2)	(3)	(4)
	±200	±250	±200	±250
Above cutoff	0.043***	0.045***	-0.044***	-0.038***
	(0.010)	(0.009)	(0.008)	(0.007)
Control group mean	0.047	0.047	0.066	0.067
State fixed effects	Yes	Yes	Yes	Yes
Threshold fixed effects	Yes	Yes	Yes	Yes

Observations	1,032	1,084	1,032	1,084

Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table IA8: Impact of new roads on lending activity, robustness

The table presents robustness on the reduced form estimates from Equation 2 of the effect of new rural roads on lending activity within these villages. Panel A present results wherein we drop four villages in Uttarakhand with habitations. Panel B restricts the sample to villages such that above and below cutoff classifications do not differ between PMGSY population and Census 2001 population. Panel C presents results with same slopes on either side of the population threshold and different intercept around the cutoff while panel D allows for same slope and intercept around the population threshold. Panel E presents results from baseline Table 4 without winsorizing our dependent variable while panel F presents results restricting the sample to new borrowers. In all panels, the dependent variable, *NetDisburse/Inc*, is the net loan amount disbursed divided by household income of each borrower. For each borrower, we compute the net loan amount disbursed as loan amount disbursed minus any repayment made by the end of the calendar year 2014. Our sample consists of individuals who had a loan with the bank by the end of the calendar year 2014. We include villages in Odisha and Uttarakhand that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All specifications include state and threshold fixed effects and baseline borrower-level controls for age, land ownership, household assets, education and gender. Panel A reports the odds ratio from a logit framework for estimation while all the remaining panels use Ordinary Least Squares (OLS) estimation and reports the coefficient estimates. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. We report bootstrapped standard errors below point estimates.

	PanelA:Dropvillageswithhabitations	
Bandwidth	±200	±250
	(1)	(2)
Above cutoff	0.024**	0.029***
	(0.012)	(0.011)
Control group mean	0.082	0.084
Controls	Yes	Yes
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	1,019	1,071

	Panel B: Same slope and different intercept around the cutoff	
Bandwidth	± 200	± 250
	(1)	(2)
Above Cutoff	0.028**	0.030***
	(0.012)	(0.011)
Control group mean	0.083	0.085
Controls	Yes	Yes
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	1,032	1,084

Continued...

	PanelC:Sameslopeandinterceptaroundthecutoff	
Bandwidth	± 200	± 250
	(1)	(2)
Above Cutoff	0.028**	0.030***
	(0.012)	(0.011)
Control group mean	0.083	0.085
Controls	Yes	Yes
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	1,032	1,084

	PanelD:Nowinsorization	
Bandwidth	±200	±250
	(1)	(2)
Above cutoff	0.025**	0.030***
	(0.012)	(0.011)
Control group mean	0.084	0.086
Controls	Yes	Yes
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	1,032	1,084

	PanelE:Rulingoutevergreening	
Bandwidth	±200	±250
	(1)	(2)
Above cutoff	0.023*	0.028**
	(0.013)	(0.011)
Control group mean	0.079	0.081
Controls	Yes	Yes
State fixed effects	Yes	Yes
Threshold fixed effects	Yes	Yes
Observations	959	1,005

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$