Roads and Loans

Sumit Agarwal, Abhiroop Mukherjee and S. Lakshmi Naaraayanan*

January 2018*

Abstract

Does the flow of financing respond to productivity shocks even in the day-to-day lives of the world's poor? We answer this question by examining the response of bank financing to a shock to rural road connectivity in India, which improved labor productivity. The program prioritized road connectivity for villages above explicit population thresholds, thereby allowing us to exploit discontinuities in the probability of treatment to identify our effects. We find large financing responses to rural road connectivity – 54% more villagers receive loans, and the average amount lent to them is 15-20% higher, for villages right above the threshold compared to those just below. Roads seem to disproportionately benefit those villagers who lack assets or income, but have basic education.

JEL classification: G21, G28, O16, O18

Keywords: Bank Lending, Household Finance, India, Economic Development

*Agarwal: Georgetown University; Mukherjee and Naaraayanan: Hong Kong University of Science and Technology. We are grateful to Nicholas Barberis, Utpal Bhattacharya, Chiman Cheung, Vidhi Chhaochharia, Darwin Choi, Sudipto Dasgupta, Hanno Lustig, Christopher Malloy, Wenlan Qian, Sheridan Titman, Baolian Wang, Daniel Wolfenzon, and Rik Sen 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 basic tenets of finance is that money should flow to its most productive uses. So when productivity changes, the flow of financing should respond in adjustment. Naturally, then, a very large literature is devoted to understanding this issue; and this literature has indeed been helpful in understanding and influencing very important policy issues of our time. But does financing really respond to productivity changes – not just for the largest corporations or the most well-off households in countries with developed financial markets – but also in the day-to-day lives of the world's poor? In this paper we use a shock to productivity, arising from a large road-building initiative in India, to shed light on this issue.

In many parts of the world, infrastructure projects – in particular roads – are thought to be key to unlocking productivity increases among the growing population of surplus laborers in agrarian villages. Hundreds of thousands of miles of such roads are being built in Asia, Africa, South America and in Eastern Europe. A large chunk of China's trillion dollar 'One Belt One Road' project – branded the biggest investment project in history – is devoted to such road building. But the type of productive opportunities policy makers often talk about as examples of trickle down benefits – for example, opening or expansion of small businesses like village grocery shops, or changing crop patterns from subsistence cereal farming to more profitable market-based crops – very often require availability of financing (e.g., King and Levine, 1993). It is typically assumed in much of the policy discourse that such financing to households will automatically follow once roads are built, facilitating the best use of these new productive opportunities.

Many rural and agrarian economies, however, suffer from chronic problems of financing. Formal finance is largely absent in many parts of the world, and informal money lenders charge usurious interest rates. So a central question today in finance is: Will financing indeed flow in response to changing productive opportunities created by these mega infrastructure projects? 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 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, perhaps due to their lack of collateralizable assets, but can now find a way in?

One reason behind the lack of more research on this area – despite its clear importance in terms of welfare implications for a major chunk of world population – is the difficulty in accurately identifying the 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. Then one could compare changes in the treated and untreated areas to estimate the impacts of the program. Infrastructure, of course, is not placed randomly in practice, making comparisons with untreated areas problematic.

Our paper finds a way to progress by exploiting a policy directive surrounding a major public road construction program in India. The objective of this nationwide program was to provide all-weather road connectivity to hitherto unconnected villages. Aggarwal (2015) and Shamdasani (2016) examines 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 a transition from subsistence to market-oriented farming. We examine whether some of these productivity improvements were accompanied by financing flows – as theory predicts – and also bring in to focus the distributional impacts of infrastructure-building.

The roads program we study created a nearly random comparison group for policy evaluation, by explicitly focused on building new roads to connect all villages above explicit population thresholds at a point in time. By doing so, program rules created discontinuities in the probability of treatment at multiple village population thresholds, which we exploit using a discontinuity design. 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 quasi-random. The results of our analysis thus provide convincing estimates of the impact of connectivity on the access to, and – given our data on repayments – productive utilization of, formal financial services.

Our empirical analysis is made possible by our access to a proprietary loan-level dataset from one of India's largest private lenders on loans made in the largely rural district of *Ganjam* in the eastern state of *Odisha*. The fact that the data is from a *private* lender is of special interest for the following reason. Governments financing some of these infrastructure projects face budget constraints - and this is especially true in poorer countries - so that they often cannot both finance infrastructure-building and provide loan financing to reap its productivity benefits. This creates the need to examine whether private sector financiers will indeed provide financing to respond to the predicated new productive opportunities in a way that theory - and policy makers - often expect them to (e.g., Beck, 2012). If private profit motives result in flow of increased financing to areas with recently improved infrastructure, then not only will this signal good news for the feasibility of government infrastructure projects actually reaping benefits for the target population, but also indicate an organic opportunity for the type of public-private partnership that Demirguc-Kunt and Levine (2008) and Carter et al. (2013) talk about.

We begin our analysis by showing that a previously unconnected village right above the population cutoff is about 45% more likely to get a road than one right below in our bank loan sample. By 2014, when we measure differences in loan outcome, every one of the 13 villages above the cutoff that were not connected when the bank started lending get connected already. There is no evidence of population criteria manipulation with the cutoffs of 500 and 1000 we use, and villages above and below cutoff do not show any stark discontinuous patterns with respect to a variety of other characteristics such as electricity connections, presence of telegraph offices etc.

Next, we examine the effect of connectivity on financing outcomes, using the population-based discontinuity as our instrument. Our evidence suggests that these effects are economically large. Our bank lends to 5.1% of villagers living in unconnected villages; this coverage jumps by over 50% to about 8-9% for villages just above the population threshold. Total loan disbursement is 15% higher for villages above the threshold, and lending net of repayment shows a similar jump of around 28%. These latter results are after controlling for the entire granular set of borrower characteristics that the bank cares about and collects hard information on; therefore, they are economically large. Our results are robust to a variety of specifications, including the use of different functional forms and econometric techniques in our population-based discontinuity specification.

In the next set of tests, we exploit our detailed data structure to examine loan characteristics within each loan type (loan types are identified by loan use in our data). We find that loans given to villagers in connected villages are similar in terms of default probability – even somewhat *less likely* to default. These loans have very similar maturity (about a year for most loans in our data). Next, we impute interest rates on loans granted by jointly examining the loan cash-flows (we have repayment installment data), grant amounts, and maturities. Interest rates charged on similar loans across the cut-off are do not differ, either in terms of economic magnitude or statistical significance.

Next, we dig deeper within our loans data. Interestingly, we uncover evidence that almost all our results on loan amounts being higher above the population cutoffs come from productive loans, which are loans taken out – as classified by our bank – for small and micro business expansions, asset acquisition, working capital needs and such. On the other hand there is some evidence of *lower* loan amounts granted for consumption uses such as those taken out to finance consumption needs, marriage and festival expenses. This is consistent with some reallocation of credit, from consumption uses to productive uses, when road connectivity improves. This particular result is informative on wealth effects. Our baseline increases in financing to connected villages is also consistent with a wealth effect story: if wealth increased for connected villagers relative to unconnected ones, and this was driving our results instead of increased availability of loans to finance newly productive activities, we would have expected consumption loans to also respond. This is not the case. Further, within productive loans, there is evidence of increases in all types, but our results are strongest economically for micro enterprise loans and crop loans. This seems consistent with policy makers' views on benefits flowing in the form of opening or expansion of small businesses like village grocery shops, or changing crop patterns from subsistence to more profitable market-based crops.

We next perform a placebo test, to show that our effects are not coming from something mechanical associated with our population cutoffs in our data, such as from fundamental differences between above- and below-cutoff villages. Here, instead of looking at a sample of unconnected villages around the cutoff – villages that actually get productivity shocks in our sample in terms of road connectivity – we look at *already connected* villages around the same cutoffs. None of our results obtain in this sample, either in terms of economic magnitude or statistical significance, the latter in spite of improved power from a larger sample size in this test.

Finally, we examine the distributional consequences of connectivity from our lending sample. This is a critical step in understanding the trickle-down effects of development, as well as for the financial inclusion literature (Beck and Demirguc-Kunt, 2008).It might be helpful to think of these tests in terms of backing out – from financing data – which pockets of population benefit more from road connectivity. Our data allows us to focus on individual-level differences. Here 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). More suggestive evidence supporting this view comes from the fact that more benefits accrue to those without assets typically used as collateral in rural India – land and jewellery. Our evidence, however, 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 existing land – used as collateral – and therefore their borrowing capacity, our effects would have been stronger among the landed.

Interestingly, we also find that those with some school education and those with lower incomes seem to benefit more from connectivity. Our evidence on education also provides clear and direct justification for results in Mukherjee (2011) and Adukia et al. (2017), who show that connectivity increases school enrollment, and that children stay in school longer in connected villages. If villagers saw benefits of new infrastructure accrue more visibly to the school-educated, they would be encouraged to invest more in education.

Note however that while our setting buys us the advantage of relatively clean identification, we cannot distinguish between whether the effect we observe is a demand-side story (as roads increase productivity villagers demand more loans) or a supply-side story (the bank actively seeks out lending opportunities to newly connected villagers), given the reduced form nature of what we do. Talking to bank officials however, indicates more support for a demand-side story. 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 lower than the local money lenders', 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 that the bank does not quite need to actively go out to these villages and market. Instead most of these branches employ people from surrounding villages; these people spread the word on the bank's operations to their neighbors.

Our paper contributes to the growing literature on the role of financing in development (King and Levine, 1993, Black and Strahan, 2002, Burgess and Pande, 2004, Levine, 2005, Beck, Demirguc-Kunt and Peria, 2007, Demirguc-Kunt and Levine, 2008a, 2008b, Beck, 2012, Demirguc-Kunt, Feyen, and Levine, 2013, Beck, Lu and Yang, 2014, among others).

Many of these papers have established important results on the effect 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; instead, we look at whether private-sector financing *responds* in the way many policy makers expect it to when productive opportunities expand through improved connectivity. In this sense our paper is close to a contemporaneous paper by Das et al. (2017), who examine aggregate industrial financing by district around a different road construction program in India. Both our papers find similar evidence that financial flows appear to be going to regions seeing connectivity improvements, but differ in terms of the type of program and the type of financing we examine. While Das et al. (2017) examine financing for industries around the upgradation of a major road network connecting India's major cities, we look at the availability of financing for poor households that get access to a proper ("all-weather") road for the first time.

We also contribute 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 proprietary rural bank dataset allows us to focus on detailed financing outcomes.

Also, while there is a strong sense among policy makers and the media that investing in infrastructure leads to economic growth, 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 cross-country empirical studies often find weak and suggestive evidence that infrastructure reduces inequality. Within-country studies, however, offer mixed evidence. For example, Khandker, Bakht, and Koolwal (2009), find that the poorest households benefitted the most from road improvement projects in Bangladesh, whereas Artadi and Sala-i-Martin (2004) find that infrastructure spending may have contributed to income inequality in Africa. They show that income inequality is associated with corruption and tunneling of funds designated for infrastructure. Similarly, Khandker and Koolwal (2010) find that richer households benefitted more from increased access to paved roads than poorer households in Bangladesh. Given these mixed results, there is a clear need for a robust and well-identified empirical framework that could identify the impact of road construction on local inequality. Our paper fills in this gap, using a policy setting that allows clean identification through a population-based discontinuity design. The results of our analysis thus provide convincing estimates of the impact of connectivity on different village-level financing outcomes.

A population-based discontinuity design based on PMGSY has also been used by Mukherjee (2011) to examine school enrolments, and Asher and Novosad (2017) who show that PMGSY led to a reallocation of village labor from agriculture to wage labor. Aggarwal (2015) also uses PMGSY to examine adoption of agricultural technology and school dropout rates. Our paper is different from these studies on at least two counts: first, our outcome of interest is very different – how financing responds to productivity shocks in a setting that is novel, yet relevant to a substantial chunk of world population living in developing countries. Second, our unique individual-loan-level dataset allows us to study who benefits from such financing – which is an independently important issue from an inequality and development policy point of view.

The rest of the paper is organized as follows. Section 2 details our data, Section 3 describes our setting and how we exploit it in our empirical strategy, Section 4 presents our main results, Section 5 presents results on the distributional impact of roads, Section **??** discusses external validity and implications of our findings, and Section 6 concludes.

2 Data and Summary Statistics

2.1 Data

The main dataset we use is a proprietary rural-bank-account level dataset that we have obtained from one of India's largest publicly traded banks. One of the main obstacles limiting research of questions like ours is the lack of availability of granular financing data at the individual level, particularly in the case of rural areas. Our data is for the district of Ganjam in the eastern state of Odisha. This dataset contains information on individual accounts and transactions in loans over the period 2009-2014. Given our focus in this paper, we consider non-Joint Liability Group (JLG) loans. JLG loans are microfinance loans, classified by our bank as consumption loans.[1] The data also contains relatively detailed demographic information, such as the borrower's gender, education, assets, and income at the time of the bank account opening. The information on assets could be on the number of dwellings owned, the type of dwelling (brick or mud), number of livestock etc.

We obtain data on road construction in India, from the website of *Pradhan Mantri Gram Sadak Yojna* (PMGSY), the road-building program in India we study.[2] The data includes detailed information on road sanction and completion dates, which we use a program to scrape. The PMGSY data is structured to consist of information both at the habitation-level and at the road-level. The correspondence between habitation and roads is many-to-many, as several roads serve multiple habitations, and habitations may be connected to multiple roads. We define our treatment at the village level wherein a village is considered as "treated" under PMGSY if at least one habitation in the village – which was previously unconnected to the paved "all-weather" road network – received a (completed) road during our sample period. We conduct our analysis at the village level because many villages have only one habitation, and even in cases they do not, multiple habitations were often aggregated up to the village level for PMGSY road sanctions.[3]

We hand match the administrative road data to our proprietary dataset at the village level. We successfully match over 90% of habitations listed on the PMGSY website to their corresponding census villages. Overall, we are able to match 207 villages in 7 blocks across both the datasets.[4] We also use data on demographics and village-level amenities (such electricity, roads, schools etc) from the 2001 Population Census and the previously listed PMGSY webpage.

2.2 Summary Statistics

Table 1 shows the summary statistics for the sample of villages matched to the bank-account level dataset. The bank data we use contains cash-flow information on each loan granted. Loans can be bullet (one-time repayment), but are mostly structured to be paid back in

[1]Such a classification of microfinance lending is specific to our bank.
[2]The data is publicly available at http://omms.nic.in.
[3]A typical census village comprises of one to three habitations. Overall in 2001 Population Census, approximately one third of all villages consists of only a single habitation.
[4]In Census enumeration, Block is an administrative equivalent of Sub-district comprising of several villages or village clusters.

weekly installments, so we do not differentiate between categories (differentiating creates challenges with number of observations when we focus on our narrow set of villages within bandwidth). We present means and standard deviations (in parentheses) for our main variables of interest. The columns display results for borrowers in villages around the population threshold (for our pooled sample) and without a paved access road in 2001. Our bank loan sample consists of individuals who had a loan with the bank by the end of the calendar year 2014.

Panel A presents main loan characteristics observed in our dataset. Columns 1 through 3 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 4 through 6 presents means expanding the sample to include villages within 250 of the population thresholds. *Extensive Margin* is an indicator variable that takes on the value one if an individual in the village received a loan from the bank. *Loan amount* is the average total loan amount disbursed by the bank while *Net Disbursement* is the average nett loan amount disbursed. For each borrower, we compute the average net loan amount disbursed as the average of loan amount disbursed minus any repayment made by the end of the calendar year 2014. *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.

We see that around 5% of the village population received traditional (non-JLG) loans from the bank. Majority of these loans were given for productive uses; others were loans to finance consumption or repay village moneylenders. Although a majority was for productive uses, the average size of a consumption loan is still similar to a productive loan. The typical loan granted is between 32,000 to 35,000 Indian Rupees (slightly above USD 500). Loan maturity is about 2 years on average, while the average interest rate is 15%. Defaults are very rare in our sample, with only 0.4% 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 charges usurious interest rates.

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 borrowers incomes and assets; and within assets, on their land and jewellery holdings.

In terms of borrower characteristics, the average borrower in our sample is 38 years old. Productive loans are mostly given out to men. More than two-thirds who receive a loan have a high school education. Not surprisingly, this is a much higher level of education than the underlying population, which had an average literacy rate of 63% in the 2001 census. The average income for borrowers in our data is about 7416 Indian Rupees (slightly more than 110 USD) per month, translating to a little over 89,000 Indian rupees per year. Again, not surprisingly, this is slightly below the average for Ganjam district (Rs.12,742 per month in 2004-05, as per Ganjam District Plan Report 2010-11), because our sample is made up of relatively small villages, and therefore does not include residents of larger villages and the city of Behrampur, who have much higher incomes.

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 transport infrastructure. Factors such as political favoritism, and 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 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 the aim was to construct roads to connect them to the closest market center, identified by the program as the "center of activities for marketing agricultural produce and inputs, servicing of agricultural implements, health, higher education, postal, banking services etc."

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 treatment is determined by extraneous factors, 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 under the program on financing outcomes using a discontinuity design.

3.2 Empirical strategy

We first test for threshold manipulation under the PMGSY program. This is important to understand whether, for example, a powerful politician was systematically classifying villages with populations below the threshold as being above it, such 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. If this is the case however, one can potentially detect this 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 Ganjam 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 zoom in around these thresholds of 500 and 1000, and formally test the discontinuity of the running variable around the thresholds, following McCrary (2008). In panel (a), for the threshold of 500, the point estimate for the discontinuity is 0.469, with a standard error of 0.782 while in panel (b), for 1000, the estimate is -0.011 with a standard error of 0.598. To ensure that our results are not driven by a lack of statistical power when we look at the thresholds individually, in Panel (c) we plot the figure for our pooled sample, combining the thresholds into one cutoff, and re-scaling village population by subtracting the relevant threshold. The point estimate for the discontinuity is still -0.219, with a standard error of 0.183. In all of these cases, we fail to reject the null hypothesis of no discontinuity in the running variable.

Second, 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 baseline covariates either at the population thresholds (Imbens and Lemieux, 2008).

In Figure 3, we examine the geographic clustering of above- versus below-cutoff villages in the district of Ganjam, and find no evidence. In Figure 4, we examine a simple scatter plot of means of various village characteristics by different population bins around the threshold, to check for discontinuities of baseline covariates, and find no such evidence.

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. This is not surprising – even in our raw data, every single village above the cutoff gets a road, while 40% of those below the cutoff remain unconnected at the end of our sample. Of course this does not constitute a formal test of power for our instrument, which we proceed to do next.

We estimate the following RDD specification for our first stage, where we examine the effect of population cutoffs on actual PMGSY road construction:

$$\begin{aligned} Road_{i,v} &= \beta_0 + \beta_1 1[Population\_above\_threshold] \\ &+ \beta_3 1[500 - h \leq pop_{i,v} < 500 + h] + \beta_4 1[500 - h \leq pop_{i,v} < 500 + h] * (pop_{i,v} - 500) \\ &+ \beta_5 1[1000 - h \leq pop_{i,v} < 1000 + h] + \beta_6 1[1000 - h \leq pop_{i,v} < 1000 + h] * (pop_{i,v} - 1000) + \epsilon_{i,v} \end{aligned} \tag{1}$$

where $Road_{i,v}$ measures whether the village v – unconnected as of 2009 – received a PMGSY road by 2014, $pop_{i,v}$ is the baseline village population as recorded in 2001 census, c

is the population threshold, and η_js are the threshold fixed effects. We restrict our sample to villages with population within a certain bandwidth around the threshold, such that $pop_{i,v}$ $\in$ [c-h, c+h], where h is the value of the bandwidth around threshold c.

We look at all unconnected villages in 2009, which ensures that we are indeed capturing the effect of *newly constructed* roads. 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*. Therefore, to keep our design clean, we do not use any time series information in our formal tests; instead, we take a snapshot of our data at the last available year-end, and exploit the discontinuity based on where our borrowers live.[5].

In our baseline specifications, we cluster standard errors conservatively by allowing for correlations among the three partitions of the population support created by PMGSY program rules, i.e., 250-500, 500-1000, and above 1000, following Bertrand, Mehta and Mullainathan, 2004. That is, we not only allow for correlations within villagers living in the same village (which village-level clustering achieves), but are more stringent by also allowing all our outcomes to be correlated among villagers living in similar-sized villages, which might have been affected together by underlying PMGSY priority rules or any other unobservable at the population-group-level. In Table IA.3 in the internet appendix, we show that our results are not an artifact of clustering choice: they retain significance if we do not cluster at all and instead use heteroscedasticity and autocorrelation-robust standard errors.

Since our bank loan data contains very few small villages to start with (not surprisingly, private banks find it more fruitful to lend to larger habitations with existing roads), we choose 200 and 250 as our bandwidths for estimation purposes. Our results remain almost identical in terms of economic magnitudes if we restrict bandwidth to 150, but the number of villages falls to 9, and the resultant decline in statistical power makes our estimates imprecise.

Even then, we are left with a total of 11 villages in our bank loan sample within a bandwidth of 200 (12 within 250) that received a paved road after the bank launched its lending program in the district; so to gain statistical power, we do not estimate the thresholds of 500 and 1000 separately. Nevertheless, in Table IA.4 (panels E and F), we allow the impact

[5]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

of the two population thresholds under consideration being different. It turns out that our results come from both thresholds, but the impact of roads is higher for the population threshold of 500.

In addition, 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 in our regressions. In Table IA.4, we show that our results are robust to alternative functional forms that relax piece-wise linearity further: they are robust to the underlying relation between outcome variables being quadratic, and also to the slopes of the linear functions being different on either side of the cutoff.[6]

We estimate the effect of roads on our financing variables by running a reduced form specification, using population-based cutoffs as below:

$$\begin{aligned} Ln(1+Y_{i,v}) &= \gamma_0 + \gamma_1 1[Population\_above\_threshold] \\ &+ \gamma_2 1[500-h \leq pop_{i,v} < 500+h] + \gamma_3 1[500-h \leq pop_{i,v} < 500+h] * (pop_{i,v}-500) \\ &+ \gamma_4 1[1000-h \leq pop_{i,v} < 1000+h] + \gamma_5 1[1000-h \leq pop_{i,v} < 1000+h] * (pop_{i,v}-1000) + \epsilon_{i,v} \end{aligned} \tag{2}$$

where $Ln(1+Y_{i,v})$ are the borrower level outcomes of interest.

4 Results

In this section, we describe and discuss the main results and robustness checks. We first show that rural road construction indeed responded to PMGSY program population thresholds, and that they led to increased lending activity in affected villages. We then analyse loan performance in terms of default behavior, as well as maturity and interest rates. We also explore the channel through which this occurs, and find evidence that financing flows towards more productive opportunities in newly connected villages.

[6] Estimating non-linear specifications is problematic from the point of view of statistical power, so we only show these robustness results for a bandwidth of 250 in the Table. Lowering bandwidth to 200 in these specifications leads to economically similar but imprecise estimates.

4.1 Do population cutoffs predict road construction?

Table 2 formalizes the visual evidence in Figure 5 by presenting first stage estimates from 3.2 in both the census sample and the bank loan sample. For the census sample, the estimates imply there is a 21-28 percentage point increase in the probability of treatment around the cutoff. These estimates are very similar to estimates reported in Asher and Novosad (2017), and precise, even though we use data only from one district.

Next, we concentrate on our bank loan sample. First, in Figure 6 we plot heat maps of bank activity across the 209 different villages in Ganjam where our bank is active by 2014. In the next figure (Figure 7), we plot the locations of our 12 sample villages that were not connected originally when the bank started its lending activity in Ganjam and are within a population bandwidth of 250 around the thresholds of 500 and 1000. Of note is that within this sample, our villages do not cluster geographically, relative to where the bank is active.

We find that our estimates on the cutoff predicting connectivity are even higher in this sample: they imply that an average villager living in an initially unconnected village above the threshold has a 45-55% more chance of getting connected by the end of our sample than a villager living in a village below threshold. Overall, the results from this table confirm that there is a significant increase in the probability of treatment around the population threshold.

4.2 The extensive margin: Number of borrowers

Table 3 presents our population-based discontinuity estimates of the impact of new roads on the extensive margin for our bank loan sample, using Equation . 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. The dependent variable, *ExtMargin*, is an indicator variable that takes on the value one if an individual in the village received a loan from our bank.

We find that a new road is associated with a significant increase in number of villagers who receive financing. While our bank lends to 5.1% of village population in a typical village just below the thresholds, this number jumps to 7.9% in a village right above, using the more conservative coefficient estimate. This is a 54% jump in coverage right above the threshold, relative to right below. Note that when we expand the cutoff to 250 from 200, the economic magnitude of our results tend to go down a bit while statistical significance goes up: this is consistent with the effect of road building being slightly weaker in the larger village that gets added on as a result (something we also observe in table IA.4, where we see our effects being stronger for the 500 population threshold rather than 1000), but we gain more observations which improves statistical power.

Note that bank branches are set up where villagers from surrounding villages come to take loans, and that bank officials indicate to us that there is no official policy of going actively out to different villages and seek loan customers. Instead, these branches employ people from surrounding villages; these people spread the word on the bank's operations to their neighbors. In this setting, the jump in the number of customers we see around the cutoff seems more 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 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.

4.3 Loan quantities

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

The dependent variable in columns 1 and 3, *TotalDisb*, is the natural logarithm of one plus average total loan amount disbursed by the bank while in columns 2 and 4, *NetDisburse*, is the natural logarithm of one plus net loan amount disbursed. For each borrower, we compute the net loan amount disbursed as the loan amount disbursed (if a villager gets only one loan this is the amount of that loan, if he gets multiple, we add the loan values within person) minus any repayment made by the end of the calendar year on those loans. So the latter measure gives us a sense of outstanding loans to the average villager, above vs below cut-off, on the bank's balance sheet.

The granularity of information in the bank data allows us to precisely control for borrower-level characteristics that the bank cares about in its lending decisions. Looking at these characteristics, we find that villagers with higher incomes and more collateralizable assets – land and jewellery – are likely to get higher loan amounts from our bank. Younger and more educated villagers also seem to get higher loan amounts, but these results are not consistently statistically significant. Females get lower loan amounts than men. One possible explanation for this latter result could be that these agrarian societies are gender-biased 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 registered.

Interestingly, however, our coefficient on the population cutoff term suggests that – even after controlling for these differences – the expansion of lending activity extends to the intensive margin. Not only do more villagers living in villages above cutoffs get loans, they also get significantly *larger* loans. In terms of economic magnitudes, net amount lent to an average villager above the population cutoffs is 28-32% higher than to an average villager below the cutoff.

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

4.4 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 loan-type fixed effects in our regression.

Table 5 presents coefficient estimates from Equation 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 (300-700 for the 500 threshold and 800-1200 for the 1000 threshold) while Columns 4–6 presents estimates expanding the sample to include villages within 250 of the population thresholds.

When we examine loan structure, we generally find we find that loans made out to villagers likely to live in newly-connected villages are of very similar maturity. This is particularly true when the economic magnitude of the coefficient on maturity is put into perspective by benchmarking against the control group mean, that is, the average maturity in below-threshold villages.

In order to measure the loan performance, we create two measures: (1)Total loan amount that was overdue (*Overdue amount*), and (2)% Overdue amount captures this overdue amount as a fraction of loan amount disbursed (*% Overdue amount*). The evidence from the table generally suggests that individuals in villages above the threshold had better repayment behaviour than those in villages below, although our estimates are not precise. Overall, loan performance seems largely similar or (magnitude-wise) slightly better for loans made to villagers more likely to have received a new road. Note that here the economic magnitudes of our coefficient of interest is large relative to the control group means, but one reason behind the lack of significance is that the control group means themselves indicate a very low level of default. Default is very rare in our entire sample – both above and below the thresholds – so that we have no power to accurately estimate differences in it.

4.5 Loan interest rates

In this section we present discontinuity estimates for average interest rates (Table 6). The dependent variable is *AvgInterestRate*, the averaged interest rate across loans for each borrower. Again, to ensure that we compare interest rates only on similar loans, we add loan-type 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, both in terms of economic magnitude and statistical significance.

So overall, controlling for loan type, the loans made to newly-connected villagers looked very similar to those 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.

4.6 Channels: Evidence from loan uses

Next, 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 such as business expansion, asset acquisition, and working capital needs, versus other non-productive uses such as consumption needs, marriage and festival expenses.

Table 7 presents our estimates of the impact of new roads on loan use. Panel A presents 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 Panel B presents estimates expanding the sample to include villages within 250 of the population thresholds. In both panels, columns 1–2 presents results for Productive loans while columns 3–4 present results for Non-Productive loans. We partition the loan sample based on whether the financing was provided for productive uses (*Productive Loans*) such as business expansion, asset acquisition, and working capital needs or other uses such as consumption needs, marriage and festival expenses (*Non-productive Loans*).

We find that the higher in financing indeed accrued to productive purposes within newly connected villages. Interestingly, we find *lower* financing flows towards consumption loans in these same villages. Our results show that almost all of the approximately 30% higher credit to connected villages is through increase in loans for productive purposes. Economic magnitudes of coefficient on non-productive loans almost offsets the higher amount of productive loans; still the overall lending expansion is about 30% given that non-productive loans are rarely provided by the bank.

This suggests that our main results in Table 4 are not being driven purely by wealth effects. Had these been wealth effects, we should also have seen increases in consumption loans being made out to the newly-connected villages.

We further examine what kind of activities are financed by the bank when they finance productive uses in the connected villages. We classify productive loans based on whether they were disbursed for farming activities ("Crop Loans") and to set up or improve existing micro-businesses like village shops("Micro Loans"), or whether they were for other productive purposes like livestock loans and working capital loans.

Table 8 presents these estimates. Panel A presents 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 Panel B presents estimates expanding the sample to include villages within 250 of the population thresholds.

We find that the increase in financing on both margins indeed accrue almost exclusively to "Crop and Micro" loans within these villages. This is consistent with the type of activity one would expect to see with new road connectivity: a change in cropping pattern from low yield but storage time-insensitive cereals to high-paying but time-sensitive cash crops like vegetables, and the setting up of small village shops (like groceries selling vegetables).

In the Internet Appendix (IA), we examine the robustness of our results. While in the main paper we present results within the bank loan sample, where we can control for granular borrower characteristics, this can raise a selection concern regarding individuals in these villages who our bank could have lent to, but does not. In order to see whether this has any effect on our results, in Tables IA.1-IA.2, we redo our analysis by retaining all villagers in our analysis, with those whom the bank does not lend to coded as zeros. Our results remain largely similar. The magnitude of our estimated effects in these tables is consistent with the cumulative effects from the extensive margin (Table 3) and the intensive margin (Tables 4, 7 and 8). In results presented in Table IA.3, we show that our results remain very similar if we estimate them as instrumental variable regressions, using the population thresholds as instruments, instead of in reduced form.

4.7 Falsification test

In this subsection we explore the possibility that factors other than the road treatment may be spuriously driving our results.

In our placebo exercise, we run our baseline specification for the set of villages which are *already connected* to a road network in 2001. For this sample, therefore, there is no discontinuous increase in road treatment at the population threshold, although our estimation methodology remains identical. Estimates from this exercise are reported in Table 9.

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 not due to other discontinuous differences in villages around the cutoffs whose effect we spuriously attribute to new roads.

4.8 Timing of road construction

While exact timing of road construction is endogenous, it is interesting to examine whether and how exposure to more years of new rural roads affects the composition of financing within these villages. For each village in our sample, we create an exposure measure, a continuous variable which captures the total number of years since the road was completed.[7] In this test, we are simply interested in finding out whether the effects of road building on financing show up immediately, or are more gradual. To do so, we deviate from our population-discontinuity based-design in the rest of the paper, and simply examine timing since construction of a new road for all initially unconnected villages (as of 2009, constrained to a bandwidth of 250) in Figure 8.

The figure plots the difference between lending amounts to newly connected villages minus that to villages that are not connected to the network at all over time. The horizontal axis plots time relative to road completion while the vertical axis presents differences in net disbursement between treated and control villages by the end of each year. We find that most of our financing effects from new rural roads around the population threshold take about one year to materialize: there is a jump in lending at that point, which gradually goes back to similar levels afterwards.

This is consistent with the view that a lot of the lending activity around new road construction is to *facilitate villagers to transition* into higher-productivity uses of their time: for example, changing their crop pattern. They need the loan to buy seeds and fertilizers to change over form cereal to cash crop cultivation, for example. But once they have done so, the higher profits from their changed pursuits allow them to pay off their loans within a relatively short span of time.

5 Distributional consequences of connectivity: Evidence from loan financing

We have thus far established the causal impact of rural roads on lending flows. New rural roads leads to greater access to finance within these villages and most of the new loans are disbursed for productive purposes. In this section, we examine the heterogeneity of the treatment effects based on baseline borrower characteristics. If we assume that our aggregate evidence that financing indeed flows to more productive uses also applies to population sub-samples, then 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 difference in default behavior or interest

[7] For our sample of villages, the earliest possible year to receive a road is 2009 while the latest year is 2014. By construction, the lower limit of this variable is censored at 0 and upper at 5.

rates of note across different types of borrowers.

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 through 3 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) while columns 4 through 6 presents estimates expanding the sample to include villages within 250 of the population thresholds.

Results from Table 10 suggest that new roads seem to alleviate collateral constraints among borrowers. Low asset households – those without significant cultivable land ownership – typically get smaller loans on average, which is perhaps not surprising. But these households benefit disproportionately more from connectivity. 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 existing land – used as collateral – and therefore borrowing capacity, our effects would have been stronger among those with higher assets.

We also observe that loan disbursements are higher in newly connected villages for lower income villagers, but those with some education (the borrower attended a village school). These results on the educated are consistent with Mukherjee (2011) and Adukia et al. (2017), who respectively show that PMGSY increases school enrollment, and that children stay in school longer and perform better on standardized exams 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.

Finally, we also examine gender differences, and find that women are less likely to get loans across the board, but particularly so in the newly connected villages. Again, this could reflect either a gender bias on the loan supply side or on the loan demand side, as outlined previously. Women who do get loans in newly connected villages do seem somewhat less likely to default, but given the very low default rates in the overall sample, this difference is not of an economically significant magnitude in percentage point terms.

Overall, we find that new roads seem to disproportionately benefit villagers who have received some basic education but do not have collateralizable assets.

6 Conclusion

Increasing infrastructure investments are a key component of growth strategy in many countries, and a particular focus of policy now, given Chinas massive "One Belt One Road" investment project. 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 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, and seems to flow disproportionately to villagers with basic education who lack collateralizable assets. Our results have important implications for understanding trickle down benefits of infrastructure-building and its distributional consequences.

7 References

Adukia, A., Asher, S. and Novosad, P., 2017, Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction, Working Paper.

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, 2015, Do Rural Roads Create Pathways out of Poverty? Evidence from India, Working Paper.

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.

Aterido, R., Thorsten Beck, and Leonardo Iacovone, 2013, Gender and Finance in Sub-Saharan 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 Mller (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., A. Demirg-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.

Caldern, Csar, and Luis Servn, 2004, The effects of infrastructure development on growth and income distribution, No. 270., World Bank Publications.

Carter, Laurence W., Florizone, and Richard Emile Joseph, 2013, A winning framework for public-private partnerships : lessons from 60-plus IFC projects, IFC smart lessons brief, World Bank, Washington DC.

Das, Abhiman, Ejaz Ghani, Arti Grover, William R. Kerr, and Ramana Nanda, 2017, The GQ Infrastructure Project and Financial Development, World Bank Interim Report.

Demirguc-Kunt, A., and Ross Levine, 2008a, Finance, financial sector policies, and long-run 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., 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), 615635.

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.

Khandker, Shahidur R., and Gayatri B. Koolwal, 2010, How infrastructure and financial institutions affect rural income and poverty: evidence from Bangladesh, The Journal of Development Studies 46(6), 1109-1137.

King, R.G. and Levine, R., 1993, Finance and growth: Schumpeter might be right, The quarterly journal of economics, 108(3), 717-737.

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

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.

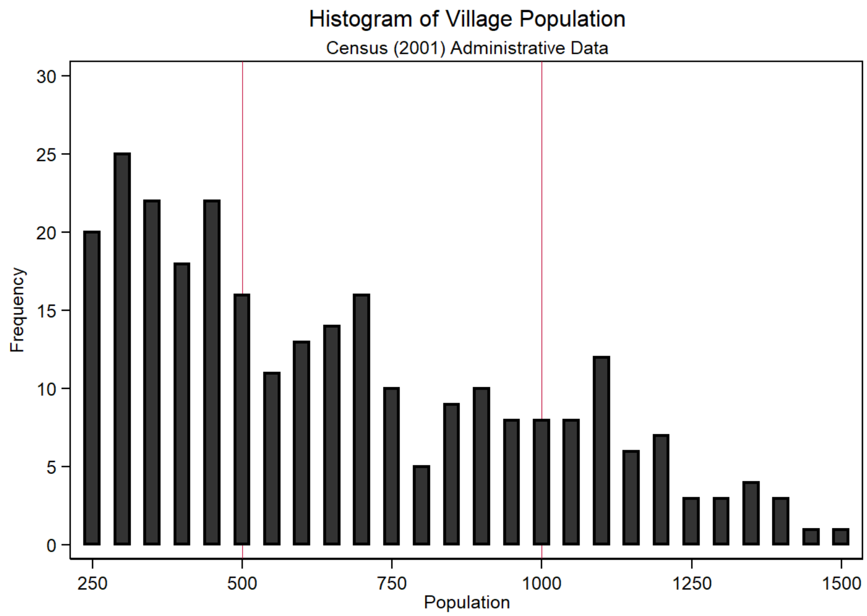
Muravyev, A., Dorothea Schfer and Oleksandr Talavera, 2007, Entrepreneurs' gender and financial constraints: Evidence from international data, Discussion Papers of DIW Berlin 706, German Institute for Economic Research.

Shamdasani, Y., 2016, 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.

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.

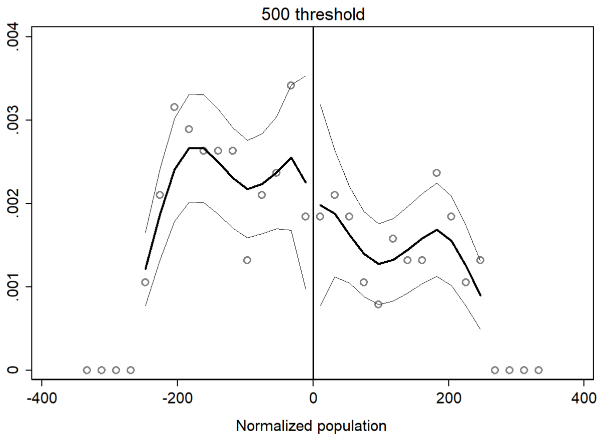
Figure 1: Distribution of running variable



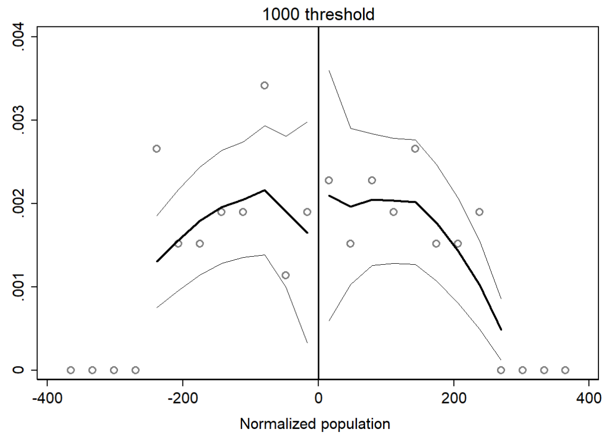
(a)

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. that did not have paved roads at the start of our sample.

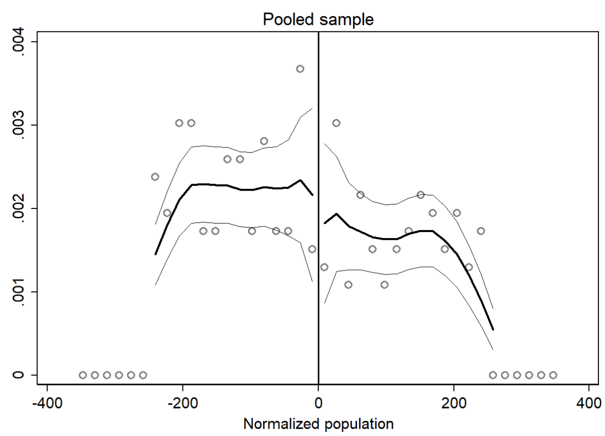
Figure 2: McCrary Test for discontinuity in the running variable



(a)



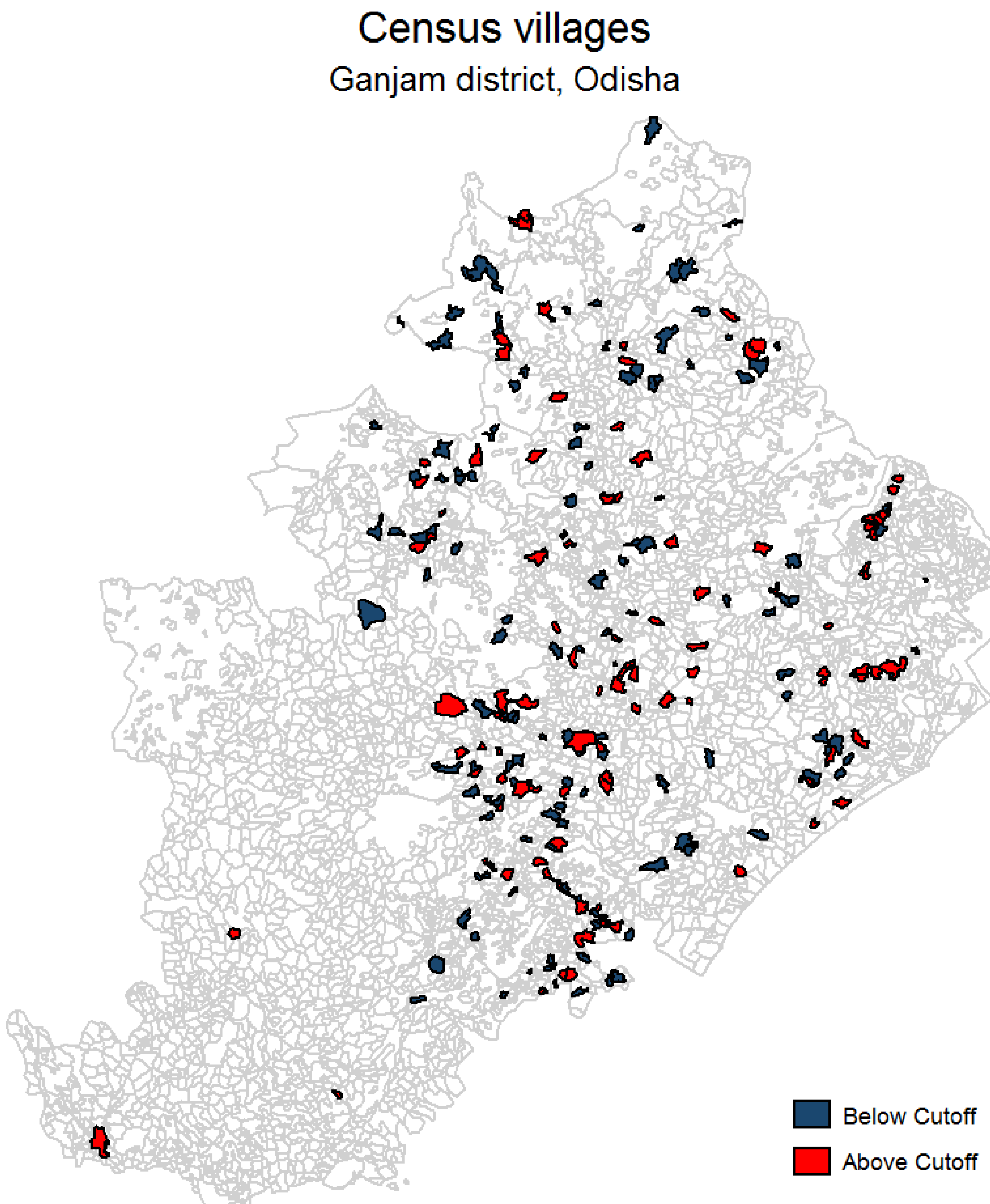
(b)



(c)

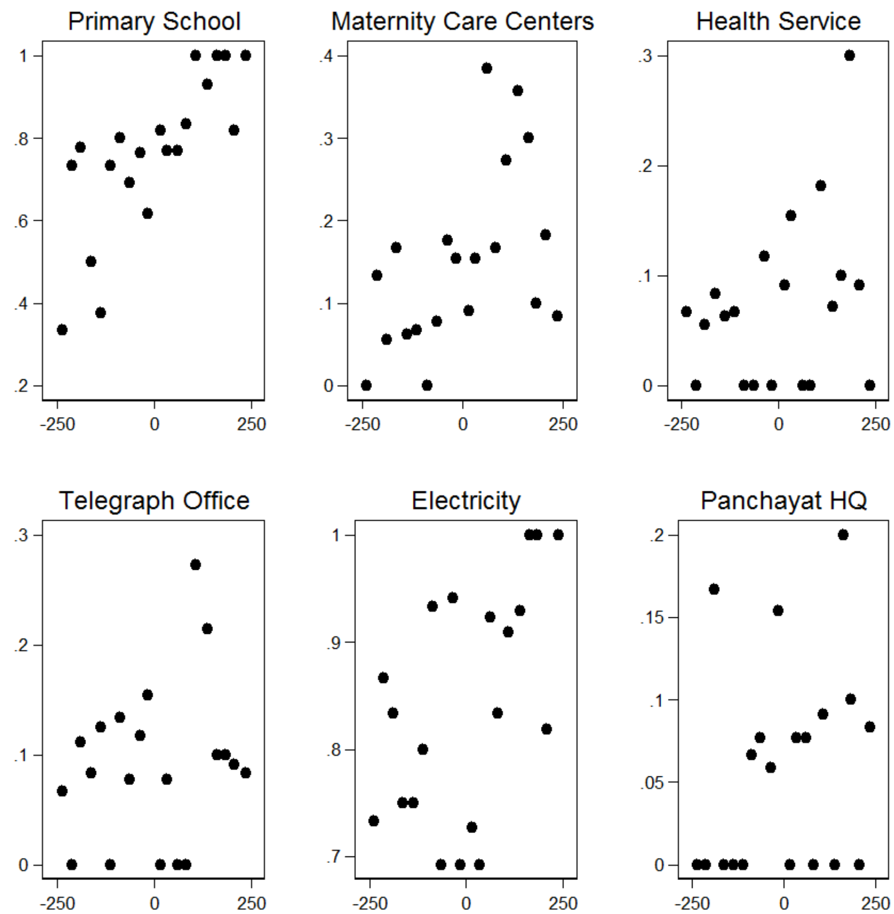
Notes: These figures plot non-parametric regressions of the distribution following McCrary (2008), testing for a discontinuity at zero. The village population is normalized by subtracting the population threshold, either 500 or 1000. Our sample consists of villages that did not have paved roads at the start of our sample. Panel (a) plots the figure for villages with populations within 250 of the population threshold i.e. 750-1250 for the 1000 threshold while Panel (b) plots the figure for villages with populations within 250 of the population threshold i.e. 250-749 for the 500 threshold. In panel (a), the point estimate for the discontinuity is 0.469, with a standard error of 0.782 while in panel (b) the estimate is -0.011 with a standard error of 0.598. Panel (c) plots the figure for our pooled sample. The point estimate for the discontinuity is -0.219, with a standard error of 0.183.

Figure 3: Distribution of villages based on cutoff in Ganjam District, Odisha



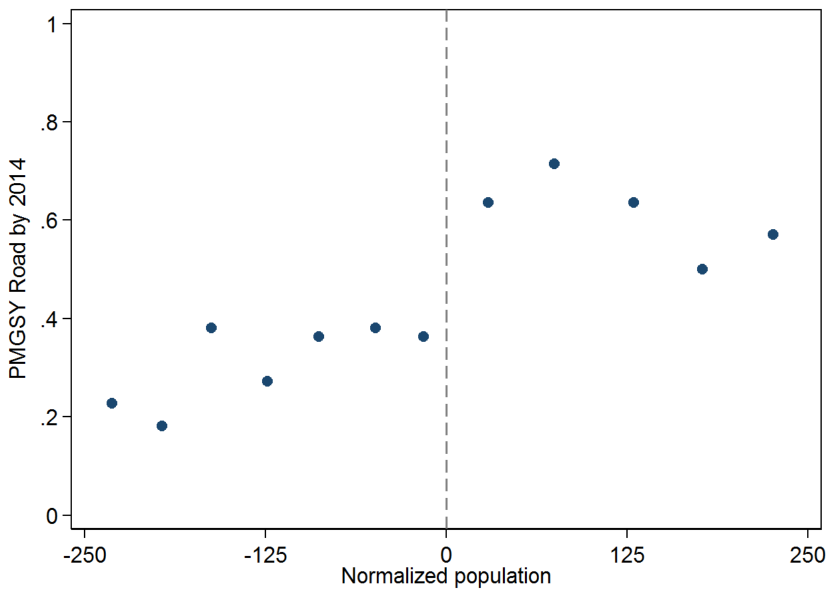
Notes: The figure shows the distribution of unconnected villages in Ganjam district, Odisha. The sample consists of villages that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. Blue shaded regions represent villages right below the population cutoff while red shaded regions display villages right above the population cutoff.

Figure 4: Balance of baseline village characteristics



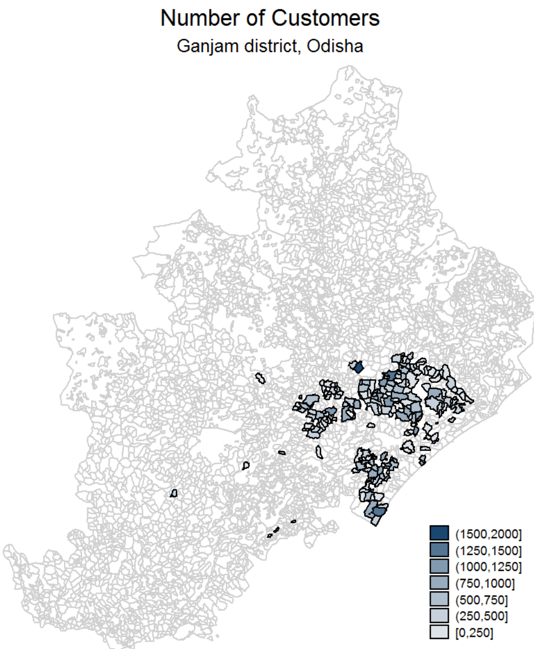
Notes: The figure plots means of baseline village characteristics over normalized popuation. Points to the right of zero are above treatment thresholds, while points to the left of zero are below treatment thresholds. The bin width is 25 on either side of the threshold and each point represents approximately fifteen observations. 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

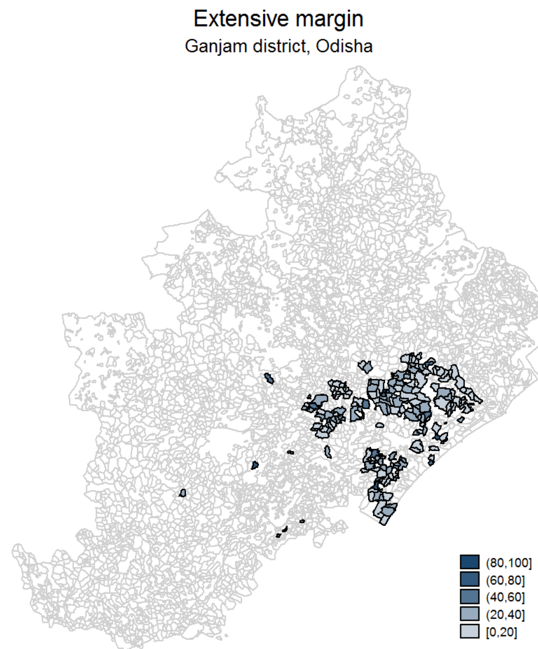


Notes: The figure plots the probability of receiving a road under a PMGSY by 2014 over village population as recorded in the 2001 Population Census. The sample consists of villages that did not have paved roads at the start of our sample. We normalize the baseline population by subtracting the cutoff.

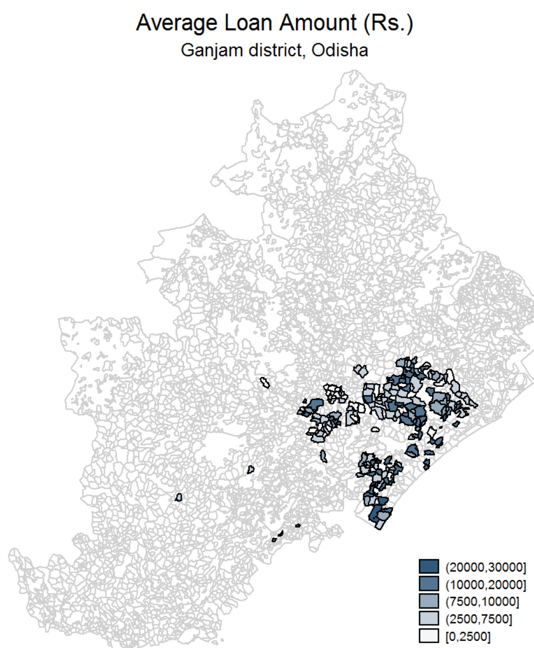
Figure 6: Distribution of bank lending sample



(a)



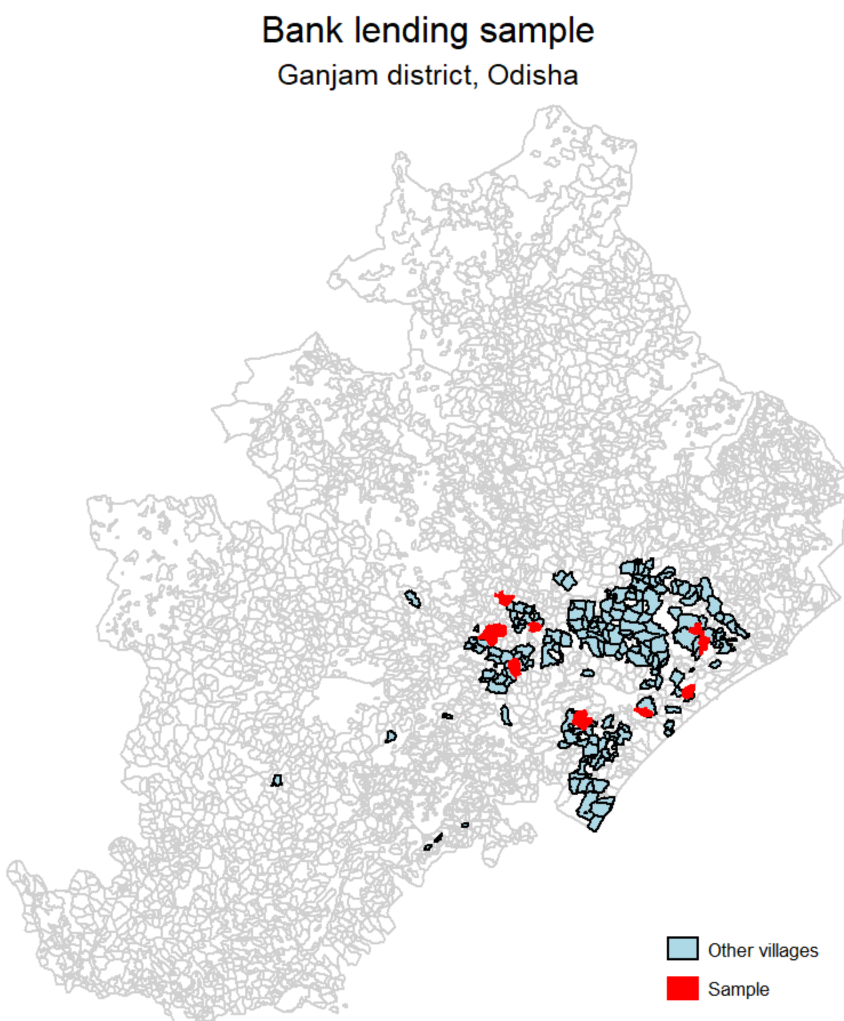
(b)



(c)

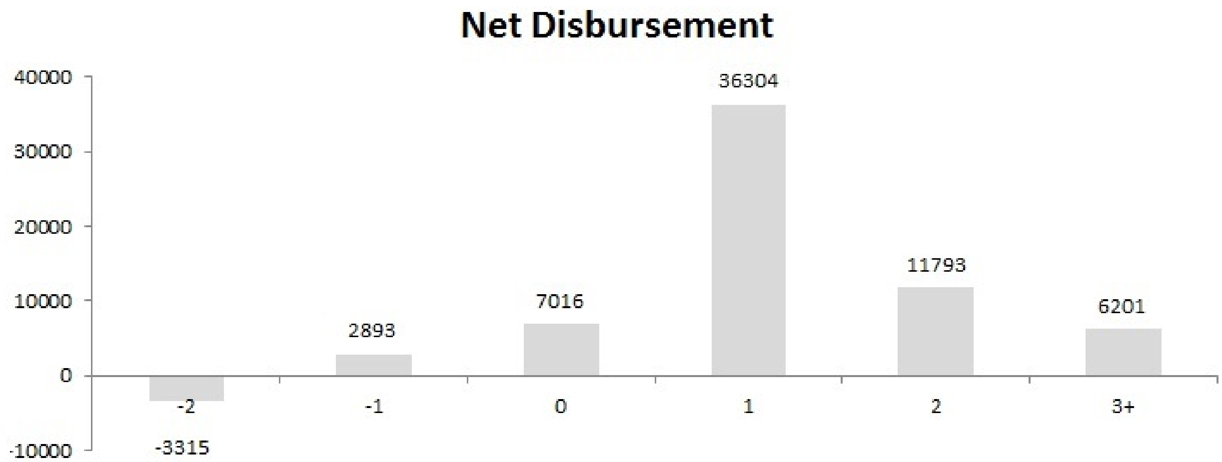
Notes: The figure shows the geographic dispersion of the lending activity of the bank sample. Panel(a) presents distribution of customers across the bank sample. Panel(b) presents the fraction of individuals within the village who receive a loan from the bank. Panel(c) presents the average loan amounts across different villages in our sample. In each panel, darker shades represent greater intensity.

Figure 7: Distribution of bank lending sample



Notes: The figure displays the geographic distribution of bank lending sample we use relative to banking activites in other villages in our sample.

Figure 8: Timing of road construction and lending activity



(a)

Notes: The figure shows the lending activity of the bank relative to the year of road construction. The horizontal axis plots time relative to road completion while the vertical axis present net disbursement by end of each calendar year in rupees. *Net Disbursement* is the average net loan amount disbursed. For each borrower, we compute the net loan amount disbursed as the average of loan amount disbursed minus any repayment made by the end of each calendar year 2014.

Table 1: Summary Statistics : Bank Loan Sample

The table presents means and standard deviations (in parentheses) for our main 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 through 3 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 4 through 6 presents means expanding the sample to include villages within 250 of the population thresholds. Columns 1 and 4 presents mean (standard deviation in parentheses) for Non-JLG (non-microfinance) loans while columns 2 and 5 present statistics for Productive loans. Columns 3 and 6 present summary statistics for Crop & MEL Loans. We partition the loan sample based on whether the financing was provided for productive uses (*Productive Loans*) such as business expansion, asset acquisition, and working capital needs and based on whether they were disbursed for farming activities specifically for the successful harvest of the Paddy Crop ("Crop Loans") or whether they were disbursed to improve the existing business income by increasing stock or better management of existing working capital needs ("Micro Loans"). *Extensive Margin* is an indicator variable that takes on the value one if an individual in the village received a loan from the bank. *Loan amount* is the total loan amount disbursed by the bank while *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. In order to measure the loan performance, we create *% Overdue amount* which captures the fraction of loan amount disbursed that was overdue by 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 *High School Education* which takes the value of one if the borrower is a high school graduate, 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 who had a loan with the bank by the end of the calendar year 2014.

Bandwidth	±200			±250		
	(1) All Loans	(2) Productive Loans	(3) Crop/MEL Loans	(4) All Loans	(5) Productive Loans	(6) Crop/MEL Loans
			Panel A: Loan Characteristics			
Extensive margin (%)	4.8	4.5	1.1	4.4	4.1	1.0
	(0.21)	(0.20)	(0.10)	(0.20)	(0.19)	(0.09)
Loan amount (Rs.)	35,361	36,426	39,825	35,092	36,246	39,778
	(14,043)	(10,928)	(10,667)	(14,121)	(11,041)	(10,703)
Net Disbursement (Rs.)	21,107	22,244	34,649	20,831	21,983	34,451
	(12,565)	(12,609)	(10,539)	(12,710)	(12,774)	(10,818)
Loan maturity (Years)	1.90	1.98	1.98	1.90	1.98	1.98
	(0.27)	(0.10)	(0.16)	(0.27)	(0.10)	(0.16)
Interest Rate (%)	15.0	15.6	15.6	15.0	15.6	15.6
	(3.0)	(0.9)	(0.0)	(3.0)	(0.9)	(0.0)
Overdue Amount (%)	0.39	0.37	0.00	0.38	0.36	0.00
	(6.13)	(5.55)	(0.0)	(6.01)	(5.44)	(0.0)
			Panel B: Borrower Characteristics			
Age (Years)	37	37	36	37	37	36
	(9)	(9)	(9)	(9)	(9)	(9)
Female (%)	34	33	28	34	33	28
	(47)	(47)	(45)	(47)	(47)	(45)
Schooling (%)	78	78	92	77	77	90
	(41)	(42)	(28)	(41)	(42)	(30)
Annual Income (Rs.)	111790	113821	128908	112104	114681	130739
	(77914)	(78699)	(85142)	(79510)	(80132)	(85180)

Table 2: 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 probabiltiy of receiving a road under PMGSY by 2014. The dependent variable is a indicator variable that takes on the value one if a village has received a PMGSY road before 2014. From this sample, we drop all the villages which received a road before 2009. Columns 1 and 2 presents results for the sample of villages in 2001 Population Census while columns 3 and 4 presents results for the bank loan sample. The first and the third column presents results for villages with populations within 200 of the population threshold (300-700 for the 500 threshold and 800-1200 for the 1000 threshold). The second and fourth columns expand the sample to include villages within 250 of the population thresholds. Both the samples consists of villages that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. The bank loan sample includes all the individuals in a particular village. All regressions include threshold fixed effects. Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

	Census sample		Bank Loan sample	
	(1)	(2)	(3)	(4)
Bandwidth	±200	±250	±200	±250
Above Cutoff	0.283***	0.229***	0.552***	0.454***
	(0.048)	(0.071)	(0.120)	(0.112)
F-Statistic	34.44	10.38	21.19	16.37
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.22	0.25	0.20	0.22
Observations	230	260	8141	9260

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

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

The table presents reduced form estimates of the effect of new rural roads on lending activity within the villages. 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 in columns 1 and 2 is is an indicator variable that takes on the value one if an individual in the village received a loan from the bank. Our sample consists of all individuals within villages that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All regressions include threshold fixed effects. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

Bandwidth	± 200	± 250
	(1)	(2)
Above Cutoff	0.044*	0.028***
	(0.025)	(0.007)
Control group mean	0.051	0.051
Threshold FE	Yes	Yes
R^2	0.008	0.009
Observations	8141	9260

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. Columns 1 and 2 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 columns 3 and 4 presents reduced form estimates expanding the sample to include villages within 250 of the population thresholds. The dependent variable in columns 1 and 3, *TotalDisb*, is the natural logarithm of one plus total loan amount disbursed by the bank while in columns 2 and 4, *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. Our sample includes villages that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All regressions include threshold fixed effects. The specification also includes baseline borrower-level controls for age, income, collateral, level of education and gender. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

Bandwidth	±200		±250	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
Above Cutoff	1.137***	1.609***	0.864**	1.386***
	(0.220)	(0.166)	(0.402)	(0.314)
Age(years)	-0.023	-0.034**	-0.025	-0.036**
	(0.018)	(0.012)	(0.020)	(0.015)
Income	0.005***	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Land & Jewel	1.613*	1.315*	1.604*	1.309*
	(0.882)	(0.764)	(0.863)	(0.749)
School Education	0.634*	0.561	0.562**	0.493
	(0.374)	(0.389)	(0.285)	(0.300)
Female(1=Yes)	-3.049***	-2.781***	-2.947***	-2.689***
	(0.198)	(0.241)	(0.122)	(0.140)
Control group mean	5.46	4.96	5.46	4.96
Controls	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.20	0.20	0.19	0.19
Observations	759	759	792	792

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 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 columns 4 through 6 presents reduced form estimates expanding the sample to include villages within 250 of the population thresholds. 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 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. The sample only consists of villages that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All regressions include threshold and loan purpose fixed effects. The specification also includes baseline borrower-level controls for age, income, collateral, level of education and gender. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported.Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

Bandwidth	±200			±250		
	(1) Ln(Maturity)	(2) ODAmount	(3) %OD Amount	(4) Ln(Maturity)	(5) ODAmount	(6) %OD Amount
Above Cutoff	-0.003	5.936	-0.057	-0.008***	5.101	-0.090
	(0.007)	(47.306)	(0.067)	(0.001)	(48.858)	(0.118)
Control group mean	0.88	32.3	0.10	0.88	32.3	0.10
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes	Yes	Yes
Loan purpose FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.92	0.10	0.039	0.92	0.089	0.038
Observations	759	759	759	792	792	792

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. The sample only consists of villages that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All regressions include threshold and loan purpose fixed effects. The specification also includes baseline borrower-level controls for age, income, collateral, level of education and gender. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

Bandwidth	± 200	± 250
	(1)	(2)
Above Cutoff	0.001	-0.002
	(0.000)	(0.004)
Control group mean	0.080	0.080
Controls	Yes	Yes
Threshold FE	Yes	Yes
Loan purpose FE	Yes	Yes
R^2	0.96	0.96
Observations	759	792

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

Table 7: 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. Panel A 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 Panel B presents reduced form estimates expanding the sample to include villages within 250 of the population thresholds. In both panels, columns 1 and 2 presents results for Productive loans while columns 3 and 4 presents results for Non-Productive loans. We partition the loan sample based on whether the financing was provided for productive uses (*Productive Loans*) such as business expansion, asset acquisition, and working capital needs or other uses such as consumption needs, marriage and festival expenses (*Non-productive Loans*). The dependent variable in columns 1 and 3, *TotalDisb*, is the natural logarithm of one plus total productive(non-productive)loan amount disbursed by the bank while in columns 2 and 4, *NetDisburse*, is the natural logarithm of one plus total net productive(non-productive) loan amount disbursed. For each borrower, we compute the net productive(non-productive) loan amount disbursed as the average of productive(non-productive) loan amount disbursed minus any repayment made by the end of the calendar year 2014. Our sample includes villages that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All regressions include threshold fixed effects. The specification also includes baseline borrower-level controls for age, income, collateral, level of education and gender. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

	Panel A: Bandwidth ±200			
	Productive Loans		Non-Productive Loans	
	(1)	(2)	(3)	(4)
	TotalDisb	NetDisburse	TotalDisb	NetDisburse
Above Cutoff	1.203***	1.692***	-0.999	-1.478***
	(0.087)	(0.029)	(0.638)	(0.344)
Control group mean	5.22	4.71	6.57	5.81
Controls	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.21	0.21	0.19	0.16
Observations	759	759	759	759

	Panel B:Bandwidth ±250			
	Productive Loans		Non-Productive Loans	
	(1)	(2)	(3)	(4)
	TotalDisb	NetDisburse	TotalDisb	NetDisburse
Above Cutoff	0.990***	1.533***	-0.171	-0.891
	(0.281)	(0.189)	(1.469)	(0.927)
Control group mean	5.22	4.71	6.57	5.81
Controls	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.20	0.20	0.18	0.16
Observations	792	792	792	792

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

Table 8: impact of new roads on productive loan types

This table presents reduced form estimates from Equation 2 quantifying the effect of new rural roads on type of productive loans disbursed. Panel A 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 Panel B presents reduced form estimates expanding the sample to include villages within 250 of the population thresholds. We classify productive loans based on whether they were disbursed for farming activities specifically for the successful harvest of the Paddy Crop (”Crop Loans”) or whether they were disbursed to improve the existing business income by increasing stock or better management of existing working capital needs (”Micro Loans”). Other prductive loans include livestock loans, and retailer loan. In both panels, columns 1 and 2 presents results for Crop & Micro loans while columns 3 and 4 presents results for Other productive loans. The dependent variable in columns 1 and 3 is the natural logarithm of one plus total Crop and Micro (Other Productive)loan amount disbursed by the bank while in columns 2 and 4 it is the natural logarithm of one plus total net Crop and Micro (Other Productive) loan amount disbursed. For each borrower, we compute the net Crop and Micro (Other Productive) 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. Our sample consists of villages that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All regressions include threshold fixed effects. The specification also includes baseline borrower-level controls for age, income, collateral, level of education and gender. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

	Panel A: Bandwidth ±200			
	Crop & Micro Loans		Other Productive Loans	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
Above Cutoff	1.330**	1.321**	-0.057	0.440
	(0.669)	(0.668)	(0.770)	(0.728)
Control group mean	0.99	0.97	4.23	3.73
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
R^2	0.065	0.065	0.14	0.14
Observations	759	759	759	759

	Panel B:Bandwidth ±250			
	Crop & Micro Loans		Other Productive Loans	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
Above Cutoff	1.567***	1.565***	-0.491	0.053
	(0.097)	(0.088)	(0.193)	(0.117)
Control group mean	0.99	0.97	4.23	3.73
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
R^2	0.059	0.060	0.14	0.13
Observations	792	792	792	792

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

Table 9: Placebo on Connected Villages

The table presents reduced form estimates from Equation 2 of the effect of new rural roads on loan disbursed on a placebo sample. Our sample consists of villages that were already connected at baseline and hence the PMGSY thresholds were not applicable to them. Panel A 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 Panel B presents reduced form estimates expanding the sample to include villages within 250 of the population thresholds. We classify productive loans based on whether they were disbursed for farming activities specifically for the successful harvest of the Paddy Crop (”Crop Loans”) or whether they were disbursed to improve the existing business income by increasing stock or better management of existing working capital needs (”Micro Loans”). Other prductive loans include livestock loans, and retailer loan. In both panels, columns 1 through 3 presents results for Non-JLG loans while columns 4 through 6 present results for Crop & Micro loans loans. The dependent variable in column 1 and 4 is an indicator variable that takes on the value one if an individual in the village received a loan from our bank. The dependent variable in columns 2 and 5 is the natural logarithm of one plus average total Non-JLG (Crop & Micro) loan amount disbursed by the bank while in columns 3 and 6 it is the natural logarithm of one plus average net Non-JLG (Crop & Micro) loan amount disbursed. For each borrower, we compute the average net Non-JLG (Crop & Micro) loan amount disbursed as the average of Non-JLG (Crop & Micro) loan amount disbursed minus any repayment made by the end of the calendar year 2014. Our sample consists of individuals who had a Non-JLG (Crop & Micro) loan with the bank by the end of the calendar year 2014. All regressions include threshold fixed effects. Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

	Panel A: Bandwidth ±200					
	All Loans			Crop & Micro Loans		
	(1) ExtMargin	(2) TotalDisb	(3) NetDisburse	(4) ExtMargin	(5) TotalDisb	(6) NetDisburse
Above Cutoff	-0.018	-0.189	-0.171	-0.005	-0.038	-0.038
	(0.027)	(0.282)	(0.302)	(0.011)	(0.098)	(0.097)
Control group mean	0.053	0.55	0.49	0.018	0.19	0.19
Controls	No	No	No	No	No	No
Threshold FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.007	0.007	0.011	0.001	0.001	0.001
Observations	35858	35858	35858	35858	35858	35858

	Panel B:Bandwidth ±250					
	All Loans			Crop & Micro Loans		
	(1) ExtMargin	(2) TotalDisb	(3) NetDisburse	(4) ExtMargin	(5) TotalDisb	(6) NetDisburse
Above Cutoff	-0.021	-0.220	-0.181	0.000	0.006	0.006
	(0.027)	(0.278)	(0.275)	(0.014)	(0.144)	(0.142)
Control group mean	0.058	0.60	0.55	0.022	0.23	0.23
Controls	No	No	No	No	No	No
Threshold FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.001	0.001	0.002	0.000	0.000	0.000
Observations	40939	40939	40939	40939	40939	40939

Standard errors in parentheses

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

Table 10: Impact of new roads, by borrower characteristics

The table presents reduced form estimates from Equation 2 of the heterogenous effects of new rural roads by borrower characteristics. Columns 1 through 3 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 columns 4 through 6 presents reduced form estimates expanding the sample to include villages within 250 of the population thresholds. The dependent variable in column 1 and 4 is the natural logarithm of one plus total loan amount disbursed by the bank. The dependent variable in columns 2 and 5 is the natural logarithm of one plus total net loan amount disbursed 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 Age, logarithm of one plus assets, logarithm of one plus income, indicator for School Education and indicator for whether the gender of the borrower is female. Our sample consists of individuals who had a loan with the bank by the end of the calendar year 2014. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. All regressions include threshold fixed effects. Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

Bandwidth	± 200			± 250		
	(1) TotalDisb	(2) %ODAmount	(3) AvgIntRate	(4) TotalDisb	(5) %ODAmount	(6) AvgIntRate
Above Cutoff	4.353*	0.005	0.002	4.680*	-0.064	0.007
	(1.259)	(1.170)	(0.009)	(1.263)	(1.273)	(0.009)
Age(Years)	-0.007*	-0.004	-0.000*	-0.007*	-0.004	-0.000
	(0.002)	(0.006)	(0.000)	(0.002)	(0.006)	(0.000)
Ln(1+Assets)	2.187***	-0.142	-0.001	2.187***	-0.140	-0.001
	(0.172)	(0.171)	(0.002)	(0.162)	(0.170)	(0.002)
Ln(1+Income)	6.896***	0.253	0.005	6.895***	0.263	0.006
	(0.446)	(0.214)	(0.008)	(0.443)	(0.234)	(0.008)
Female(1=Yes)	-0.894**	0.185	0.003	-0.895**	0.185	0.003
	(0.194)	(0.320)	(0.002)	(0.199)	(0.318)	(0.002)
School Education	-0.146	-0.525	-0.002	-0.146	-0.524	-0.003
	(0.210)	(0.732)	(0.003)	(0.208)	(0.732)	(0.003)
Above Cutoff x Ln(1+Assets)	-1.176**	0.138	0.001	-1.149**	0.129	0.001
	(0.249)	(0.077)	(0.002)	(0.209)	(0.071)	(0.002)
Above Cutoff x Ln(1+Income)	-3.010**	-0.274	-0.002	-3.495***	-0.257	-0.012
	(0.345)	(0.130)	(0.007)	(0.350)	(0.122)	(0.010)
Above Cutoff x Age (Years)	-0.040	-0.008	-0.000	-0.045	-0.007	-0.000
	(0.026)	(0.019)	(0.000)	(0.029)	(0.019)	(0.000)
Above Cutoff x School Education	1.495**	0.443	0.003	1.405**	0.429	0.003
	(0.304)	(0.779)	(0.005)	(0.217)	(0.765)	(0.006)
Above Cutoff x Female	-1.067**	-0.183*	-0.006	-0.791**	-0.221	-0.010*
	(0.236)	(0.057)	(0.002)	(0.131)	(0.092)	(0.003)
Control group mean	5.46	0.10	0.08	5.46	0.10	0.08
Threshold FE	Yes	Yes	Yes	Yes	Yes	Yes
Loan purpose FE	No	Yes	Yes	No	Yes	Yes
R^2	0.29	0.11	0.96	0.28	0.11	0.96
Observations	759	759	759	792	792	792

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

Internet Appendix: Additional tables

Table IA1: 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. Columns 1 and 2 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 columns 3 and 4 presents reduced form estimates expanding the sample to include villages within 250 of the population thresholds. The dependent variable in columns 1 and 3, *TotalDisb*, is the natural logarithm of one plus total loan amount disbursed by the bank while in columns 2 and 4, *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. Our sample includes villages that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. We add back individuals who did not receive loans from the bank by the end of calendar year 2014. All regressions include threshold fixed effects. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

Bandwidth	±200		±250	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
Above Cutoff	0.461*	0.480*	0.295**	0.327**
	(0.261)	(0.248)	(0.068)	(0.057)
Control group mean	0.53	0.48	0.53	0.48
Controls	No	No	No	No
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.0075	0.0074	0.0091	0.0090
Observations	8141	8141	9260	9260

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

Table IA2: Impact of new roads on lending activity - loan types

The table presents reduced form estimates from Equation 2 of the effect of new rural roads on lending activity within these villages. Panels A, and C present 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 Panel B and D present estimates expanding the sample to include villages within 250 of the population thresholds. In all panels, columns 1 and 2 presents results for Productive loans while columns 3 and 4 present results for our Non-Productive loans. The dependent variable in columns 1 and 3 is the natural logarithm of one plus total loan amount disbursed by the bank while in columns 2 and 4 it is the natural logarithm of one plus 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 add back all the individuals in a particular village who do not take out a loan from the bank. All regressions include threshold fixed effects. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. All regressions include threshold fixed effects. Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

	Panel A: Lending Quantities by loan type, Bandwidth ±200			
	Productive Loans		Non-Productive Loans	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
Above Cutoff	0.434*	0.454**	0.044	-0.027
	(0.242)	(0.229)	(0.442)	(0.318)
Control group mean	0.51	0.46	0.64	0.56
Controls	No	No	No	No
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.0087	0.0087	0.0035	0.0027
Observations	8141	8141	8141	8141

	Panel B: Lending Quantities by loan type, Bandwidth ±250			
	Productive Loans		Non-Productive Loans	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
Above Cutoff	0.280***	0.312***	0.207	0.109
	(0.048)	(0.037)	(0.322)	(0.234)
Control group mean	0.51	0.88	0.64	0.56
Controls	No	No	No	No
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.010	0.012	0.0067	0.0056
Observations	9260	9260	9260	9260

Continued

	Panel C: Lending Quantities by productive loan type, Bandwidth ±200			
	Crop & Micro Loans		Other Productive Loans	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
Above Cutoff	0.210***	0.208***	0.230	0.252
	(0.002)	(0.002)	(0.246)	(0.233)
Control group mean	0.096	0.094	0.41	0.36
Controls	No	No	No	No
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.0021	0.0022	0.0091	0.0090
Observations	8141	8141	8141	8141

	Panel D: Lending Quantities by productive loan type, Bandwidth ±250			
	Crop & Micro Loans		Other Productive Loans	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
Above Cutoff	0.206***	0.205***	0.081*	0.114***
	(0.008)	(0.007)	(0.042)	(0.032)
Control group mean	0.096	0.094	0.41	0.36
Controls	No	No	No	No
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.0026	0.0026	0.010	0.0100
Observations	9260	9260	9260	9260

Standard errors in parentheses

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

Table IA3: Impact of new roads on lending activity, instrumental variables

The table presents regression discontinuity estimates from instrumental variables (IV) regression of the effect of new rural roads on lending activity within these villages. Panels A present IV estimates of baseline Table 3 while panel B present IV estimates for Table 4. Panel C present IV estimates of Table 5 while panels D and E present IV estimates of Table 7. Lastly, panels F and G present IV estimates of Table 8. The dependent variable for panel A is an indicator which takes the value of one if the individual received a loan from the bank and zero otherwise. The dependent variable for panels B through G, in columns 1 and 3 is the natural logarithm of one plus total loan amount disbursed by the bank while in columns 2 and 4 it is the natural logarithm of one plus net loan amount disbursed. 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. Our bank loan sample consists of individuals who had a loan with the bank by the end of the calendar year 2014. We add back all the individuals in a particular village who do not take out a loan from the bank. We include villages that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All regressions include threshold fixed effects. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

	Panel A: Extensive margin	
Bandwidth	± 200	± 250
	(1)	(2)
PMGSY Road	0.079*	0.061***
	(0.048)	(0.017)
Control group mean	0.051	0.051
Threshold FE	Yes	Yes
R^2	0.008	0.009
Observations	8141	9260

Continued

	Panel B: Lending activity			
Bandwidth	±200		±250	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
PMGSY Road	2.206***	3.121***	2.075***	3.328***
	(0.380)	(0.309)	(0.383)	(0.505)
Control group mean	5.46	4.96	5.46	4.96
Controls	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.14	0.10	0.14	0.089
Observations	759	759	792	792

Continued

	Panel C: Loan maturity and quality					
Bandwidth	±200			±250		
	(1) Ln(Maturity)	(2) ODAmount	(3) %OD Amount	(4) Ln(Maturity)	(5) ODAmount	(6) %OD Amount
PMGSY Road	-0.006	11.142	-0.106	-0.019***	12.281	-0.216
	(0.011)	(72.175)	(0.105)	(0.008)	(98.101)	(0.284)
Control group mean	0.88	32.3	0.10	0.88	32.3	0.10
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes	Yes	Yes
Loan purpose FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.92	0.10	0.040	0.92	0.090	0.040
Observations	759	759	759	792	792	792

Continued

	Panel D: Lending Quantities by loan type, Bandwidth ±200			
	Productive Loans		Non-Productive Loans	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
PMGSY Road	2.335***	3.283***	-1.938**	-2.867***
	(0.177)	(0.105)	(0.971)	(0.490)
Control group mean	5.22	4.71	6.57	5.81
Controls	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.14	0.10	0.11	0.034
Observations	759	759	759	759

	Panel E: Lending Quantities by loan type, Bandwidth ±250			
	Productive Loans		Non-Productive Loans	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
PMGSY Road	2.377***	3.682***	-0.410	-2.141*
	(0.238)	(0.701)	(2.747)	(1.198)
Control group mean	5.22	4.71	6.57	5.81
Controls	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.13	0.079	0.17	0.073
Observations	792	792	792	792

Continued

	Panel F: Lending Quantities by productive loan type, Bandwidth ±200			
	Crop & Micro Loans		Other Productive Loans	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
PMGSY Road	2.581**	2.564**	-0.110	0.854
	(1.009)	(1.008)	(1.210)	(1.162)
Control group mean	0.99	0.97	4.23	3.73
Controls	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.15	0.12	0.15	0.12
Observations	759	759	759	759

	Panel G: Lending Quantities by productive loan type, Bandwidth ±250			
	Crop & Micro Loans		Other Productive Loans	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
PMGSY Road	3.763***	3.758***	-1.180*	0.127
	(0.930)	(0.951)	(0.654)	(0.212)
Control group mean	0.99	0.97	4.23	3.73
Controls	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.16	0.13	0.16	0.13
Observations	792	792	792	792

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

Table IA4: Impact of new roads on lending activity, robustness

The table presents robustness on the reduced form estimates from Equation 3 of the effect of new rural roads on lending activity within these villages. Panel A present results allowing for heteroscedasticity robust standard errors. Panel B presents results examining differential impacts of the two population thresholds under consideration while Panel C presents results allowing for different slopes on either side of the population threshold and allows for non-linear specification around the cutoff. The dependent variable in columns 1 and 3 is the natural logarithm of one plus average total loan amount disbursed by the bank while in columns 2 and 4 it is the natural logarithm of one plus net loan amount disbursed. For each borrower, we compute the net loan amount disbursed as the average of loan amount disbursed minus any repayment made by the end of the calendar year 2014. For panel C, we restict our analyses to the bandwidth of 250 around the population threshold. Our bank loan sample consists of individuals who had a loan with the bank by the end of the calendar year 2014. The sample includes villages that did not have paved roads at the start of our sample as recorded in the 2001 Population Census. All regressions include threshold fixed effects. For each regression, the outcome mean for the control group (villages with population below the threshold) is also reported. All regressions include threshold fixed effects. Standard errors are clustered allowing for correlations among partitions around the thresholds of the population support. They are reported below the point estimates.

	Panel A: Heteroscedasticity robust standard errors			
Bandwidth	±200		±250	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
Above Cutoff	1.137	1.609**	0.864	1.386**
	(0.793)	(0.753)	(0.743)	(0.706)
Control group mean	5.46	4.96	5.46	4.96
Controls	Yes	Yes	Yes	Yes
Thredhold FE	Yes	Yes	Yes	Yes
R^2	0.20	0.20	0.19	0.19
Observations	759	759	792	792

	Panel B: Differential impact by threshold around the cutoff			
Bandwidth	±200		±250	
	(1) TotalDisb	(2) NetDisburse	(3) TotalDisb	(4) NetDisburse
Above Cutoff (500)	1.671**	1.708***	1.688**	1.731***
	(0.609)	(0.534)	(0.629)	(0.560)
Above Cutoff (1000)	0.077	0.692**	0.073	0.726**
	(0.384)	(0.335)	(0.271)	(0.288)
Control group mean	0.53	0.48	0.53	0.48
Controls	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.24	0.24	0.23	0.22
Observations	759	759	792	792

Continued

	Panel C: Different slopes and quadratic term			
	Different Slopes		Quadratic	
	(1)	(2)	(3)	(4)
	TotalDisb	NetDisburse	TotalDisb	NetDisburse
Above Cutoff	0.318	0.926*	0.328	0.936*
	(0.613)	(0.559)	(0.537)	(0.488)
Control group mean	5.46	4.96	5.46	4.96
Controls	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes
R^2	0.23	0.22	0.23	0.22
Observations	792	792	792	792

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