Market Access and Structural Transformation: Evidence from Rural Roads in India*

Sam Asher[†] Paul Novosad[‡]

JOB MARKET PAPER

Please click here for the latest version.

January 11, 2016

Abstract

More than one billion people worldwide live in rural areas without access to the paved road network. How does the lack of such infrastructure affect rural employment and economic outcomes? We construct a comprehensive, high spatial resolution dataset of 825 million individuals in rural India to estimate the impact of a national rural road construction program that has built paved roads to over 100,000 villages since 2000. Program rules provide discontinuities in the probability of treatment at multiple village population thresholds, which we exploit using a fuzzy regression discontinuity design. Road construction to previously unconnected villages leads to a 10 percentage point reduction in the share of households and workers in agriculture, with an equivalent increase in wage labor market participation. This sectoral reallocation is concentrated among males and households with low levels of land, precisely those groups who have the lowest costs and highest returns to sectoral reallocation. Labor reallocation to wage labor is strongest in locations close to major cities, suggesting the importance of access to urban markets in the process of structural transformation. Rather than facilitating growth of nonfarm firms in treated villages, rural roads enable workers to access external labor markets. We also provide evidence for gains to multiple measures of economic outcomes. Our results suggest that poor rural transportation infrastructure is a major constraint on the sectoral allocation of labor in low income countries.

JEL Codes: O12/O18/J43.

*First version March 2013. We are thankful for useful discussions with Abhijit Banerjee, Lorenzo Casaburi, Melissa Dell, Ed Glaeser, Doug Gollin, Ricardo Hausmann, Rick Hornbeck, Clement Imbert, Lakshmi Iyer, Radhika Jain, Asim Khwaja, Michael Kremer, Sendhil Mullainathan, Rohini Pande, Simon Quinn, Gautam Rao, Andrei Shleifer, Seth Stephens-Davidowitz, Andre Veiga, Tony Venables and David Yanagizawa-Drott. We are indebted to Taewan Roh and Kathryn Nicholson for exemplary research assistance. This project received financial support from the Center for International Development and the Warburg Fund (Harvard University) and the IGC. All errors are our own.

[†]University of Oxford and Harvard University, samuel.asher@economics.ox.ac.uk

[‡]Dartmouth College

1 Introduction

Labor productivity in agriculture is generally significantly lower than in other sectors of the economy (Caselli, 2005; Gollin et al., 2014; Restuccia et al., 2008). This is particularly true in developing countries, whose economies are also characterized by high population shares living in rural areas and working in agriculture (McMillan et al., 2014). Researchers going back to Lewis (1954) and beyond have suggested that labor market imperfections prevent labor from reallocating away from agricultural cultivation towards higher productivity activities.[1] This paper focuses on one particular friction: the poor state of transportation infrastructure in low-income countries.[2] One billion people, or thirty-one percent of the world's rural population, live in settlements more than 2 km from a paved road. Ninety-eight percent of people lacking such access to outside markets and government services live in developing countries (World Bank, 2015). This paper examines the labor market consequences of high rural transport costs by estimating the causal effects of a $37 billion rural road construction program, which has provided over 100,000 Indian villages with paved connections to the wider road network. In order to estimate impacts at the village level, we assemble socioeconomic microdata on every individual in rural India (825 million observations). We find that road construction leads to reallocation of labor out of agriculture, suggesting that poor rural transportation infrastructure is a major barrier to the efficient allocation of labor in low income countries.

[1] Lewis (1954) argued that a subsistence wage above the marginal product of labor would prevent the efficient allocation of labor across sectors. Ranis and Fei (1961) formalize the Lewis model, modeling the friction as coming from agricultural workers' being paid their average rather than marginal product. Harris and Todaro (1970) argue that workers will only arbitrage expected wages across sectors or locations, leading to an urban-rural wage gap if higher urban wages are associated with greater search costs such as periods of unemployment. Other research has suggested that barriers to the reallocation of labor could result from credit constraints (Banerjee and Newman, 1993), informational frictions (Banerjee and Newman, 1998; Bryan et al., 2014), the cost of human capital acquisition (Caselli and Coleman, 2001), and insurance networks that discourage movement out of rural areas (Munshi and Rosenzweig, 2016).

[2] Gollin and Rogerson (2014) develop a multi-sector, multi-region model to show that the size of the agricultural workforce is increasing in transport costs. A calibration using data from Uganda shows that high transport costs is a quantitatively important determinant of the allocation of labor across sectors.

Due to the scarcity of high spatial resolution data and the endogeneity of road placement, the economic impacts of rural roads have proven difficult for researchers to assess. The high costs and potentially large benefits of infrastructure investments mean that road construction is likely to be correlated with both economic and political characteristics of locations.[3,4] We overcome this challenge by taking advantage of a large-scale natural experiment in an Indian national rural road construction program, which by 2015 had built over 100,000 roads to over 185,000 villages at a cost of nearly $40 billion. The implementation guidelines produce exogenous variation in road construction by generating discontinuities in the probability of road construction at two village population thresholds (500 and 1000). We exploit these population thresholds to estimate the economic impact of rural roads using a fuzzy regression discontinuity design.

To utilize village-level variation in road construction, we construct a high spatial resolution dataset that combines household and firm microdata with village aggregates describing amenities, infrastructure and demographic information. This is the first research to take advantage of the Government of India's recent socioeconomic census. We assemble the microdata from the 2012 Socioeconomic and Caste Census (SECC), which contains economic data for every individual and household in rural India. In assembling and analyzing microdata for every rural household and individual, we are able to test hypotheses that would be impossible with aggregate data or household surveys, joining a growing body of economic research that utilizes comprehensive administrative data to investigate otherwise elusive re-

[3] Brueckner (2014) uses international oil price movements to show that investment in infrastructure responds strongly to economic growth. Burgess et al. (2015) show that the ethnic homelands of Kenya presidents receive greater road investments, although this effect disappears during periods of democracy. Harding (2015) finds that road construction increases electoral support for incumbents in Ghana, while Blimpo et al. (2013) show in the cross section that politically marginalized areas across West Africa have lower levels of road infrastructure.

[4] Recent work has suggested that rural roads can have a significant effect on local economic outcomes via their impacts on agricultural land values (Jacoby, 2000; Shrestha, 2015), household income (Jacoby and Minten, 2009), and agricultural market prices (Casaburi et al., 2013). See below for a more detailed discussion of this literature.

search questions (Einav and Levin, 2014).

We find that rural roads lead to large movements of workers out of agriculture: a road is associated with a 10 percentage point decrease in agricultural cultivation and an equivalent increase in income from wage labor. These effects are driven by villages close to large cities, where a new rural road represents a larger proportional decrease in total transportation costs to external demand for rural labor and production. Roads also lead to an increase in measures of household welfare: we estimate an 8 percent increase in household earnings and a 20 percent increase in the share of households living in houses with a solid roof and walls. Gains in income are supported by an increase in the growth rate of night light luminosity following road construction. Looking within the village, these impacts are most pronounced among groups with the lowest costs and highest potential gains from participation in labor markets: households with small landholdings and male workers.

We argue that our results are best explained by the increased access of rural households to labor markets beyond the village. We consider two alternatives to this story: (i) increases in agricultural productivity reduce demand for labor, and (ii) within-village nonfarm sectoral growth induces movement out of agriculture. We find no evidence for increases in the size of landholdings or increases in agricultural mechanization and consolidation, suggesting that investments in agriculture have not led to reduced demand for agricultural labor. There is also no evidence to support a large increase in nonfarm economic activity in treated villages. Finally, we show that labor reallocation out of agriculture is greatest in areas with high rural-urban wage gaps, where we expect the highest returns to participation in external labor markets.

This paper contributes to multiple strands of research in economics. First, we contribute to a large literature seeking to understand the determinants of structural transformation in the process of development. It is well established that across the developing world, labor productivity outside agriculture is much higher than within agriculture (Gollin et al., 2014;

McMillan et al., 2014). This paper provides evidence that poor transportation infrastructure is an important barrier to the reallocation of labor out of agriculture and entrance into wage labor markets. This should not be surprising: for sectoral arbitrage to occur, there must be both an "agricultural productivity gap" (Gollin et al., 2014) and sufficiently low costs to reallocating labor, land and capital that it is profitable to do so. This paper lends support to the argument that transportation costs are an important component of labor market search costs in developing countries and can pose a major barrier to the spatial and sectoral allocation of labor (Bryan et al., 2014; Bryan and Morten, 2015).

Our work also complements a related literature examining the constraints to labor market participation in developing countries. Workers in low income countries are far more likely be either self-employed or work in informal firms, which have been shown to have low growth and productivity relative to firms in the formal sector (La Porta and Shleifer, 2014). The majority of self-employment and informality is in the agricultural sector. We show that transport infrastructure provision can play a large role in increasing labor force participation and earnings, consistent with existing research on electrification (Dinkelman, 2011).

Second, we add to a growing body of research that seeks to estimate the causal effects of transport infrastructure in low- and middle-income countries. Utilizing various creative identification strategies, these papers find economically meaningful effects of transportation projects across a wide range of outcomes. Transportation infrastructure has been shown to raise the value of agricultural land (Donaldson and Hornbeck, 2015), increase agricultural trade and income (Donaldson, 2012), reduce the risk of famine (Burgess and Donaldson, 2012), increase migration (Morten and Oliveira, 2014) and accelerate urban decentralization (Baum-Snow et al., 2015). However, results have also proven somewhat mixed: there is evidence that reducing transportation costs can increase (Ghani et al., 2015; Storeygard, 2014), decrease (Faber, 2014) or leave unchanged (Banerjee et al., 2012) growth rates in local economic activity. These papers have largely focused on highways and railroads.

We add to this literature by providing some of the first causal estimates of the impact of smaller scale roads to rural areas, as well as providing detailed estimates of the response of households (rather than firms or aggregate measures of economic activity) to the construction of transport infrastructure. Our findings complement existing studies by documenting the impact of roads on the allocation of labor across sectors.

Third, we contribute causal estimates to the literature that examines the economic impacts of rural roads specifically. Such intra-regional roads differ in multiple ways from inter-regional transport infrastructure such as railroads and major highways. As they do not affect transport costs between cities, they are unlikely to have the same impacts that the literature has found on firm location choices, productivity and income. They also lower transport costs to rural areas often lacking complementary infrastructure such as electricity. We add to this literature in several ways. This paper is the first large-scale study on rural roads that combines household microdata with exogenous variation from program rules; in this regard we join recent work that has estimated the impacts of major infrastructural investments such as dams (Duflo and Pande, 2007) and electrification (Lipscomb et al., 2013).[5] While most research has focused on agricultural outcomes, we demonstrate the large impacts that road construction can have on sectoral reallocation away from agricultural activity. We

[5] An older literature suggested that rural transport infrastructure was highly correlated with positive development outcomes (Binswanger et al., 1993; Fan and Hazell, 2001; Zhang and Fan, 2004), estimating high returns to such investments. More recent work has generally demonstrated that rural roads are associated with large economic benefits by looking at their impact on agricultural land values (Jacoby, 2000; Shrestha, 2015), estimated willingness to pay for agricultural households (Jacoby and Minten, 2009), complementarities with agricultural productivity gains (Gollin and Rogerson, 2014), and search and competition among agricultural traders (Casaburi et al., 2013). Most closely related are papers that estimate the impact of rural road programs in Bangladesh (Khandker et al., 2009; Khandker and Koolwal, 2011; Ali, 2011), Ethiopia (Dercon et al., 2009), Indonesia (Gibson and Olivia, 2010), Papua New Guinea (Gibson and Rozelle, 2003) and Vietnam (Mu and van de Walle, 2011). Existing research on the PMGSY demonstrates a strong relationship between PMGSY road construction and changes in human capital formation, agricultural technology adoption and price dispersion (Aggarwal, 2015). Other papers also suggest that the lack of rural transport infrastructure may be a significant contributor to rural underdevelopment. Wantchekon and Stanig (2015) provide evidence that transport costs are a strong predictor of poverty across sub-Saharan Africa. Fafchamps and Shilpi (2005) offer cross-sectional evidence that villages closer to cities are more economically diversified, with residents more likely to work for wages.

also show the role that proximity to cities plays in determining the impacts of transport infrastructure investments. Finally, much of this literature has generated estimates from very small samples; our large sample both argues for a higher degree of external validity and allows us to investigate how location and household characteristics mediate the effects of rural connectivity.

The rest of the paper proceeds as follows: Section 2 provides the conceptual framework for how rural road construction may affect local economic activity and labor force participation. Section 3 provides a description of the rural road construction program. Sections 4 and 5 describe the data construction and empirical strategies. Section 6 presents results and discussion. Section 7 concludes.

2 Conceptual Framework

The construction of a paved road to a village may change the nature of economic activity via numerous channels. We expect roads to lower transportation costs for labor, capital and goods, as well as for information. We begin by laying out a general model of occupational choice, followed by a discussion of how it relates to various urban and trade models.

2.1 Model

There is a continuum of villages characterized by market access $a \in [0, \bar{a}]$. Within villages, there is a continuum of agents characterized by $\theta \sim \mathcal{U}[0, 1]$, where θ captures the relative productivity of a worker in cultivation as compared to manual labor. Each individual maximizes earnings by deciding between two occupations: cultivation and manual labor. Occupation-specific earnings in agriculture (y) and labor market participation (w) are represented by the following equations:

$$y = \theta g(a)$$

$$w = w(a)$$

Agents choose cultivation if earnings in agriculture are higher than in wage labor: $\theta g(a) > w(a)$. There is a marginal farmer of type $\tilde{\theta} = \frac{w(a)}{g(a)}$ who is indifferent between cultivation and labor; all those with $\theta > \tilde{\theta}$ will work in cultivation. We can thus represent the share of village labor working in cultivation as

$$q = q(a) = 1 - \tilde{\theta} = 1 - \frac{w(a)}{g(a)}.$$

We understand a rural road as increasing market access a. Differentiating $q(\tilde{\theta})$ by a tells us how agricultural employment will change with an increase in market access. Intuitively, there are two potentially countervailing effects. The first is that market access could change the productivity of cultivation, e.g. through access to inputs or through higher farmgate prices. The second is that market access changes the returns to wage labor. Potential mechanisms here include lower transport costs to work in the city, lower search costs and higher in village productivity among firms that demand labor. The impact of the road will depend on the relative strength of these forces. Mathematically:

$$\frac{\partial q}{\partial a} = \frac{g'w - gw'}{g^2}.$$

As the denominator is positive, the sign of this will depend on $g'w \lesseqgtr gw'$. For a road (i.e. increase in market access) to induce movement out of agriculture, it must be that $\frac{w'}{w} > \frac{g'}{g}$. Multiplying both sides by a yields the inequality in terms of elasticities: for an increase in market access to induce movement from the cultivation to wage sector, it must be that the elasticity of wages with respect to market access is greater than the same elasticity for profits from cultivation:

$$\frac{\partial q}{\partial a} \leq 0 \leftrightarrow \varepsilon_w \geq \varepsilon_g.$$

Following Redding and Turner (2015), we assume that $w'(a) > 0$.[6] We also assume that $g'(a) > 0$. In the general framework we have set up here, we can also see that q is not necessarily monotonic in a. Taking the second derivative, we can see that q is concave when the convexity of w relative to its level is large relative to the same term for g:

$$q'' < 0 \leftrightarrow \frac{w''}{w} - \frac{g''}{g} > 2\frac{g'}{g}(\frac{w'}{w} - \frac{g'}{g}).$$

2.2 Discussion

The general framework described above does not explicitly model the forces that determine the returns to cultivation and labor market participation or how these vary with market access. In this section we describe various predictions from the urban and trade literatures regarding the impacts of rural road construction on the allocation of labor across sectors.

Models from urban economics would predict that rural road construction would facilitate movement out of agricultural and into urban labor markets. The Alonso-Muth-Mills model predicts an urban perimeter beyond which labor will only be used in agriculture, as urban earnings net of commuting costs are lower than agricultural income (Brueckner, 1987). If we think of rural road construction as a reduction in commuting costs for a given location, we expect treatment to expand the urban perimeter. This model thus predicts that labor will leave agriculture, but only in villages sufficiently close to cities to allow commuting.[7]

Trade models are more agnostic on the impact of rural road construction on sectoral allocation. If roads to previously unconnected villages can be thought of as transforming a closed

[6] This is intuitive if real wages in the village are simply the urban wage minus transport costs, or if increasing market access raises in-village labor productivity. However, it is possible to imagine alternate scenarios, e.g. where increased market access allows workers from poorer regions to enter local labor markets, driving down wages.

[7] Short-term migration is also a possibility. While not considered in the canonical urban models, it is a prominent feature of rural economic activity in developing countries such as India (Imbert and Papp, 2015) and would likely result in a similar reallocation away from agriculture. Kochar (1999) and Colmer (2015) provide evidence that workers in agriculture reallocate labor to non-agricultural activities in response to adverse weather shocks.

village economy into an open economy, we should expect that the village economy will specialize in its comparative advantage. As pointed out by Matsuyama (1992), the comparative advantage for poor places may be very well be in agriculture, but this is not necessarily the case if roads allow villages to trade with agriculturally more productive locations. Certain trade models focused on structural transformation include labor market frictions that create a wedge between agricultural and non-agricultural wages (Tombe, 2014). If road construction is interpreted as a reduction in this friction, treatment should induce greater movement of labor into agriculture as the wage increases. However, if this friction is actually between between rural and urban wages, then the take-home wage of rural commuters to outside work will increase, and the predictions for the sectoral allocation of labor are ambiguous.

Of course, it is possible that roads not only facilitate trade but also change the productivity of labor both in and out of agriculture. First, we expect agricultural productivity to increase due to lower transportation costs for importing inputs such as fertilizer and exporting agricultural output.[8] But it is unclear whether the relative productivity of labor in agriculture will rise or fall, as other rural sectors should also experience such productivity gains. Second, workers are likely to experience a fall in search and commuting costs, reducing the barriers to working outside of the village. This represents an increase in labor demand, which should translate into an increase in village wages.[9] Given that labor productivity has been shown to be higher outside of agriculture across a wide range of countries (referred to as the "agricultural productivity gap" (Gollin et al., 2014)), we find it likely that this demand comes primarily from non-agricultural activities. Road construction is thus likely to increase the ability of workers to arbitrage the productivity gap, leading to the reallocation

[8] Sotelo (2015) estimates that paving existing roads will on average boost agricultural productivity by 15% by both increasing access to inputs and raising output prices. These changes both increase productivity directly and induce greater specialization.

[9] We acknowledge the possibility that roads actually lower wages due to an increase in labor supply. Given large rural-urban wage gaps and that our sample villages are smaller, more remote and have fewer amenities than the Indian averages, we expect the labor demand effect to dominate the supply effect.

of labor (and perhaps land and capital) away from agriculture. Higher wages in the village may then cause a shift in the sectoral composition of village employment, which will depend on the slopes of the demand curves for labor in and out of agriculture. It is also possible that there is an income effect in which workers exit work in which the marginal utility of earnings is now less than that of leisure.

Due to all of the factors discussed above, we expect the impact of roads on economic activity to depend strongly on characteristics of both the village and individual. Most obviously, we expect treatment effects to be largest where the treatment intensity is greatest. There are multiple reasons to expect that this will be the case close to cities. Rural road construction likely represents the largest proportional decrease in transport costs between a village and demand for rural labor and production. Further, commuting to work in cities is only possible in areas close enough to cities that daily trips are feasible.[10]

As the major input into agricultural production is land, theory would predict that households with larger landholdings would be less likely to exit agricultural cultivation than households with smaller landholdings. However, households with smaller landholdings are also less likely to engage in cultivation as their primary income source. This would argue for examining the proportional decline in cultivation, rather than the level, as the outcome of interest when considering heterogeneity by household landholdings.[11]

The effect of road construction on occupation choice may also depend on individual characteristics such as age and gender. Given men's advantage in physical labor, we would expect them to specialize in activities that are more intensive in physical strength. This could be in

[10]Sharma and Chandrasekhar (2014) use data from the National Sample Survey to estimate that over eight million workers commute from rural to urban areas in India every day, and half as many make the opposite commute.

[11]A positive relationship between household agricultural productivity and landholding could emerge either from functioning land markets that allow more productive farmers to accumulate more land, or from market failures in land and labor that (in the limit) restrict households to farming only the land they own. Foster and Rosenzweig (2011) show that Indian farms are inefficiently small, suggesting potential unrealized gains from consolidation.

either agriculture or manual labor, depending on the technologies used for each. Attitudes against women's spending time far away from home, as well as their greater responsibilities in house work and child raising, may diminish any reallocation of female labor away from agriculture and into the labor market (Goldin, 1995).

For various reasons, the transition from agriculture may depend on age. First, younger workers they have less sector-specific (and perhaps location-specific (Bazzi et al., 2014)) experience in agriculture, and thus the opportunity cost of working outside of agriculture is lower. Second, younger workers may have lower search costs, due to such factors as the absence of children. Third, younger workers may have superior human capital. Finally, existing evidence lends credence to this prediction: studying South Korea's rapid industrialization, Kim and Topel (1995) find that non-agricultural firms almost exclusively hired new entrants to the labor force; in other words, South Korea experienced rapid structural transformation at the aggregate level with little sectoral reallocation at the individual level.

Although most trade theories predict that lowering barriers to trade will generally increase overall income, the effect of road construction on poverty is theoretically ambiguous as a road may have countervailing effects on demand for inputs and outputs. By lowering the cost of exporting, a road will increase demand for inputs (such as labor) and final products. By lowering the cost of importing, a road increases competition, potentially reducing demand for goods produced in the village. Which effect dominates is theoretically unclear. Recent work has provided strong evidence that there are likely to be losers as competition and access change the returns to different assets and skills. Depending on the distribution of these assets, roads could induce an increase in poverty, particularly if the adjustment costs are high.[12] Further, if wages rise more than output prices, labor intensive farms and firms may actually become less profitable.

[12] Autor et al. (2014) find that workers in U.S. manufacturing industries most exposed to Chinese competition garner lower earnings and experience more job churn, with greater losses for workers with low wages and tenure, precisely those whose work is most substitutable with Chinese labor.

Road construction could also influence migration decisions via multiple, potentially cross-cutting mechanisms. There may be net migration towards areas with rural roads, which are now more appealing places to live. However, roads may lower the cost of migration from rural areas and thus induce greater outmigration (Morten and Oliveira, 2014; Bryan et al., 2014). In the presence of migration, changes in the composition of local economic activity and poverty may be attributable to changes in the composition of the population, rather than to sectoral reallocation or higher earnings for the baseline residents of the village.

3 Context and background

The Pradhan Mantri Gram Sadak Yojana (PMGSY) – the Prime Minister's Village Road Program – was launched in 2000 with the goal of providing all-weather access to unconnected habitations across India.[13] The focus was on the provision of new feeder roads to localities that did not have paved roads, although in practice many projects under the scheme upgraded pre-existing roads. As the objective was to connect the most locations to the external road network at the lowest possible price, link routes (terminating at a village) were to be given priority over through routes (those passing through a village to another larger road).

National guidelines determine prioritization of road construction under the PMGSY. Most importantly for this paper, road construction is supposed to occur first in large habitations, as defined by the 2001 Population Census. Originally, the stated goal was to provide all habitations with populations greater than 1000 with connectivity by 2003 and all habitations with population greater than 500 with connectivity by 2007. These thresholds were to be lower in desert and tribal areas, as well as hilly states and districts affected by left-

[13] Habitations are defined as clusters of population whose location does not change over time. They are distinct from, but form parts of, revenue villages used by the Economic and Population Censuses. In this paper, we aggregate all data to the level of the revenue village. See National Rural Roads Development Agency (2005) for more details.

wing extremism.[14] These rules were to be applied statewise, meaning that states that had achieved connectivity of all larger habitations could proceed to constructing roads in smaller localities. However, program guidelines also laid out other rules that states could use to determine allocation. Smaller villages could be connected if they lay in the least-cost path of connecting a priority habitation. Groups of habitations could combine their populations if they lay within 500 meters of each other. Members of Parliament and state Legislative Assemblies were also allowed to make suggestions that would be taken into consideration when approving construction projects. Finally, measures of local economic importance such as the presence of a weekly market were also considered relevant.

Although funded and overseen by the federal Ministry of Rural Development, responsibility for road construction is delegated to state governments. District Rural Road Plans were drafted for every district in India, delineating a "core network" of roads that would be required to connect every habitation to the paved road network at the lowest possible cost. Funding comes from a combination of taxes on diesel fuel (0.75 INR per liter), central government support and loans from the Asian Development Bank and World Bank. By 2015, over 400,000 km of roads had been constructed, benefiting 185,000 villages – 107,000 previously lacking an all-weather road – at a cost of more than $37 billion.[15]

4 Data

In order to estimate the economic impacts of PMGSY road construction, it was necessary to construct a unique village-level dataset that combines administrative data from the PMGSY program with aggregate and micro-data from multiple sources. In this section we describe

[14] Our calculations suggest that few states followed these guidelines in the first five years of the program, and some states never adopted them at all. As explained later, we restrict our sample to the set of states that did follow these population thresholds.

[15] Author's calculations from official PMGSY administrative data. We use an exchange rate of 48.01 INR per USD, the average for the period between January 1, 2000 and January 1, 2015.

the data sources and collection process.

4.1 PMGSY

Program administrative data on the PMGSY are generated and reported through the Online Management and Monitoring System (OMMS), the software used in program tracking and implementation. Variables include road sanctioning and completion dates, cost and time overruns, contractor names, and quality monitoring reports.

PMGSY data are posted online at either the habitation or the road level.[16] There is a many-to-many correspondence between habitations and roads: roads serve multiple habitations, and habitations may be connected to multiple roads. A census village typically comprises between one and three habitations; approximately 200,000 villages, one third of the total, consist of only a single habitation. For the purposes of this paper, all variables are aggregated to the level of the census village, the geographic unit at which we measure outcomes. We consider a village to be treated by the PMGSY if at least one habitation in the village received a completed PMGSY road by the year before data collection.

4.2 Socioeconomic census

The primary outcomes presented in this paper come from individual- and household-level microdata from a national socioeconomic census. Beginning in 1992, the Government of India has conducted multiple household censuses in order to determine eligibility for various government programs (Alkire and Seth, 2013). In 1992, 1997 and 2002, these were referred to as Below Poverty Line (BPL) censuses. We obtained the anonymized microdata to the 2002 BPL Census from the Ministry of Rural Development. This dataset contains age, gender, education and caste group (at the individual level) as well as various measures of household

[16] All data are publicly available at http://omms.nic.in. The variables used in this paper were assembled from data scraped in January 2015.

economic activity and well-being. Households that were automatically considered above the poverty line were not included in this dataset.

The fourth such census, the Socioeconomic and Caste Census (SECC), was launched in 2011 but primarily conducted in 2012. This survey departed from the previous methodologies by collecting data on all households, even if they demonstrated characteristics that would exclude them from eligibility under various government schemes targeted at the poor.[17] To increase the likelihood of collecting data on all individuals and households, it is based on the National Population Register (NPR) from the 2011 Population Census.

The Government of India has made the SECC publicly available on the internet in PDF and Excel form. In order to construct a useful microdataset, we scraped over two million files, parsed the embedded into text data, and translated these from twelve different India languages into English. At the individual level, these data contains variables describing age, gender, occupation, caste group, disability and marital status. Data on occupations are written freeform in the SECC; after translation we cleaned and matched these descriptions to the 2004 National Classification of Occupations. At the household level, this dataset contains variables describing housing, landholdings, agricultural assets, household assets and sources of income. We are able to match these data to our other datasets at the village level. This dataset is unique in describing the economic conditions of every person and household in rural India, at a spatial resolution unavailable from comparable sample surveys.

4.3 Economic and population censuses

The Indian Ministry of Statistics and Programme Implementation (MoSPI) conducted the 4th and 5th Economic Censuses respectively in 1998 and 2005.[18] The Economic Census

[17]It is often referred to as the 2011 SECC, as the initial plan was for the survey to be conducted between June and December 2011. However, various delays meant that the majority of the surveying was conducted in 2012, with urban surveys continuing to undergo verification at the time of writing. We therefore use 2012 as the relevant year for the SECC.

[18]The 6th Economic Census, conducted primarily in 2012, has not yet been released at the time of writing.

is a complete enumeration of all economic establishments except those engaged in crop production and plantation; there is no minimum firm size, and both formal and informal establishments are included.

The Economic Census records information on the town or village of each establishment, whether ownership is public or private, the number and demographic characteristics of employees, the sources of electricity and finance, and the caste group of the owner. The main product of the firm is also coded using the 4-digit National Industrial Classification (NIC), which corresponds roughly to a 4-digit International Standard Industrial Classification (ISIC) code. More detailed information on income or capital is not included. The main strengths of the data are its comprehensiveness and rich detail on spatial location and industrial classification of firms. We obtained location directories for the Economic Censuses, and then used a series of fuzzy matching algorithms to match villages and towns by name to the population censuses of 1991 and 2001.[19] We were able to match approximately 93% of villages between 1998 and 2005.

We also make extensive use of data from the Population Censuses of 1991, 2001 and 2011. In addition to basic demographic data, the Population Census contains variables describing local public infrastructure (roads, electricity, schools and hospitals) and household assets, all aggregated to the village level.

4.4 Other data

In addition to the socioeconomic, population and economic censuses, we use cross-sectional data from the 68th (2011-12) Round of the National Sample Survey (Employment/Unemployment), which contains far fewer villages and individuals than our socioeconomic census data, but includes data on earnings, place of work and time use across primary and secondary occupa-

[19]The Economic Census of 1998 was conducted with the house listing for the 1991 population census, while the 2005 Economic Census used codes from the 2001 population census.

tions. Using village populations backed out from the sample weights, we match observations from the National Sample Survey to the rest of our village-level data.

We use village and town latitude and longitude from ML Infomap to generate measures of straight line distances from villages to cities and highways as a proxy for market access. Highway GIS data come from both OpenStreetMap and the National Highways Authority of India.[20]

We downloaded gridded average annual night light data from the web site of the National Oceanic and Atmospheric Administration, and matched the grid cells to constituency polygons and election years.[21] Night lights are a proxy for economic growth that have the advantage of high resolution and objective measurement over a 20+ year period (Henderson et al., 2011). Their weakness is that they may be biased by factors affecting light but not output, such as electricity supply. Villages are assigned the value of the pixel in which their centroid is located. For years in which we have observations from multiple satellites, we take the average. We definite light growth as the annual increase in log luminosity, with the value 1 added to the level before taking logs in order to not lose observations of luminosity level 0.

4.5 Summary statistics

We matched PMGSY data to economic, population and poverty census data at the village level. In order to generate a village correspondence across multiple datasets, we developed a Hindi-language fuzzy matching algorithm to match differently spelled village names. We successfully match over 85% of habitations listed in the PMGSY to their corresponding population census villages.

Table 1 shows village-level summary statistics for the entire sample of Indian villages.

[20]We gratefully acknowledge Ejaz Ghani, Arti Goswami and Bill Kerr for generously sharing the GIS data on the Golden Quadrilateral highway network with us.

[21]We calibrated the data to best rationalize the changing sensitivity of luminosity sensors over time and across satellites; but this calibration does not affect results as all our specifications include year fixed effects. Luminosity is measured on a top-coded 64 point scale.

The first column shows results for villages without a paved approach road in 2001, the second column for villages with a paved approach road, and the third column for the pooled sample. Over 25% of villages without paved roads in 2000 received a PMGSY road by 2012.[22] Across a wide range of variables, villages without roads have lower levels of other amenities. They are further from towns, have higher illiteracy rates and are half as likely to be electrified at baseline. Inhabitants of unconnected villages are also much more likely to work in agriculture: 83% of workers in villages lacking paved roads worked in agriculture in 2001, compared to 74% in villages with paved roads. These differences lend further evidence to our assertion of endogenous placement of transport infrastructure, and thus the need for careful empirics to identify the causal effect.

Figure 1 provides a visual representation of the major datasets used in this project, along with year-by-year counts of the number of villages receiving PMGSY roads for the years of this investigation (2000 - 2012). It demonstrates that PMGSY construction is negligible before our baseline data in 2001, then slowly ramps up to a peak of almost 16,000 villages connected annually in 2009 before slowing down slightly.

5 Empirical Strategy

The impacts of infrastructural investments have often proved challenging for economists to assess. First, the high cost and large potential returns of such investments mean that few policymakers are willing to allow random targeting. Political favoritism, economic potential and pro-poor targeting would lead infrastructure to be correlated with other government programs and economic growth, leading to bias in OLS estimates of road construction. Second, data are rarely available at the level of road construction, particularly in the case of

[22] Nearly 21% of villages that were recorded as having a paved road in the 2001 Population Census also received PMGSY roads by 2012. This appears to have been due both to measurement error in the Population Census variables and to upgrades that were performed on existing roads.

rural roads. Third, the impacts of infrastructure are likely to depend on local and regional economic factors, necessitating a large sample to have sufficient power for tests of heterogeneity. In this section we describe the empirical strategy for the estimation of unbiased estimates of the impact of the PMGSY road construction program.

Identification comes from the guidelines by which villages are prioritized for PMGSY road construction. State implementing officials were instructed to target habitations in the following order: (i) habitations with population greater than 1000; (ii) habitations with populations greater than 500; and (iii) habitations with populations greater than 250. Even if selection into PMGSY treatment is partly determined by political or economic factors, these factors are not likely to change discontinuously at these population thresholds. If these rules were followed to any degree by state officials, the likelihood of PMGSY treatment will discontinuously increase at these population thresholds, making it possible to estimate the effect of the program using a fuzzy regression discontinuity design.

Under the assumption of continuity at the treatment threshold, the fuzzy RD estimator (Imbens and Lemieux, 2008) estimates the local average treatment effect (LATE) of receiving a new road, for a village with population equal to the threshold:

$$\tau = \frac{\lim_{pop \to T^{+}} \mathbb{E}[Y_v | pop_v = T] - \lim_{pop \to T^{-}} \mathbb{E}[Y_v | pop_v = T]}{\lim_{pop \to T^{+}} \mathbb{E}[newroad_v | pop_v = T] - \lim_{pop \to T^{-}} \mathbb{E}[newroad_v | pop_v = T]}, \tag{1}$$

where pop_v is the baseline village population, T is the threshold population, and $newroad_v$ is an indicator variable for whether village v received a new road in the sample period. The treatment effect can be interpreted as the discontinuous change in the outcome variable at the population threshold (the numerator) divided by the discontinuous change in the probability of treatment (the denominator).[23] The LATE estimated by our empirical design

[23] Our design is a "fuzzy" regression discontinuity design (RDD) because the change in the probability of treatment at the threshold is less than one. Due to both other program rules guiding road prioritization and imperfect compliance with program rules, neither is the probability of treatment below the threshold zero nor the probability of treatment above the threshold one.

is specific to the complier set, namely those villages whose treatment status would be zero with population below the threshold and one with population above.

Our estimation follows the recommendations of Imbens and Lemieux (2008), Imbens and Kalyanaraman (2012) and Gelman and Imbens (2014). Our preferred specification uses local linear regression to control for the running variable (village population) on either side of the threshold. We restrict our sample to those villages whose population is within a certain bandwidth around the threshold, formally $pop_v \in [T-h; T+h]$, where h is the value of the bandwidth around threshold T. We calculate an optimal bandwidth of 85.3 following Imbens and Kalyanaraman (2012) and use a triangular kernel that places the most weight on observations close to the cutoff, as in Dell (2010).[24] Controls and fixed effects are not necessary for identification, but their inclusion increases the efficiency of the estimator.

We begin by estimating the following reduced form fuzzy RDD specification:

$$Y_{v,j} = \beta_0 + \beta_1 1\{pop_{v,j} \geq T\} + \beta_2 pop_{v,j} + \beta_3 pop_{v,j} * 1\{pop_{v,j} \geq T\} + \zeta X_{v,j} + \eta_j + \epsilon_{v,j}, \quad (2)$$

where $Y_{v,j}$ is the outcome of interest, T is the population threshold, $pop_{v,j}$ is baseline village population, $X_{v,j}$ is a vector of village controls measured at baseline, and η_j is a group fixed effect. Village controls and fixed effects are not necessary for identification but improve the efficiency of the estimation. The change in outcome $Y_{v,j}$ for a village at the population threshold T is captured by $\beta_1 + \beta_3 * T$. For ease of exposition, we subtract the threshold value T from the population variable, such that $T = 0$, and β_1 fully describes the change in outcome $Y_{v,j}$ at the treatment threshold.

We make the following choices when estimating this model. In the first stage regression, in which we estimate the change in the probability of treatment, $Y_{v,j}$ is a dummy variable that

[24] Results are robust to alternate bandwidths, as described below. Following the methodology of (Calonico et al., 2014), the optimal bandwidth is 63.0, which does not appreciably change the results. Results using alternative weighting functions and thresholds are available from the authors upon request.

takes on the value one if the village has received a PMGSY road before 2012, the year of our primary outcome data.[25] For regressions in which we estimate the reduced form effect of road prioritization (i.e. being to the right of the population threshold) on economic outcomes, we discuss the definition of outcome variables as we present the results in Section 6. The vector of village controls, $X_{v,j}$, contains various village characteristics as measured in the 2001 Population Census: indicators for village amenities (primary school, medical center and electrification), the log of total agricultural land area, the share of agricultural land that is irrigated, distance in km from the closest census town, share of workers in agriculture, the illiteracy rate and the share of inhabitants that belong to a scheduled caste. For η_j, we use district-cutoff fixed effects.[26]

We understand the reduced form effect of road priority to be treatment effect of a new road times the change in the probability of road treatment at the population threshold. To estimate the treatment effect directly, we use the following fuzzy RDD specification in which we instrument for treatment ($newroad_{v,j}$) with our road priority dummy $1\{pop_{v,j} \geq T\}$.

$$Y_{v,j} = \gamma_0 + \gamma_1 newroad_{v,j} + \gamma_2 pop_{v,j} + \gamma_3 pop_{v,j} * 1\{pop_{v,j} \geq T\} + \zeta X_{v,j} + \eta_j + \upsilon_{v,j}. \quad (3)$$

We estimate this equation using two stage least squares, where the first stage comes from Equation 2, with $newroad_{v,j}$ as the dependent variable.

As the objective of this paper is to estimate the economic impacts of receiving a paved road for the first time, we restrict our sample to villages that did not have a paved road at the start of the program.[27] The PMGSY used multiple population thresholds to determine road

[25] This is the year that most data was collected for the SECC. When estimating outcomes measured in a different year, such as in the Population Census, we use the appropriate year of measurement for that particular set of regressions.

[26] Results are robust to alternative specifications using state or district fixed effects, and are available from the authors upon request.

[27] While unconnected villages were to be prioritized over those that already had some paved road, many already connected villages still received roads under the program. This is partly because road upgradation was also allowed under the rules and partly because program rules were not entirely followed. We define our

prioritization: 1000, 500 and 250. Very few villages around the 250 population threshold received roads by 2012, so we limit our sample to villages with populations close to 500 and 1000. Further, only certain states followed the population threshold prioritization rules as given by the national guidelines of the PMGSY. We worked closely with the National Rural Roads Development Agency to identify the state-specific thresholds that were followed and define our sample accordingly. Our sample is comprised of villages from the following states, with the population thresholds used in parentheses: Chhattisgarh (500, 1000), Jharkhand (1000), Madhya Pradesh (500, 1000), Maharashtra (500, 1000), Orissa (500, 1000), Rajasthan (500), and Uttar Pradesh (500).[28] To maximize power, we pool our samples, using the same optimal bandwidth (85.3) for villages close to the 500 and 1000 thresholds.

The fuzzy regression discontinuity approach identifies the treatment effect of rural road construction under the assumption that crossing the population threshold affects the probability of receiving a road, and nothing else of significance. We follow Imbens and Lemieux (2008) in testing for discontinuities in baseline covariates and in the density of the running variable at the population thresholds. Other threats to identification that rely on outcome variables are discussed below, in Section 6.2.

We first show that there are no discontinuities in baseline village characteristics. Table 2 presents the mean values for various village baseline characteristics, including the set of controls that we use in all regressions. Unsurprisingly, there are differences between the villages above and below the population threshold, as many village characteristics are correlated with village size. Reassuringly, however, we find no significant differences once we control for the covariates used in the fuzzy RDD specification. Figure 3 shows how our control variables and other village characteristics vary at the cutoff, plotting the residuals after controlling

sample of unconnected villages to be those that were recorded as lacking a paved road in either the 2001 Population Census (whose village amenities were recorded in 2000) or the PMGSY administrative data.

[28]Students of Indian geography will notice that these states are concentrated in north India. Southern states generally have far superior infrastructure and thus had few unconnected villages to prioritize. Other states such as Bihar had many unconnected villages but did not comply with program guidelines.

for the set of controls (excluding the one in question, running variable controls and the road priority dummy) and fixed effects used in our main specification against normalized village population. The black lines show a linear fit, estimated separately on either side of the cutoff, and the grey lines give the 95% confidence interval. Again, no significant differences in village characteristics can be observed.

We also investigate the possibility of manipulation of the running variable. We find evidence of considerable manipulation of village population in the official program data.[29] To resolve this issue, we instead use village population from the 2001 Population Census. Figure 2 displays two representations of the distribution of village populations in our sample, using data from the Population Census. In the left panel, there are no noticeable discontinuities at the PMGSY population cutoffs. We test this formally by testing for a discontinuity in the running variable (village population) around the population threshold for the pooled sample, following McCrary (2008). We estimate a discontinuity of .03 with a standard error of .04, failing to reject the null hypothesis of no discontinuity in the running variable.

We next examine the first stage, showing that there is a large and highly significant jump in the probability of road construction by 2012 at the population cutoff. Table 3 presents first stage estimates of the change in probability of treatment across different bandwidths h. The estimates are highly stable. Across bandwidths, there is a 13 percentage point increase in the probability of treatment around the cutoff. Figure 4 shows these results graphically for the optimal threshold as a scatterplot of population bin means. This graph confirms the results from Table 3: at the population threshold, there is a significant increase in the probability of treatment of approximately 13 percentage points.

[29] Figure A2 shows the distribution of village population as reported to the PMGSY, with implementation cutoffs indicated with vertical lines. There are noticeable discontinuities in density at the implementation cutoffs, suggesting that selection into treatment is not as good as random around these population cutoffs—for example, villages that are politically connected or more strategic may be able to report their population as just above 1000, even if it is not in reality. If this is occurring, the RDD approach cannot distinguish the effect of a new road from the effect of being politically connected.

6 Results

In this section, we describe and discuss the main results (Section 6.1), robustness (Section 6.2) and the evidence on the mechanism (Section 6.3). We first show that rural road construction leads to a reallocation of labor out of self-employment in agriculture and into manual labor. This result is driven by villages close to major cities, as well as households and individuals with high potential returns to labor market participation: households with small landholdings and male workers. We then show that these results are not driven by alternate programs or the impact of roads on data quality. We then consider multiple mechanisms that could explain these results, finding that that the evidence best supports increased access to labor markets outside of the village.

6.1 Main results

We begin by estimating the effect of rural roads on household economic activity, as reported in the SECC. As approximately 93% of households in our sample report their primary source of income to be either agricultural cultivation or manual labor, we focus our investigation on these categories. Outcomes $Y_{v,j}$ are defined at the village level to be the share of households reporting their primary income source as either cultivation or manual labor.

We find that rural road construction is associated with a significant occupational reallocation out of agricultural cultivation and into manual labor. Table 4 presents regression discontinuity estimates of the impact of road construction on the share of households reporting cultivation and manual labor as their primary sources of income. For robustness, we present results across six different bandwidths ranging from 50 to 100.[30] We find that road construction is associated with a large and stable reduction in the share of households in cultivation (point estimates range from 8.4 to 11.1 percentage point reduction, depending

[30]This range contains both the optimal IK (85.3) and CCT (63.0) bandwidths.

on bandwidth) and corresponding increasing in the share of households in manual labor (9.9 to 12.3 percentage point increase). Figure 5 presents the reduced form estimates graphically with the optimal bandwidth, demonstrating the significant drop in the share of households in cultivation to the right of the population cutoff.

Our analysis from this point forward restricts itself to using the optimal bandwidth of 85.3. Occupational data allow us to test whether this sectoral reallocation of income is the result a sectoral reallocation of labor. Table 5 presents regression discontinuity estimates from Equation 3 of the effect of road construction on occupational choice. The first two columns present the impact of road construction on household income source, the same outcomes from the previous table but using the optimal bandwidth. The second two columns use the same specification to estimate the impact of road construction on individual occupation in agriculture and manual labor. We find similar results in the occupational data and in the income source data: an 9.3 percentage point decrease in the share of employed working age (21-60) adults who list agriculture as their occupation. This is accompanied by an increase of 8.4 percentage points in the share of working adults in manual labor. These point estimates are very close to the household-level income source estimates of -9.6 and 10.9 percentage point changes for cultivation and manual labor, respectively, suggesting that our income source results are driven primarily by the reallocation of labor across sectors rather than by changes to wages alone.

We next test our hypothesis that villages that gain the most market access will demonstrate the largest effects. Table 6 presents estimates of the road treatment by distance to urban areas. We calculate this measure as the straight line distance between a village and the centroid of the nearest major city. Villages that are below median distance from cities experience much more movement out of agriculture than those further away. This result is robust to whether we define major cities as those of 100,000+ or 500,000+ population (as measured in the 2001 Population Census), though the heterogeneity is more pronounced (the estimates are statistically difference) in the latter case. To investigate this result further, we estimate the impact of rural road construction at quartiles of distance from these large (500,000+ inhabitants) cities. Table A4 presents these results. Panel A demonstrates that the first stage does not vary significantly across different quartiles of distance to major cities. Panel B presents the reduced form estimates of road priority on cultivation, and Panel C the regression discontinuity results. We find that our results are diminishing with distance from major cities and the average effect is driven by the first quartile (villages less than 63 km from large cities).[31,32]

Supporting evidence is found when considering an alternate specification that considers distance to highways. Here we take advantage of the Golden Quadrilateral, India's premier highway network that was upgraded between 2001 and 2007 to connect the country's four largest metropolitan areas.[33] As above, we divide our sample into villages that were above and below median distance to the Golden Quadrilateral, restricting ourselves to only states through which it passes. Table A10 presents reduced form results of the effect of road priority on cultivation share.[34] Column 1 repeats our main reduced form finding that road priority is associated with a significant decrease in the share of households in cultivation, in this case for the sample of states through which the Golden Quadrilateral passes. Column 2 adds an interaction of road priority with below-median distance to the Golden Quadrilateral. This specification suggests that road treatment only lowers cultivation in places close to the Golden Quadrilateral. Column 3 shows similar findings for the role of urban proximity in

[31] We test for the equality of the reduced form estimates using seemingly unrelated regressions. We find that the estimate for the first quartile is statistically different from two of the coefficients at the five percent level and the third (quartile 3) at the ten percent level.

[32] These results are consistent with cross-sectional evidence that in villages close to urban centers, there exist a wider range of industries (Fafchamps and Shilpi, 2005) and workers are more likely to be in the nonfarm sector and work for wages (Deichmann et al., 2009).

[33] For a richer description of this project and estimates of its impact on firm location and productivity, see Ghani et al. (2015).

[34] We use the reduced form rather than RDD specification here in order to be able to interact the road priority dummy below-median distance to the Golden Quadrilateral and large cities, as it is these interactions that are interested in testing here.

driving our results. Column 4 estimates the model with interactions of proximity to the GQ and proximity to cities. This kills the effect of the GQ: while treatment effects are significantly larger in locations close to cities, we find no evidence that highways are the perfect substitutes for urban proximity that was suggested by the specification in Column 2.

Recent evidence has demonstrated that a reduction in transportation costs can lead to significant increases in outmigration from rural areas (Bryan et al., 2014; Morten and Oliveira, 2014). Although we are not able to measure migration choices directly, we examine the closest proxy in our data: village-level population growth. Table A6 presents the impact of rural road construction on total annualized village population growth between the 2001 and 2011 Censuses. We find no evidence of significant outmigration in response to road construction. Given the lack of large population growth effects, we interpret our findings of sectoral reallocation as the result of changes in occupational choice and not compositional effects due to selective migration.

Theory suggests that those who exit agricultural cultivation in favor of labor market opportunities will be those for whom the losses of agricultural income are smallest and labor market gains are largest. By using household level census data, we are able to examine the distribution of treatment effects across subgroups with different factor endowments. As the dominant sector of the rural Indian economy is agriculture, land endowments may play a major role in determining which households respond most to a rural road. We first establish in Table A7 that the probability of owning land does not respond to rural road construction, nor does the share of households with landholdings of various sizes. We take this as prima facie evidence that the landholding distribution does not respond to our treatment, and thus ex post observed landholdings can be treated as a baseline variable upon which to conduct heterogeneity analysis. Table 7 presents our main specification, estimating the effect separately by size of landholdings. We find that movement out of cultivation is strongest in households with small landholdings ($\leq$ 1 acre), precisely those that stand to gain the most from increased labor force participation.[35] This decrease in cultivation for those with small landholdings (13 percentage points) is much stronger as a percentage of the control group mean: our estimates suggest that 27% of households with less than one acre of land exit cultivation, compared to just 7% in households with greater than one acre of land.[36] These results are consistent with recent work finding that the inheritance of land in India can significantly reduce rates of migration and participation in non-agricultural occupations (Fernando, 2014) and suggest that the lack of paved roads may be one cause of the inefficiently small size of many farms in rural India (Foster and Rosenzweig, 2011).[37]

We next examine the heterogeneity of the treatment effect as a function of age and gender, finding that roads have a significantly larger effect on male agricultural employment than female. As age and gender are individual characteristics, we use the same outcome as in Table 5: the share of employed population working in agriculture, defined as any occupation listing agriculture or farming in its description. While we find no differences in the impact of road construction by age of workers, we do find that the entire effect is driven by males.[38] Due to the large standard errors on our estimates, the differences are not statistically significantly different. These estimates could be the result of a male physical advantage in non-agricultural work or attitudes against women's working far away from

[35] We can reject equality between the estimates for households with below one acre compared to those with greater than one acre; a test for equality of the reduced form estimates yields a p-value of .02.

[36] It is important to note that productivity in cultivation will only depend on landholdings if there are market failures such that it is more productive to work on one's own land. An extensive literature investigates common failures in agricultural land and labor markets in low income countries. See, for example, de Janvry et al. (1991).

[37] These effects suggest that road construction may be a progressive investment in that those with the least agricultural wealth (as proxied by landholding) show the largest labor market effects. We also estimate effects separately by caste group (Table A8). There is considerable evidence of discrimination against low caste groups in rural labor markets (Ito, 2009). We find that road construction lowers the share of scheduled caste households in agriculture by 16.6 percentage points, as compared to 7.6 for general caste households. As a share of the control group average, road construction induces 63 percent of low caste households to leave agriculture, as opposed to 16 percent of high caste households.

[38] The lack of differential effect by age is in contrast to the findings of Kim and Topel (1995), who document that the rapid movement of the South Korean labor force out of agriculture occurred with very little reallocation at the individual level, instead being driven by new entrants to the labor force.

home that may prevent reallocation of female labor away from agriculture (Goldin, 1995). It is beyond the scope of this paper or abilities of our data to investigate the cause for this differential effect by gender, but this finding suggests this would be a fruitful topic for future research.[39]

Finally, we use asset and income data to examine whether the labor market impacts of rural road construction is associated with improvements in economic outcomes. Table 9 presents estimates of PMGSY road construction on earnings and assets using our main regression discontinuity specification. The SECC categorizes the monthly income of the highest earning member of the household into three groups: less than 5,000 rupees, between 5,000 and 10,000 rupees, and more than 10,000 rupees per month.[40] Using data from the National Sample Survey, we generate expected earnings based on the group means.[41] We find that road construction leads to an increase of 326 rupees in monthly earnings, significant at the 10% level, representing an 8% increase over the control group. This change is the result of an estimated 2.6 percentage point decrease in the share of households in the lowest income category (not significant) and 1.5 percentage point increase in the share of households in the highest category. We mostly find no significant changes to the assets measured in our data, except for a 5.5 percentage point increase in the share of households living in a home with a solid roof and walls.[42]

[39]One possibility is that women actually increase their agriculture activity as the men spend more time away from the farm and the village. A proper exploration of this would likely utilize richer data than we have, including time use data that would allow for examination of the intensive margin of labor reallocation.

[40]The average exchange rate in 2012, the year in which most of the SECC was collected, was 53.5 Indian rupees per US dollar. At this rate, 5,000 rupees is the equivalent of $93.50 and 10,000 rupees the equivalent of approximately $187.

[41]We calculate these means to best match our sample and the definition of our earnings variables in the SECC. We restrict ourselves to the highest earning member of each household, in villages that did not have a paved road at baseline. We calculate mean earnings of 3,076 rupees per month for those in the lowest income category, 6,373 rupees in the middle category and 22,357 in the highest category.

[42]These measures of economic outcomes may underestimate the welfare effects of increased nonfarm labor market participation if wages serve as insurance against agricultural risk, as demonstrated by Kochar (1999). It is also possible that rural road construction lowers the volatility of agricultural income, as demonstrated Allen and Atkin (2015) with respect to Indian highway construction. Another way that we may fail to accurately estimate the impact on welfare is due to our inability to measure consumption, which many researchers have argued is a better proxy for welfare than income. See Meyer and Sullivan (2003) for a

Finally, we consider whether the impacts of road construction are due to a level change in economic activity and outcomes, or a change in the growth rate. Most research on transport infrastructure has not been able to differentiate between these hypotheses. Our primary identification strategy is likewise ill-suited to differentiate between level and growth effects, as we have exogenous variation in the extensive margin (road construction by 2012) rather than the intensive margin (years since road construction). Evidence from historical railroad construction in China points towards there being a level but not growth effect (Banerjee et al., 2012). In order to test for an impact on the growth rate, we turn to nighttime luminosity from satellite images. Since Henderson et al. (2011) demonstrated the robust relationship between the annual average of luminosity of lights at night and GDP, night lights have become an increasingly common way of measuring total economic activity in settings where other data are not available at such a high spatial or temporal resolution. Unlike our other data, night lights have observations for every year, allowing us to estimate a panel specification with both year and village fixed effects. Our hypothesis is that villages grew significantly faster after road construction. We thus create a binary treatment variable that takes on the value one for any year after road construction. We limit our sample to 9 observations for every village that received a paved road before 2010: 4 years preceding road completion, the year of road completion, and four years after.[43] Table 10 shows these results of this estimation. We find that luminosity grew 2.5 log points faster after road construction as compared to before (Column 1). Figure 6 plots the coefficients in Column 2, which estimates the effect on luminosity of each year relative to the year of road completion. There is a visible increase in the growth rate in the year following road construction, an effect that persists over time. This finding provides a potential explanation for the marginally significant impacts on earnings and assets that we observe in Table 9. The median number

discussion of the trade-offs between these different measures.

[43]Our night lights data run through 2013. We thus only consider villages that received roads 2009 and earlier to ensure that we have four post-treatment observations per village.

of years of treatment among villages in our sample that received a road before our 2012 data is four years (mean = 4.03). If the benefits to road construction accumulate over time, our short window between road construction and the earnings and asset data in the SECC may not be enough time for large changes in economic outcomes to accumulate.

6.2 Robustness

In this section we explore the possibility that factors other than PMGSY road construction may be driving our results. Reassuringly, we find no evidence supporting such concerns.

As a placebo exercise, we run our first stage and reduced form estimation on these outcomes for the set of villages not in our main sample, where there is not discontinuous increase in road construction at the population threshold. If other determinants of sectoral allocation varied discontinuously at the treatment threshold, we might incorrectly attribute their effects to rural road treatment. Table A3 presents the estimates of these regressions. There is no evidence of either a first stage or reduced form effect on cultivation or manual labor shares for the placebo sample, indicating that our results are not due to some other factor whose effect we spuriously attribute to PMGSY roads.

A different threat to our identification could come from any other policy that used the same thresholds as the PMGSY. In fact, one national government program did prioritize villages above 1000 population: the Total Sanitation Campaign (Spears, 2015). This program sought to incentivize rural local governments to improve sanitation by eliminating open defecation. We present three reasons why it is highly unlikely that this program is spuriously driving our results. First, there is little theoretical reason to believe that a reduction in open defecation could produce a large reallocation of labor from cultivation to wage labor market participation. Second, our estimated results are statistically indistinguishable for the villages around the 500 population cutoff as compared to those near the 1000 cutoff (see Table A1 for estimates); the Total Sanitation Campaign did not use this lower threshold. Third, in

Table A5 we present reduced form estimates of the impact of road priority on four measures of sanitation, including the share of households practicing open defecation. We find no evidence that being above the 1000 population threshold is associated with improved outcomes in any of these measures.

We consider yet another concern to the identification of the effect of rural roads on economic activity: a treatment effect on data quality. It is conceivable that road connectivity could facilitate changes in the quality of data collection, and even introduce bias if lower transport costs changes attrition. We test for this possibility by constructing a measure of data quality. Using variables that can be constructed in both the 2011 Population Census and 2012 SECC, we estimate the impact of road construction on the differences between measures from these two datasets. We consider the three variables that are measured equivalently in the two datasets: total population, population under six years of age, and the share of households owning any phone.[44] Table A2 presents these results. We find no significant differences across any of these three measures.

6.3 Mechanism

We have thus far established the causal impact of rural roads on occupation choice. The evidence clearly demonstrates that rural road construction leads to a large reallocation of labor out of cultivation and into manual labor, driven by locations with greater market access, and individuals and households who are likely to have the highest returns to labor market participation. We now test between the potential mechanisms (discussed in Section 2) by which lower transport costs may lead to such sectoral reallocation out of agriculture: (i) an increase in demand for labor from the within-village nonfarm sector, (ii) a reduction in demand for agricultural labor, and (iii) increase in demand for labor from external labor

[44] Both datasets contain information on other assets and occupation, but the definitions used are different and thus these variables are not appropriate measures of the quality of the SECC.

markets. While we cannot definitively rule out that any of these processes are at work, we argue that the evidence that follows points most strongly in the direction of participation in external labor markets.

We do not find evidence that the observed occupational reallocation is due to growth of in-village nonfarm firms. Our data allow us to estimate the growth of such activity in three different ways, the results of which we present in Panel A of Table 11. Our first two tests for the growth of the nonfarm sector comes from regression discontinuity estimates of the impact of road construction on two measures of business ownership in the SECC: the share of households owning or operating a registered enterprise, and the share of households reporting a business as their primary source of income. We find no evidence of an increase in either measure; in fact, we observe a 1.2 percentage point decrease in the share of households deriving their primary income from a business, a 21% decrease from the control group average. Our third test of this hypothesis is to estimate the OLS effect of PMGSY road construction on nonfarm employment in firms in the 2005 Economic Census.[45] For consistency with other outcomes, we define the outcome to be the share of the population employed in non-farm firms within the village. Our treatment group is all villages that received a PMGSY road before 2005. To maximize comparability, we define our control group to be those villages the received a PMGSY road between 2005 and 2009. Our estimated impact allows us to reject a 0.2 percentage point increase in the share of the population employed in non-farm firms within the villages. While these estimates are subject to concerns of selection bias, most theories of the endogeneity of road construction (e.g. political favoritism or economic potential) would predict that OLS provides upward and not downward bias in the estimated treatment effect. These three results together are strongly suggestive evidence that growth in the within-village non-farm sector is not responsible for the observed sectoral reallocation

[45] As fewer than one sixth of the roads in our sample were constructed before 2005, and the population threshold rules do not seem to have been followed, we are unable to run our RD analysis for this estimation.

out of agriculture.

We also fail to find evidence consistent with a reduction in demand for agricultural labor due to labor saving investments. Much of the existing literature on rural roads focuses on agricultural outcomes, finding evidence that connectivity results in increased agricultural land values (Jacoby, 2000), increased productivity (Sotelo, 2015) and lower market prices for agricultural output (Casaburi et al., 2013). Our primary findings – that rural roads lead to a reduction in the share of households and individuals deriving their income from agriculture – appear to be at odds with this literature. However, productivity increases (which we are unable to test for) could lead to a reduction in agricultural workers if road construction triggers labor saving investments in agriculture. For example, Bustos et al. (2015) find that technical change in soy production in Brazil was strongly labor saving, leading to a reduction in the agricultural share of the workforce in soy growing areas. This story has the strong prediction that we should observe potentially labor-saving changes in agricultural production. Our data allow us to test for three such outcomes: ownership of mechanized farm equipment (tractors, etc), ownership of irrigation equipment, and consolidation of landholdings. Panel B of Table 11 presents regression discontinuity estimates of the impact of road construction on these outcomes. We find no evidence for increases in ownership of agricultural capital, nor for a decrease in land ownership that would suggest a consolidation of landholdings.[46] We further explore the landholding distribution in Table A7, again finding no evidence of significant changes to the landholding distribution. While we cannot observe other agricultural investments such as input use, these results strongly suggest that major labor saving investments in agriculture are not driving the impact of road construction on the sectoral allocation of labor.[47]

[46]In fact, for large landholders, we observe significant negative decreases in ownership of agricultural equipment. See Table A11 for full results.

[47]Although all of our results must be interpreted as local average treatment effects and we should be cautious in extrapolating to other samples, these results seems especially sensitive to concerns of external validity. Capital-intensive agriculture is practiced in many parts of India, and it is hard to imagine this would be possible without sufficiently low transport costs that farmers can export their surpluses beyond

Having eliminated both within-village nonfarm growth and labor-saving agricultural investments as the mechanisms by which roads cause the reallocation of labor out of agriculture, it appears likely that increased access to external labor markets is driving our results. We provide three further pieces of evidence that point in this direction. The first comes from the heterogeneity of treatment effect by size of the urban-rural wage gap (mean urban wages minus mean rural wages). We use data from the National Sample Survey (68th Round) to estimate the gap in average earnings of low skill workers (those with below primary education), following Munshi and Rosenzweig (2016). To reduce noise, we estimate these gaps at the division level.[48] Theoretically, we should expect that if roads are enabling workers to gain employment in nearby urban areas, we should expect to see the largest exit from agriculture in areas with the largest urban-rural wage gaps. In Panel C of Table 11, we estimate the impact of road construction on the share of households in agriculture, dividing the sample by median urban-rural wage gap (39.08 rupees). We find that treatment effects are close to zero in the low wage gap sample, while in the high wage gap sample, a road leads to a 23.5 percentage point reduction in the share of households in agriculture. A seemingly unrelated regression test of the reduced form coefficients shows that these two estimates are statistically different from each other, with a p-value of 0.04.

Our second piece of evidence comes from the National Sample Survey (2011-12) question on location of work: in villages that did not have a paved road at baseline, workers living in rural areas are 43% more likely to report working in an urban area when their village has received a road by 2011.[49]

Finally, a road alone does not allow rural inhabitants to access urban labor markets:

the village. In fact the region best known for such agriculture, Punjab, is not in our sample because nearly all villages already had roads at the start of the PMGSY, thus excluding them from our analysis.

[48]A division is an administrative unit that is larger than a district but smaller than a state. The median division has four districts in it.

[49]See Table A12 for these results broken out by distance to a major city. Given the far smaller sample of the NSS, we do not have the power to generate regression discontinuity estimates using these data.

transportation must be available. We should thus expect to see an increase in transportation services in precisely those areas where workers are exiting agriculture at the highest rates. While there is no dataset containing high frequency data on traffic in rural areas across India, we do have data in the 2011 Population Census on whether villages are served by transport. Using our main RDD specification, we estimate the impact of road construction on the availability of scheduled bus services at the village level. Table A9 presents these results, both for the full sample and for each quartile of distance to a major city of 500,000+ population. The results show that for the quartile of villages closest to a large city, a rural road increases the probability of bus service by 32 percentage points. Point estimates for the other quartiles are statistically indistinguishable from zero. Assuming that bus services only operate where it is economically feasible to do so, we interpret this to say that rural roads only generate sufficient demand for bus service when they connect rural areas to nearby urban markets. This helps to shed light on recent research by Raballand et al. (2011), whose experiment on bus subsidies demonstrated that in rural Malawi, there was no price at which there was sufficient demand to make bus routes profitable. Our findings lend credence to the possibility that low population densities and incomes in many rural areas may limit the profitability of transport services, and by extension the returns to road construction, but also suggest that this is not the case for rural areas that are sufficiently close to major urban markets.

7 Conclusion

Access to the outside world via paved roads, easily taken for granted in many rich countries, is far from a reality for many of the world's rural poor. High transportation costs potentially inhibit gains from the division of labor, economies of scale and specialization. Recent work has begun to demonstrate the role of trunk infrastructure (railroads and highways) on economic activity. However, despite the emphasis of both theorists and development policy-makers on the importance of transportation costs, little is known of the economic effects of road provision on rural economic activity and outcomes.

In this paper we estimate the economic impacts of the Pradhan Mantri Gram Sadak Yojana, a large-scale program in India with the objective to provide universal access to paved "all-weather" roads in rural India. To do so, we assemble microdata describing every rural inhabitant and household in the country, joining a growing body of economic research utilizing government administrative datasets. We exploit discontinuities in the probability of road construction at village population thresholds to estimate the impact of this program. We find that road construction leads to a large reallocation of labor out of agriculture and into (manual) labor markets. The results are strongest in locations close to large cities, where we expect commuting and short-term migration to be most profitable. Rather than facilitating growth of the nonfarm sector in rural areas, road construction appears to facilitate the access of rural labor to external employment. These labor market outcomes are associated with a nearly ten percent increase in earnings.

The question of why so many workers remain in low productivity agriculture when higher wages are available in other locations and sectors is a classic one in development economics. Our findings suggest that the poor state of rural transportation infrastructure in developing countries must be taken seriously as a barrier to the efficient allocation of labor across space and sectors. This should not be surprising: arbitrage is only possible if the costs of reallocation are less than the gains. We do not claim to resolve the entire puzzle. Migration is an obvious alternate way of accessing labor market opportunities outside of the village. It is beyond the scope of this paper to examine the many potential barriers to migration, but recent research has suggested that transportation costs are an important factor (Morten and Oliveira, 2014; Bryan et al., 2014). We find no such evidence of a rise in migration following road construction, lending credence to research proposing factors other than the state of rural transport infrastructure to explain India's low rates of rural-urban migration (see, for example, Munshi and Rosenzweig (2016)).

Foster and Rosenzweig (2007) assert that economists do not adequately understand the flows of capital and labor between rural and urban areas in developing countries. This paper adds to a growing literature on the linkages in labor markets across space, and suggests that transportation infrastructure may be an important determinant of such flows. However, a limitation of this paper is that we cannot study these flows directly. We hope that future research will shed light on the nature and causes of labor flows across sectors and between rural and urban areas. There is increasing evidence that rural workers are an important component of urban labor supply (Imbert and Papp, 2015). If so, the impacts of improving transportation linkages between rural and urban areas will also be felt by urban inhabitants, with potentially large consequences for both urban wages and firm behavior.

Many researchers have puzzled over India's low rates of urbanization and structural transformation when compared to other developing countries. This paper provides evidence that workers can participate in non-agricultural labor markets without moving to cities when market access to urban areas is sufficiently high. India's high population densities and superior infrastructure may help to explain why its structural transformation has approximately matched the speed of sub-Saharan Africa while urbanizing much more slowly (see Figure A3). At the same time, our results suggest that India's low rate of structural transformation when compared to China may be due in part to its much lower rate of investment in transportation infrastructure.[50] More research is needed to understand the policies that have enabled the structural transformation away from low productivity agriculture in certain low-income countries and not in others.

[50] Over the period 1992 to 2011, China spent 8.2% of GDP on infrastructure compared to India's 4.7% (Dobbs et al., 2013). For roads specifically, China spent 3.4% of GDP in 2003, compared to just 0.4% for India in 2004 (authors' calculations).

Table 1
Summary statistics

	No Road	Paved Road	Total
Primary school	0.797	0.912	0.864
	(0.403)	(0.283)	(0.343)
Medical center	0.226	0.480	0.373
	(0.418)	(0.500)	(0.484)
Electrified	0.329	0.618	0.497
	(0.470)	(0.486)	(0.500)
Distance from town	26.03	20.36	22.74
	(24.52)	(18.95)	(21.65)
Land irrigated share	0.359	0.441	0.406
	(0.351)	(0.372)	(0.365)
Ln land area	4.873	5.451	5.208
	(1.080)	(1.147)	(1.155)
Illiterate share	0.562	0.498	0.525
	(0.165)	(0.142)	(0.155)
Ag emp share	0.825	0.744	0.778
	(0.218)	(0.233)	(0.230)
SC share	0.166	0.184	0.176
	(0.205)	(0.183)	(0.193)
Population (2001)	864.0	1734.3	1368.3
	(976.6)	(1876.5)	(1620.6)
Population (2011)	1031.0	1996.2	1590.4
	(1194.2)	(2190.3)	(1899.2)
Employment in firms (1998)	49.43	94.23	75.39
	(131.6)	(314.9)	(255.4)
Number of firms (1998)	24.39	42.19	34.71
	(53.22)	(84.60)	(73.59)
PMGSY road by 2012	0.247	0.209	0.225
	(0.431)	(0.406)	(0.418)
Observations	135568	186839	322407

Notes: This table presents means and standard deviations of baseline variables and outcomes. The first column presents summary statistics for villages without a paved road in the 2001 Population Census, the second column for villages with a paved road, and the third column for the pooled sample.

Table 2

Balance

Variable	Below threshold	Over threshold	Difference of means	t-stat on difference	RD estimate	t-stat on RD estimate
Primary school	0.89	0.84	0.06	4.49	-0.02	-0.28
Medical center	0.29	0.22	0.07	12.49	-0.03	-0.37
Electrified	0.44	0.39	0.04	3.60	-0.03	-0.39
Distance from town	22.18	23.72	-1.53	-5.15	-4.13	-1.06
Land irrigated share	0.41	0.39	0.03	3.80	-0.02	-0.48
Ln land area	4.87	4.63	0.24	6.14	0.21	1.06
Illiterate share	0.53	0.54	-0.01	-1.85	0.00	0.12
Ag emp share	0.79	0.80	-0.01	-2.27	-0.01	-0.33
SC share	0.18	0.18	0.01	3.95	0.04	1.05
N	10170	9442				

Notes: The table presents mean values for village characteristics, measured in the baseline period. The baseline period is 2001 for all variables. Columns 1 and 2 show the unconditional means for villages below and above the treatment threshold, respectively. Column 3 shows the difference of means across columns 1 and 2 and column 4 shows the t statistic for the difference of means. Column 5 shows the regression discontinuity estimate of the effect of cutoff on the baseline variable (with the outcome variable omitted from the set of controls), and column 6 is the t statistic for this last estimate, using heteroskedasticity robust standard errors. A optimal bandwidth of ± 85 around the population thresholds has been used to define the sample of villages (see text for details), such that the sample for the estimation are villages with a population in the range of 415-585 for the 500 threshold and 915-1085 for the 1000 threshold.

Table 3

First stage effect of road priority on PMGSY road treatment

	±50	±60	±70	±80	±90	±100
Road priority	0.137	0.134	0.131	0.130	0.129	0.130
	(0.018)***	(0.016)***	(0.015)***	(0.014)***	(0.013)***	(0.013)***
F Statistic	58.38	67.2	74.81	85.03	95.1	107.3
N	8840	10484	12250	13979	15762	17469
R2	.2592	.2527	.2492	.247	.2455	.2447

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

Notes: This table presents first stage estimates from Equation 3 of the effect of PMGSY prioritization on a village's probability of treatment. The dependent variable is a dummy variable that takes on the value one if a village has received a PMGSY road before 2012. The first column presents results for villages with populations within 50 of the population threshold (450-550 for the low threshold and 950-1050 for the high threshold). The second through sixth columns expand the sample to include villages within 60, 70, 80, 90 and 100 of the population thresholds. The sample consists of villages that did not have a paved road at baseline (see text for details). The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table 4

Impact of road on source of income

Panel A. Cultivation as Main Source of Income

	±50	±60	±70	±80	±90	±100
Road	-0.099	-0.111	-0.106	-0.100	-0.092	-0.084
	(0.0506)**	(0.0468)**	(0.0434)**	(0.0402)**	(0.0376)**	(0.0353)**
Outcome Mean	0.4280	0.4279	0.4276	0.4283	0.4283	0.4295
N	11506	13683	15968	18245	20533	22772
R2	0.4821	0.4733	0.4726	0.4733	0.4753	0.4771

Panel B. Manual Labor as Main Source of Income

	±50	±60	±70	±80	±90	±100
Road	0.1103	0.1230	0.1206	0.1130	0.1054	0.0993
	(0.0497)**	(0.0461)***	(0.0428)***	(0.0397)***	(0.0372)***	(0.0349)***
Outcome Mean	0.5098	0.5098	0.5101	0.5096	0.5095	0.5085
N	11506	13683	15968	18245	20533	22772
R2	0.4460	0.4347	0.4324	0.4340	20533	22772

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

Notes: This table presents regression discontinuity estimates from Equation 3 of the effect of road construction on the primary source of income. Panel A presents regression discontinuity estimates for the share of households reporting cultivation as their primary source of income while Panel B presents regression discontinuity estimates for the share of households reporting manual labor as their primary source of income. The first column presents results for villages with populations within 50 of the population threshold (450-550 for the low threshold and 950-1050 for the high threshold). The second through sixth columns expand the sample to include villages within 60, 70, 80, 90 and 100 of the population thresholds. For each regression, the outcome mean for the control group (villages with population below the threshold) is also shown. The sample consists of villages that did not have a paved road at baseline (see text for details). The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table 5
Impact of road on occupation in agriculture

	Household Income Source		Occupation	
	Cultivation	Manual Labor	Agriculture	Manual Labor
Road	-0.096	0.109	-0.093	0.084
	(0.039)**	(0.038)***	(0.047)**	(0.046)*
Outcome Mean	.4286	.5093	.4505	.4439
N	19612	19612	19525	19525
R2	.4743	.435	.3032	.2811

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

Notes: This table presents regression discontinuity estimates from Equation 3 of the effect of road construction on occupational choice. Columns 1 and 2 provide estimates of the impact of a rural road on the share of households reporting cultivation and manual labor as the primary source of income. Column 3 estimates the impact on the share of employed working age population (21-60) working in agriculture, defined as any occupation listing agriculture or farming in its description. Column 4 estimates the effect on the share of employed working age population working in manual labor (excluding agriculture). The sample consists of villages that did not have a paved road at baseline, with baseline population within the optimal bandwidth (85) of the threshold (see text for details). For each regression, the outcome mean for the control group (villages with population below the threshold) is also shown. The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table 6

Heterogeneity of impact by proximity to cities

	100k near	100k far	500k near	500k far
Road	-0.146	-0.066	-0.206	-0.023
	(0.059)**	(0.054)	(0.067)***	(0.048)
Outcome Mean	.4356	.4216	.4268	.4303
N	9806	9806	9806	9806
R2	.4412	.5159	.3988	.5108

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

Notes: This table presents regression discontinuity estimates from Equation 3 of the effect of road construction on cultivation as the primary source of income (share of households). Columns 1 and 2 estimate the treatment effect for villages below and above median (51.5 km) straight line distance to cities of 100,000+ population at baseline. Columns 3 and 4 estimate the treatment effect for villages below and above median (100.8 km) straight line distance to cities of 500,000+ population at baseline. The sample consists of villages that did not have a paved road at baseline, with baseline population within the optimal bandwidth (85) of the threshold (see text for details). The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table 7

Heterogeneity of impact by size of landholdings

	Landless	0-1 Acres	1+ Acres
Road	-0.024	-0.130	-0.047
	(0.029)	(0.058)**	(0.039)
Outcome Mean	.1218	.4867	.7213
N	19383	17203	19137
R2	.1868	.1553	.2855

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

Notes: This table presents regression discontinuity estimates from Equation 3 of the effect of road construction on cultivation as the primary source of income (share of households). The first column reports results for households reporting no agricultural land, the second column for households owning less than one acre of land, and the third column for households with one or more acre of land. The sample consists of villages that did not have a paved road at baseline, with baseline population within the optimal bandwidth (85) of the threshold (see text for details). For each regression, the outcome mean for the control group (villages with population below the threshold) is also shown. The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table 8
Heterogeneity by age and gender

	All		Male		Female	
	21-40	41-60	21-40	41-60	21-40	41-60
Road	-0.100	-0.097	-0.110	-0.117	-0.044	0.024
	(0.049)**	(0.050)*	(0.049)**	(0.050)**	(0.060)	(0.067)
Outcome Mean	.4088	.5403	.4258	.5693	.2673	.3154
N	19512	19438	19494	19426	18098	17041
R2	.2894	.3077	.2893	.3103	.228	.2452

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

Notes: This table presents regression discontinuity estimates from Equation 3 of the effect of road construction on occupational choice, examining the heterogeneity of effects by age and gender. The outcome is the share of employed population working in agriculture, defined as any occupation listing agriculture or farming in its description. The first two columns estimate the effect for both genders, columns 3 and 4 for the male employed working age population and columns 5 and 6 for the female employed working age population, considering results separately for younger (21-40) and older (41-60) workers. For each regression, the outcome mean for the control group (villages with population below the threshold) is also shown. The sample consists of villages that did not have a paved road at baseline, with baseline population within the optimal bandwidth (85) of the threshold (see text for details). The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table 9

Impact of road on household earnings and assets

	Income			Assets			
	Mean	< 5k	≥ 10k	Solid House	Refrigerator	Vehicle	Phone
Road	327.341	-0.027	0.015	0.054	0.016	-0.021	-0.043
	(194.806)*	(0.024)	(0.008)*	(0.032)*	(0.013)	(0.024)	(0.040)
Outcome mean	4073	.8711	.03579	.2724	.03344	.1421	.5111
Fixed effects	Dist x Cutoff	Dist x Cutoff	Dist x Cutoff	Dist x Cutoff	Dist x Cutoff	Dist x Cutoff	Dist x Cutoff
N	19792	19792	19792	19792	19792	19792	19792
R2	.2883	.2864	.2498	.7152	.1617	.3434	.6211

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

Notes: This table presents regression discontinuity estimates from Equation 3 of the effect of road construction on measures of earnings and assets. Columns 1, 2 and 3 present regression discontinuity estimates for three different measures of income: imputed mean earnings, share of households whose highest earning member earns less than 5,000 rupees per month and more than 10,000 rupees per month. Imputed mean earnings are based on assigning monthly earnings of 3,076 rupees to households whose highest earning member reports income of less than 5,000 rupees, 6,373 rupess to households reporting greater than 5,000 but less than 10,000 rupees and 22,353 rupees to households reporting greater than 10,000 rupees. These numbers are mean monthly earnings for earners in these wage ranges using data from the 68th Round (2011-12) of the National Sample Survey. Columns 3 through 7 present estimates for the impact of road construction on the share of households owning the following assets: a house of solid material (having both solid walls and roof), a refrigerator, any motorized vehicle, and any phone. The sample consists of villages that did not have a paved road at baseline, with baseline population within an optimal bandwidth (85) of the threshold (see text for details). For each regression, the outcome mean for the control group (villages with population below the threshold) is also shown. The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table 10

Impact of road on night lights

	(1)	(2)
Road	0.025	
	(0.003)***	
Year - Road Comp Year = -3		0.015
		(0.004)***
Year - Road Comp Year = -2		0.021
		(0.004)***
Year - Road Comp Year = -1		0.033
		(0.004)***
Year - Road Comp Year = 0		0.033
		(0.005)***
Year - Road Comp Year = 1		0.073
		(0.005)***
Year - Road Comp Year = 2		0.072
		(0.006)***
Year - Road Comp Year = 3		0.066
		(0.007)***
Year - Road Comp Year = 4		0.087
		(0.007)***
Fixed Effects	Year, Village	Year, Village
N	289611	289611
R2	.1969	.1971

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

Notes: This table presents panel estimates of the impact of road construction on village economic activity as proxied by nighttime luminosity. The sample is all villages that did not have a paved road at program baseline and received one before 2010, ensuring that we have a full 9 year panel for each village that includes four years of observations before road construction and four years after. The outcome variable is the difference in log in luminosity (growth rate). Column 1 estimates the effect of road treatment (a dummy for whether the observation is after the year of road completion) on the log difference (growth rate) of luminosity. Column 2 estimates the effect by year relative to road treatment, where 0 is the year of road completion and year -4 is the omitted variable. All regressions include year and village fixed effects. Heteroskedasticity robust standard errors are reported below point estimates, with clustering at the village level.

Table 11

Evidence on mechanism

Panel A. In-village economic activity

	Enterprise Ownership	Enterprise Income	EC05 Emp Share
Road	-0.001	-0.012	0.0004
	(0.0101)	(0.0072)*	(0.0008)
Outcome Mean	0.0101	0.0056	0.4065
N	19612	19612	13281
R2	0.0512	0.0540	0.2795

Panel B. Agricultural investments

	Mech Farm Equip	Irr Equip	Land Ownership
Road	-0.023	-0.038	0.0364
	(0.0147)	(0.0291)	(0.0393)
Outcome Mean	0.0414	0.1388	0.5831
N	19612	19612	19612
R2	0.2225	0.4190	0.3478

Panel C. Returns to participation in urban labor markets

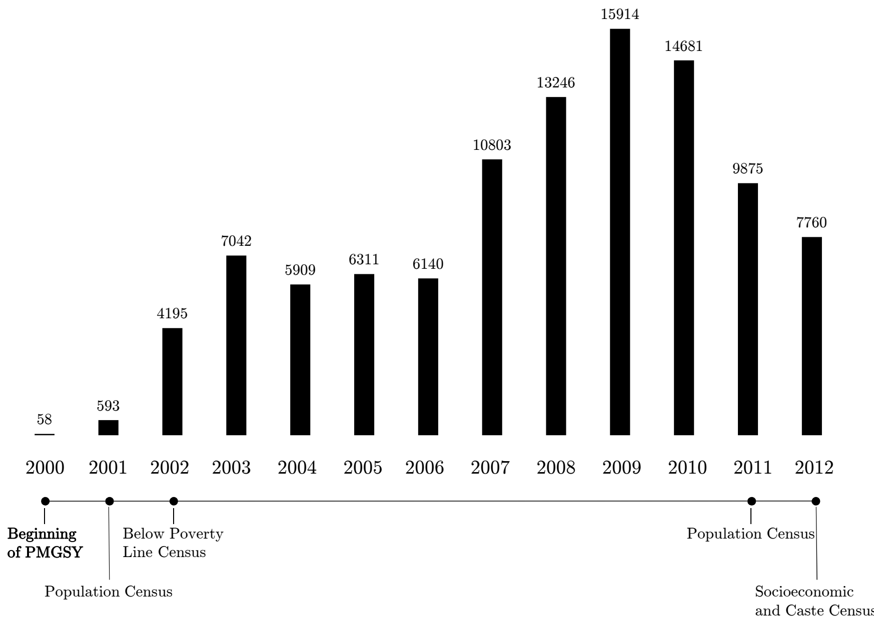
	Wage Gap Low	Wage Gap High
Road	-0.002	-0.235
	(0.0491)	(0.0899)***
Outcome Mean	0.4359	0.4207
N	8948	7894
R2	0.5017	0.3865

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

Notes: This table presents evidence on the mechanism by which road construction affects the sectoral allocation of economic activity. Panel A tests for growth in the within-village non-farm sector. Columns 1 and 2 provide regression discontinuity estimates of the impact of road construction on the share of households owning or operating an enterprise that is registered with the government and the share of households reporting a business as their primary source of income. Column 3 reports the OLS estimate of road construction on the share of village population working in non-farm establishments within the village in 2005, defined as the non-farm employment in the 2005 Economic Census divided by total population in 2001. For column 3, the sample is restricted to villages that did not have a paved road at baseline and received a paved road by 2006, with treatment defined as a dummy if a village received a road before 2005. Panel B tests for evidence that sectoral reallocation could be the result reduced demand for agricultural labor. The three columns present regression discontinuity estimates of the impact of road construction on the share of households owning mechanized farm equipment, irrigation equipment and agricultural land. Panel C tests for heterogeneity of effects by returns to participation in urban labor markets. District level rural-urban wage gaps were calculated using the 68th Round (2011-12) of the National Sample Survey, Employment and Unemployment surveys. Following Munshi and Rosenzweig (2016), the wage gap was calculated for individuals who have completed less than primary school to best estimate differential returns for the group most likely to be reallocating from low-skilled agricultural labor to manual labor. Wage gap high denotes villages in divisions with a daily wage gap higher than 39.86 Rs, the median daily wage gap for the sample of interest (see text for details). All specifications include baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste). Regression discontinuity estimates include district-cutoff fixed effects. Panel A, Column 3 includes a baseline control for 1998 non-farm village employment as a share of 2001 population, quadratic baseline population controls and district fixed effects. Heteroskedasticity robust standard errors are reported below point estimates. For each regression, the outcome mean for the control group is also shown.

Figure 1

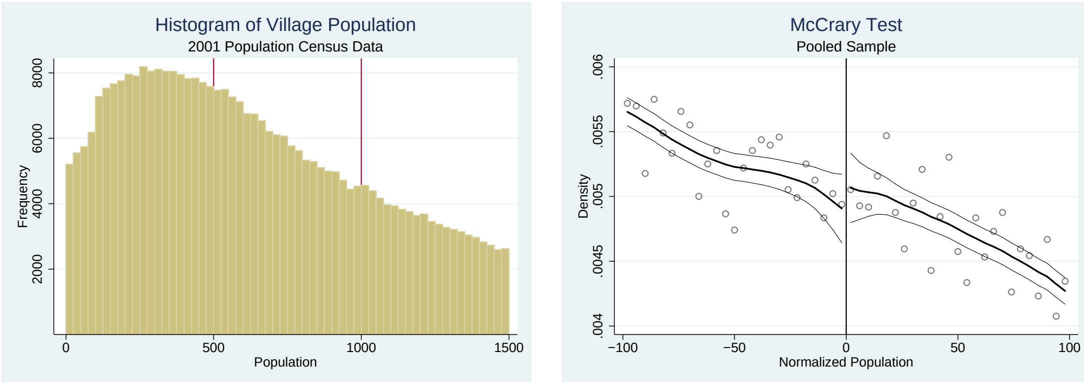
Timeline of Data Sources, with Count of Villages Treated



Notes: The figure shows when the population and poverty censuses of India used as primary data sources in this paper were conducted. Note that while the Socioeconomic and Caste Census (SECC) was intended to be conducted exclusively in 2011, and it is often referred to with this year, it was conducted primarily in 2012. The bar graph above represents the number of villages receiving PMGSY roads in each year in our full village-level dataset. Exact counts are also listed.

Figure 2

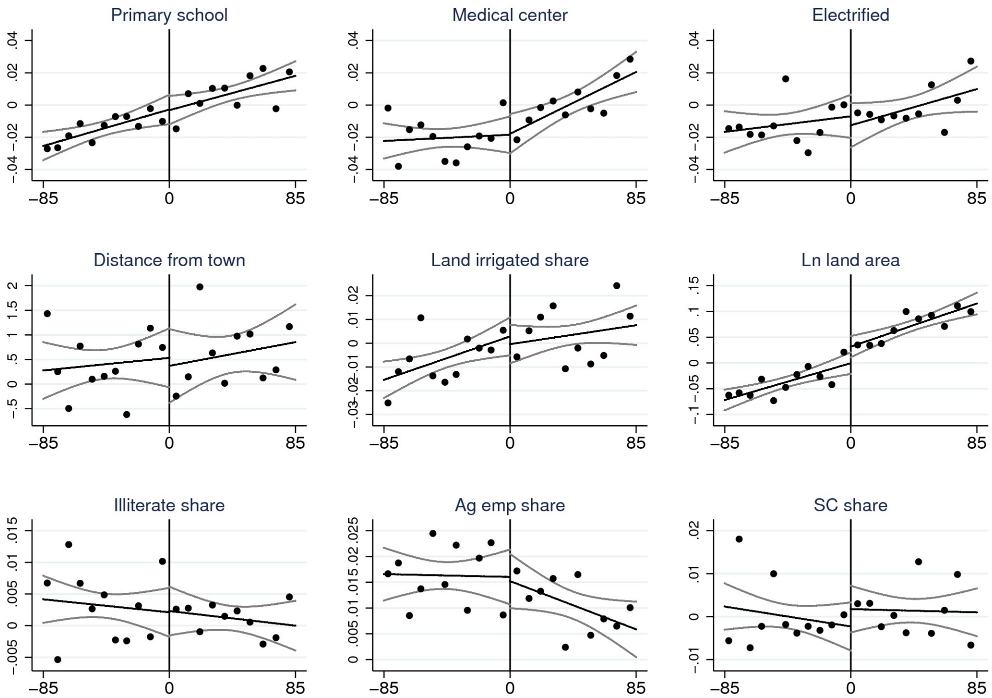
Distribution of running variable (normalized 2001 Population Census village population)



Notes: The figure shows the distribution of village population around the population thresholds. The left panel is a histogram of village population as recorded in the 2001 Population Census. The vertical lines show the program eligibility cutoffs used in this paper, at 500 and 1000. The right panel uses the normalized village population (reported population minus the threshold, either 500 or 1000). It plots a non-parametric regression to each half of the distribution following McCrary (2008), testing for a discontinuity at zero. The point estimate for the discontinuity is 0.04, with a standard error of 0.04.

Figure 3

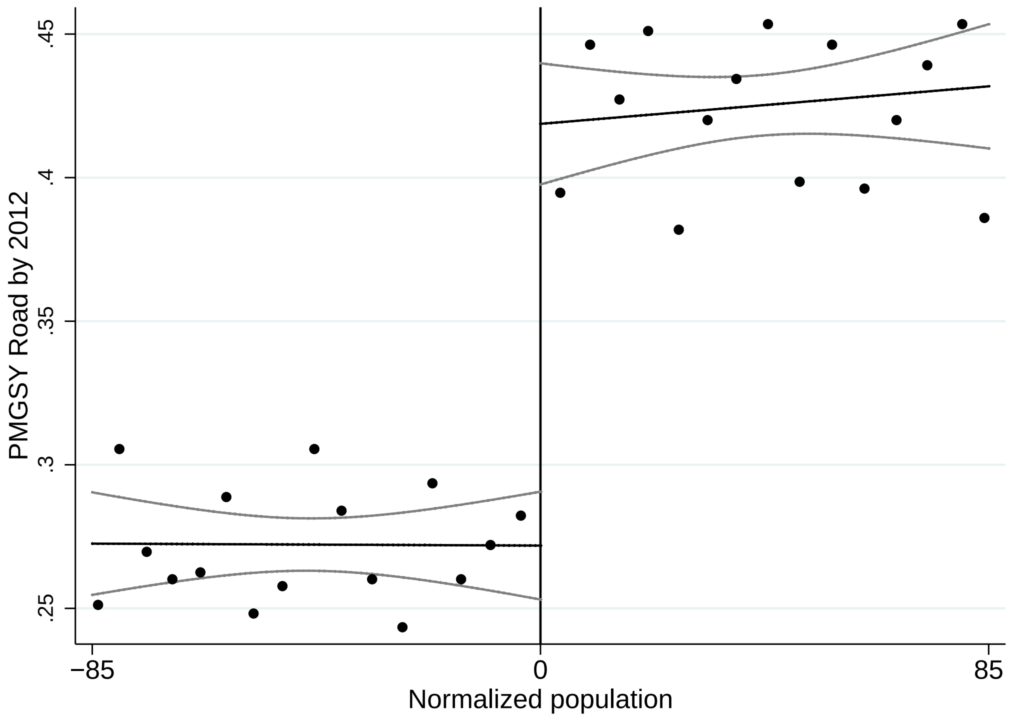
Balance of baseline village characteristics



Notes: The figures plot the conditional expectation function of baseline village characteristics, conditioning on village population. Points to the right of zero are above treatment thresholds, while points to the left of zero are below treatment thresholds. Each point represents approximately fifty observations. As in the main specification, a linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages that did not have a paved road at baseline, with baseline population within an optimal bandwidth (85) of the threshold (see text for details).

Figure 4

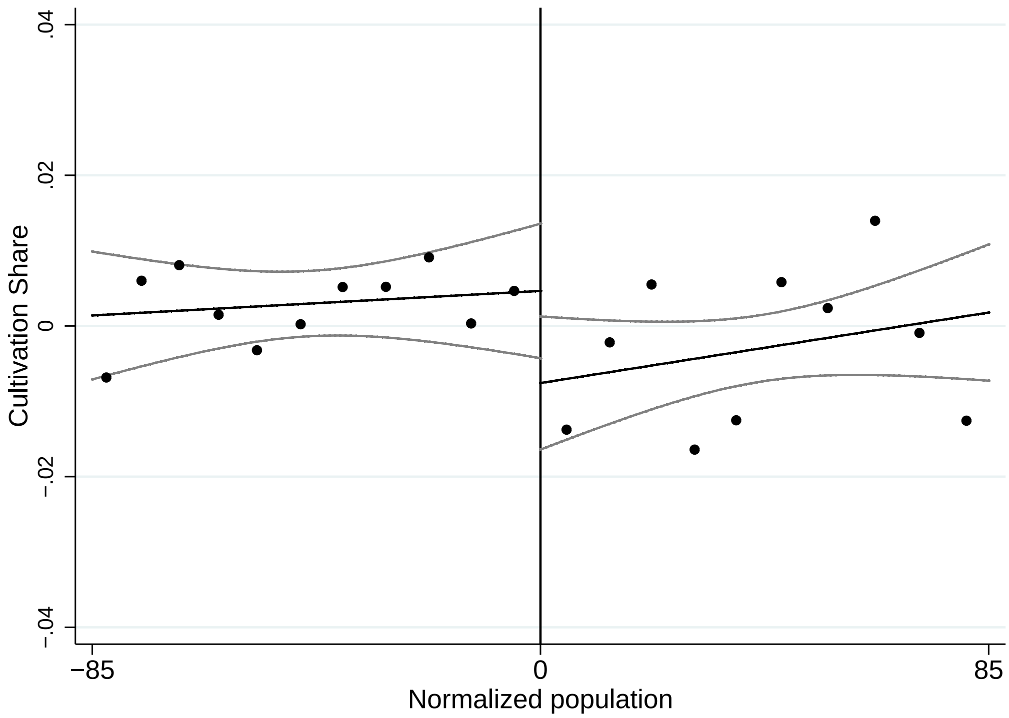
First stage: effect of priority on probability of PMGSY road



Notes: The figure plots the probability of getting a PMGSY road by 2012 over village population in the 2001 Population Census. The sample consists of villages that did not have a paved road at baseline, with baseline population within an optimal bandwidth (85) of the population thresholds (see text for details). Populations are normalized by subtracting the cutoff.

Figure 5

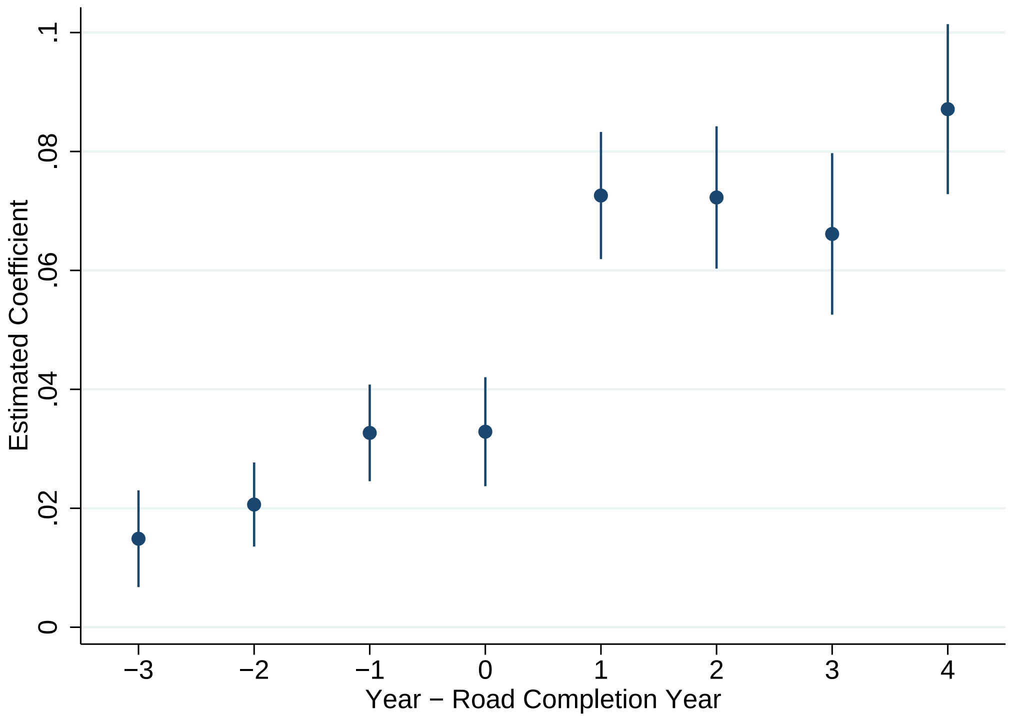
Reduced form: effect of priority on share of households reporting cultivation as primary source of income



Notes: The figure plots the residualized share of households reporting cultivation as the primary source of income (after controlling for all variables in the main specification other than population) over normalized village population in the 2001 Population Census. The sample consists of villages that did not have a paved road at baseline, with baseline population within an optimal bandwidth (85) of the population thresholds (see text for details). Populations are normalized by subtracting the cutoff.

Figure 6

Growth rate of night lights



Notes: The figure plots coefficients (with 95% confidence intervals) of the estimated growth rate (difference in log nighttime luminosity) by year relative to road construction. Year 0 is the year of road construction and year -4 is the omitted dummy variable. The sample is all villages that did not have a paved road at program baseline and received one before 2010, ensuring that we have a full 9 year panel for each village that includes four years of observations before road construction and four years after. The regression includes year and village fixed effects. Heteroskedasticity robust standard errors are clustered at the village level.

References

Aggarwal, Shilpa, "Do Rural Roads Create Pathways out of Poverty? Evidence from India," 2015.

Ali, Rubaba, "Impact of Rural Road Improvement on High Yield Variety Technology Adoption: Evidence from Bangladesh," 2011.

Alkire, Sabina and Suman Seth, "Identifying BPL Households: A Comparison of Methods," *Economic and Political Weekly*, 2013, *48* (2), 49–57.

Allen, Treb and David Atkin, "Volatility, Insurance, and the Gains from Trade," 2015.

Autor, David H, David Dorn, Gordon H Hanson, and Jae Song, "Trade Adjustment: Worker-Level Evidence," *Quarterly Journal of Economics*, 2014, *129* (4), 1799–1860.

Banerjee, Abhijit and Andrew F. Newman, "Information, the Dual Economy, and Development," *The Review of Economic Studies*, 1998, *65* (4), 631–653.

___ **, Esther Duflo, and Nancy Qian**, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," 2012.

Banerjee, AV and AF Newman, "Occupational Choice and the Process of Development," *Journal of Political Economy*, 1993, *101* (2), 274–298.

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

Bazzi, Samuel, Arya Gaduh, Alexander Rothenberg, and Maisy Wong, "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia," 2014.

Binswanger, Hans P., Shahidur R. Khandker, and Mark R. Rosenzweig, "How infrastructure and financial institutions affect agricultural output and investment in India," *Journal of Development Economics*, 1993, *41* (2), 337–366.

Blimpo, M. P., R. Harding, and L. Wantchekon, "Public Investment in Rural Infrastructure: Some Political Economy Considerations," *Journal of African Economies*, jul 2013, *22* (AERC Supplement 2), 57–83.

Brueckner, Jan K., "The Structure of Urban Equilibria: A Unified Treatment of the Muth-Mills Model," in "Handbook of Regional and Urban Economics," Vol. II 1987.

Brueckner, Markus, "Infrastructure, Anocracy, and Economic Growth: Evidence from International Oil Price Shocks," 2014.

Bryan, Gharad and Melanie Morten, "Economic Development and the Spatial Allocation of Labor: Evidence From Indonesia," 2015.

——, **Shyamal Chowdury, and Ahmed Mushfiq Mobarak**, "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 2014, *82* (5), 1671–1748.

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

——, **Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padro i Miquel**, "The Value of Democracy: Evidence from Road Building in Kenya," *American Economic Review*, 2015, *105* (6).

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli, "Agricultural Productivity and Structural Transformation. Evidence from Brazil," 2015.

Calonico, Sebastian, Md Cattaneo, and Rocio Titiunik, "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Casaburi, Lorenzo, Rachel Glennerster, and Tavneet Suri, "Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone," 2013.

Caselli, Francesco, "Accounting for Cross-Country Income Differences," in "Handbook of Economic Growth," Vol. 1 2005, chapter 9, pp. 679–741.

—— **and Wilbur John II Coleman**, "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy*, 2001, *109* (3), 584–616.

Colmer, Jonathan, "Weather, Labour Reallocation, and Industrial Production: Evidence from India," 2015.

de Janvry, Alain, Marcel Fafchamps, and Elisabeth Sadoulet, "Peasant household behaviour with missing markets: Some Paradoxes Explained," *The Economic Journal*, 1991, *101* (409), 1400–1417.

Deichmann, Uwe, Forhad Shilpi, and Renos Vakis, "Urban Proximity, Agricultural Potential and Rural Non-farm Employment: Evidence from Bangladesh," *World Development*, 2009, *37* (3), 645–660.

Dell, Melissa, "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 2010, *78* (6).

Dercon, Stefan, Daniel O. Gilligan, John Hoddinott, and Tassew Woldehanna, "The impact of agricultural extension and roads on poverty and consumption growth in fifteen Ethiopian villages," *American Journal of Agricultural Economics*, 2009, *91* (4), 1007–1021.

Dinkelman, Taryn, "The Effects of Rural Electrification on Employment: New Evidence from South Africa," *American Economic Review*, 2011, *101* (7), 3078–3108.

Dobbs, Richard, Herbert Pohl, Diaan-Yi Lin, Jan Mischke, Nicklas Garemo, Jimmy Hexter, Stefan Matzinger, Robert Palter, and Rushad Nanavatty, "Infrastructure productivity: how to save $1 trillion a year," Technical Report January, McKinsey Global Institute 2013.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," 2012.

___ **and Richard Hornbeck**, "Railroads and American Economic Growth: A "Market Access" Approach," 2015.

Duflo, Esther and Rohini Pande, "Dams," *Quarterly Journal of Economics*, 2007, *122* (2), 601–646.

Einav, Liran and Jonathan Levin, "Economics in the Age of Big Data," *Science*, 2014, *346* (6210), 1–5.

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

Fafchamps, Marcel and Forhad Shilpi, "Cities and Specialisation: Evidence from South Asia," *The Economic Journal*, 2005, *115* (503), 477–504.

Fan, Shenggen and Peter Hazell, "Returns to Public Investments in the Less-Favored Areas of India and China," *American Journal of Agricultural Economics*, 2001, *83* (5), 1217–1222.

Fernando, A. Nilesh, "Shackled to the Soil: The Long-Term Effects of Inherited Land on Labor Mobility and Consumption," 2014.

Foster, Andrew D. and Mark R. Rosenzweig, "Economic Development and the Decline of Agricultural Employment," in "Handbook of Development Economics," Vol. 4 2007, chapter 47, pp. 3051–3083.

___ **and** ___ , "Are Indian Farms Too Small? Mechanization, Agency Costs, and Farm Efficiency," 2011.

Gelman, Andrew and Guido Imbens, "Why high-order polynomials should not be used in regression discontinuity designs," 2014.

Ghani, Ejaz, Arti Grover Goswami, and William R. Kerr, "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing," *Economic Journal*, 2015.

Gibson, John and Scott Rozelle, "Poverty and Access to Roads in Papua New Guinea," *Economic Development and Cultural Change*, 2003, *52* (1), 159–185.

——— **and Susan Olivia**, "The effect of infrastructure access and quality on non-farm enterprises in rural Indonesia," *World Development*, 2010, *38* (5), 717–726.

Goldin, Claudia, "The U-shaped female labour force function in economic development and economic history," in T. Paul Schultz, ed., *Investment in Women's Human Capital*, Chicago and London: University of Chicago Press, 1995.

Gollin, Douglas and Richard Rogerson, "Productivity, transport costs and subsistence agriculture," *Journal of Development Economics*, 2014, *107*, 38–48.

———**, David Lagakos, and Michael E Waugh**, "The Agricultural Productivity Gap," *Quarterly Journal of Economics*, 2014, *129* (2), 939–993.

Harding, Robin, "Attribution and Accountability: Voting for Roads in Ghana," *World Politics*, 2015, *67* (4).

Harris, John R. and Michael P. Todaro, "Migration, unemployment and development: a two-sector analysis," *The American Economic Review*, 1970, *60* (1), 126–142.

Henderson, J Vernon, Adam Storeygard, and David N Weil, "A Bright Idea for Measuring Economic Growth," *American Economic Review*, 2011, *101* (3).

Imbens, Guido and K. Kalyanaraman, "Optimal bandwidth choice for the regression discontinuity estimator," *Review of Economic Studies*, 2012, *79* (3).

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

Imbert, Clement and John Papp, "Short-term Migration, Rural Workfare Programs and Urban Labor Markets: Evidence from India," 2015.

Ito, Takahiro, "Caste discrimination and transaction costs in the labor market: Evidence from rural North India," *Journal of Development Economics*, 2009, *88* (2), 292–300.

Jacoby, Hanan G, "Access to Markets and the Benefits of Rural Roads," *The Economic Journal*, 2000, *110* (465), 713–737.

——— **and Bart Minten**, "On measuring the benefits of lower transport costs," *Journal of Development Economics*, 2009, *89*, 28–38.

Khandker, Shaidur R. and Gayatri B Koolwal, "Estimating the Long-term Impacts of Rural Roads: A Dynamic Panel Approach," 2011.

——, **Zaid Bakht, and Gayatri B. Koolwal**, "The Poverty Impact of Rural Roads: Evidence from Bangladesh," *Economic Development and Cultural Change*, 2009, *57* (4), 685–722.

Kim, Dae-Il and Robert H Topel, "Labor Markets and Economic Growth: Lessons from Korea's Industrialization, 1970-1990," in Richard B. Freeman and Lawrence F. Katz, eds., *Differences and Changes in Wage Structures*, University of Chicago Press, 1995, chapter 7, pp. 227–264.

Kochar, Anjini, "Smoothing Consumption by Smoothing Income: Hours-of-Work Responses to Idiosyncratic Agricultural Shocks in Rural India," *The Review of Economics and Statistics*, 1999, *81* (1), 50–61.

La Porta, Rafael and Andrei Shleifer, "Informality and Development," *Journal of Economic Perspectives*, 2014, *28* (3), 109–126.

Lewis, W. Arthur, "Economic Development with Unlimited Supplies of Labour," *The Manchester School*, 1954, *22* (2), 139–191.

Lipscomb, Molly, Ahmed Mushfiq Mobarak, and Tania Bahram, "Development effects of electrification: evidence from the geologic placement of hydropower plants in Brasil," *American Economic Journal: Applied Economics*, 2013, *5* (2), 200–231.

Matsuyama, Kiminori, "Agricultural Productivity, Comparative Advantage and Economic Growth," *Journal of Economic Theory*, 1992, *58*, 317–334.

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

McMillan, Margaret, Dani Rodrik, and Iñigo Verduzco-Gallo, "Globalization, Structural Change, and Productivity Growth, with an Update on Africa," *World Development*, 2014, *63*, 11–32.

Meyer, Bruce D and James X Sullivan, "Measuring the Well-Being of the Poor Using Income and Consumption," *The Journal of Human Resources*, 2003, *38* (2003).

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

Mu, Ren and Dominique van de Walle, "Rural Roads and Local Market Development in Vietnam," *Journal of Development Studies*, may 2011, *47* (5), 709–734.

Munshi, Kaivan and M Rosenzweig, "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *American Economic Review*, 2016, *106* (1), 46–98.

National Rural Roads Development Agency, "Pradhan Mantri Gram Sadak Yojana - Operations Manual," Technical Report, Ministry of Rural Development, Government of India 2005.

Raballand, Gael, Rebecca Thornton, Dean Yang, Jessica Goldberg, Niall Keleher, and Annika Muller, "Are Rural Road Investments Alone Sufficient to Generate Transport Flows? Lessons from a Randomized Experiment in Rural Malawi and Policy Implications," 2011.

Ranis, Gustav and John C. H. Fei, "A Theory of Economic Development," *The American Economic Review*, 1961, *51* (4).

Redding, Stephen J and Matthew A Turner, "Transportation Costs and the Spatial Organization of Economic Activity," in "Handbook of Regional and Urban Economics, Vol. 5B" 2015, chapter 20.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu, "Agriculture and aggregate productivity: A quantitative cross-country analysis," *Journal of Monetary Economics*, 2008, *55* (2), 234–250.

Sharma, Ajay and S. Chandrasekhar, "Growth of the Urban Shadow, Spatial Distribution of Economic Activities, and Commuting by Workers in Rural and Urban India," *World Development*, 2014, *61*, 154–166.

Shrestha, Slesh A., "The Effect of Roads on Farmland Values: Evidence from the Topography-based Road Network in Nepal," 2015.

Sotelo, Sebastian, "Domestic Trade Frictions and Agriculture," 2015.

Spears, Dean, "Effects of sanitation on early-life health: Evidence from a governance incentive in rural India," 2015.

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

Tombe, Trevor, "The Missing Food Problem: Trade, Agriculture, and International Productivity Differences," *American Economic Journal: Macroeconomics*, 2014, *7* (3), 226–258.

Wantchekon, Leonard and Piero Stanig, "The Curse of Good Soil? Land Fertility, Roads, and Rural Poverty in Africa," 2015.

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

Zhang, Xiaobo and Shenggen Fan, "How Productive is Infrastructure? A New Approach and Evidence from Rural India," *American Journal of Agricultural Economics*, 2004, *86* (2), 492–501.

A Appendix: Additional figures and tables

Table A1

Regression discontinuity estimate of PMGSY road on cultivation as primary income source (share of households), by threshold

	Full Sample	500 Cutoff	1000 Cutoff
Road	-0.096	-0.080	-0.126
	(0.039)**	(0.052)	(0.052)**
Outcome Mean	.4299	.4439	.3822
N	19612	15071	4541
R2	.4743	.4817	.4258

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

This table presents regression discontinuity estimates from Equation 3 of the effect of PMGSY prioritization on the share of households reporting cultivation and manual labor as the primary source of income. The first column restricts the sample to villages with populations within the optimal bandwidth (85) of 500, while the second column restricts the sample to villages within the optimal bandwidth (85) of the 1000 population threshold. For each regression, the outcome mean for the control group (villages with population below the threshold) is also shown. The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table A2

Effect of PMGSY road on measures of data quality

	Population	Under 6 Pop.	Phone
Road	-14.233	0.003	-0.006
	(11.015)	(0.005)	(0.048)
Outcome Mean	11.45	-.03023	.09343
N	19612	19612	19606
R2	.1111	.6846	.1795

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

This table presents regression discontinuity estimates from Equation 3 of the effect of PMGSY treatment on a measure of data quality: the difference in the same variable between the 2011 Population Census and 2012 SECC. The first column presents the estimated effect on village population, the second for population under 6, and the third for the share of households reporting ownership of any type of phone (mobile or landline). The sample consists of villages that did not have a paved road at baseline, with baseline population within an optimal bandwidth (85) of the threshold (see text for details). For each regression, the outcome mean for the control group (villages with population below the threshold) is also shown. The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table A3

Effect of road priority on PMGSY road treatment and cultivation as primary income source (share of households), for primary and placebo sample

Panel A. Outcome : Road treatment (first stage)

	Main Sample	Placebo Sample
Road Priority	0.1516	0.0134
	(0.0124)***	(0.0129)
Outcome Mean	0.3132	0.2077
N	19612	14491
R2	0.2737	0.2569

Panel B. Outcome : Cultivation share (reduced form)

	Main Sample	Placebo Sample
Road Priority	-0.014	-0.002
	(0.0058)**	(0.0073)
Outcome Mean	0.4285	0.3949
N	19612	14491
R2	0.4953	0.4552

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

This table presents a comparison of estimates from Equation 2 of the effect of PMGSY prioritization on a village's probability of treatment and reduced form estimates of the effect of PMGSY prioritization on the share of households reporting cultivation as their primary source of income for the main sample of states that adhered to the implementation cutoffs and a placebo sample of states that did not follow the cutoffs. The first column presents estimates for the sample of states who followed the cutoff rules, while the second column presents estimates for the sample that did not follow the cutoff. The sample consists of villages that did not have a paved road at baseline, with baseline population within an optimal bandwidth (85) of the threshold (see text for details). For each regression, the outcome mean for the control group (villages with population below the threshold) is also shown. The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table A4

First stage, reduced form and RD estimate of PMGSY road on cultivation as primary source of income (share of households), by distance to urban centers

Panel A. First Stage

	Full	Quart 1	Quart 2	Quart 3	Quart 4
Road Priority	0.1517	0.1372	0.1268	0.1761	0.1623
	(0.0125)***	(0.0233)***	(0.0240)***	(0.0262)***	(0.0265)***
Outcome Mean	0.3131	0.2508	0.3069	0.3175	0.3775
N	19614	4903	4903	4903	4903
R2	0.2729	0.2931	0.3538	0.2861	0.2867

Panel B. Reduced Form

	Full	Quart 1	Quart 2	Quart 3	Quart 4
Road Priority	-0.014	-0.039	-0.013	-0.016	0.0072
	(0.0058)**	(0.0118)***	(0.0113)	(0.0115)	(0.0123)
Outcome Mean	0.4285	0.4179	0.4356	0.4348	0.4257
N	19614	4903	4903	4903	4903
R2	0.4950	0.4982	0.5365	0.5184	0.5299

Panel C. Regression Discontinuity

	Full	Quart 1	Quart 2	Quart 3	Quart 4
Road	-0.097	-0.286	-0.102	-0.092	0.0447
	(0.0388)**	(0.0977)***	(0.0891)	(0.0654)	(0.0747)
Outcome Mean	0.4285	0.4179	0.4356	0.4348	0.4257
N	19614	4903	4903	4903	4903
R2	0.4733	0.3241	0.5104	0.5014	0.5255

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

This table presents estimates of the effect of PMGSY road prioritization on a village's probability of receiving a PMGSY road before 2012 (first stage, Panel A) and share of households whose primary source of income is cultivation (reduced form, Panel B) from Equation 2. Panel C estimates the impact of a PMGSY road on the share of households in cultivation from Equation 3. The first column presents results for the full sample. The second column presents results for villages in the first quartile of straight line distance to cities with at least 500,000 inhabitants in the 2001 Population Census. Columns 3-5 are the second, third and fourth quartiles, respectively. The first quartile includes villages below 62.9 km from such cities, the second quartile up to 100.8 km, the third up to 142.9 km and the final quartile all villages with distances greater than 142.9k m. The sample consists of villages that did not have a paved road at baseline, with baseline population within an optimal bandwidth (85) of the threshold (see text for details). For each regression, the outcome mean for the control group (villages with population below the threshold) is also shown. The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table A5

Reduced form estimate of PMGSY road on major TSC variables

	Open Defecation	Latrine in Premises	Pit Latrine - with slab	Pit Latrine - without slab
Road priority	-0.006	0.007	0.003	0.000
	(0.009)	(0.009)	(0.005)	(0.003)
N	4540	4540	4540	4540
r2	0.38	0.38	0.38	0.10

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

The Total Sanitation Campaign (TSC) is stated to have "aimed to transition rural households from open defecation to use of onsite pit latrines" (Spears, 2015). The program began construction of latrines in 2001. The outcomes considered here are 2011 measures of (in order) percentages of households who report: open defecation; the existence of a latrine within premises; an in-house pit latrine with slab or ventilated improved pit; and an in-house pit latrine without slab/open pit. The sample has been restricted to villages with population within the optimal bandwidth (85) of 1000, the cutoff used by the TSC. The sample of states here come from our main PMGSY specification. The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table A6

Impact of road construction on population growth

	Full	Quart 1	Quart 2	Quart 3	Quart 4
Road	0.001	0.004	-0.003	0.002	-0.001
	(0.002)	(0.005)	(0.004)	(0.004)	(0.005)
Outcome Mean	1.018	1.017	1.019	1.019	1.017
N	18570	4582	4644	4672	4672
R2	.2546	.2399	.2938	.3038	.3253

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

This table presents regression discontinuity estimates from Equation 3 of the effect of PMGSY treatment by 2011 on annualized population growth for the period 2001 to 2011. The first column presents results for the full sample. The second column presents results for villages in the first quartile of distance to cities with at least 500,000 inhabitants in the 2001 Population Census. Columns 3-5 are the second, third and fourth quartiles, respectively. The sample consists of villages that did not have a paved road at baseline, with baseline population within the optimal bandwidth (85) of the threshold (see text for details). The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table A7

RD estimate of PMGSY road on distribution of landholdings (share of households)

	Landless	0-1 Acres	1-2 Acres	2-4 Acres	4-10 Acres	10-25 Acres	25+ Acres
Road	-0.029	0.036	0.003	-0.012	-0.015	0.008	0.008
	(0.040)	(0.031)	(0.018)	(0.017)	(0.017)	(0.009)	(0.006)
Outcome Mean	.4194	.1991	.1248	.1162	.09667	.03354	.01023
N	19553	19553	19553	19553	19553	19553	19553
R2	.3505	.4301	.2089	.2361	.3868	.3942	.1785

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

This table presents regression discontinuity estimates from Equation 3 of the effect of road construction on the share of village households with landholdings in a given range. The first column reports the estimate effect on the share of households reporting no agricultural land, followed by five columns for households owning agricultural land. The sample consists of villages that did not have a paved road at baseline, with baseline population within the optimal bandwidth (85) of the threshold (see text for details). The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table A8

RD estimate of PMGSY road on cultivation (share of households), by caste

	Scheduled Caste	Scheduled Tribe	General
Road	-0.166	0.033	-0.076
	(0.060)***	(0.053)	(0.045)*
Outcome Mean	.2624	.3362	.467
N	15424	11192	18795
R2	.2199	.4274	.3977

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

This table presents regression discontinuity estimates from Equation 3 of the effect of road construction on the share of village households whose primary source of income is cultivation. The first column reports the estimated effect on Scheduled Caste households, the second on Schedule Tribe and the third for all other households. The sample consists of villages that did not have a paved road at baseline, with baseline population within the optimal bandwidth (85) of the threshold (see text for details). The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table A9

RD estimate of PMGSY road on bus service

	Full	Quart 1	Quart 2	Quart 3	Quart 4
Road	0.110	0.346	0.328	-0.124	-0.022
	(0.076)	(0.179)*	(0.158)**	(0.143)	(0.142)
N	27201	6985	6648	6710	6858
r2	0.30	0.23	0.25	0.30	0.35

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

This table presents regression discontinuity estimates from Equation 3 of the effect of PMGSY treatment on availability of a scheduled bus service in the village. The outcome variable is an indicator variable that takes on the value 1 if the village is served by scheduled bus service in 2011, and a 0 otherwise. Column 1 presents results for the full sample, while Columns 2-5 present results by quartile of distance to a major city. Heteroskedasticity robust standard errors are reported below point estimates.

Table A10

Reduced form estimate of road priority on cultivation as primary source of income (share of households), by distance to Golden Quadrilateral and large cities

	(1)	(2)	(3)	(4)
Road priority	-0.021	-0.004	-0.024	-0.005
	(0.010)**	(0.014)	(0.011)**	(0.016)
Road Priority * GQ Near		-0.029		-0.016
		(0.020)		(0.027)
GQ Near		0.053		-0.051
		(0.107)		(0.209)
Road Priority * City Near			0.007	-0.000
			(0.011)	(0.015)
City Near			0.006	0.006
			(0.009)	(0.013)
Road Priority * GQ * City				-0.014
				(0.033)
N	6893	6893	6893	6893
r2	0.50	0.50	0.51	0.53

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

This table presents reduced form estimates from Equation 2 of the effect of PMGSY prioritization on the share of households reporting cultivation as the primary source of income. The sample is restricted to states through which the Golden Quadrilateral (GQ) highway network passes (among those who followed the population prioritization rules): Maharashtra, Orissa, Rajasthan and Uttar Pradesh. The first columns presents the main result: the effect of road priority on cultivation. The second column adds an interaction of road priority and below-median distance to the Golden Quadrilateral, as well as a control for below-median distance to the Golden Quadrilateral. Column 3 does runs the same specification for proximity to cities of 500,000+ inhabitants. Column 4 includes both interactions and the triple interaction. The sample consists of villages that did not have a paved road at baseline, with baseline population within the optimal bandwidth (85) of the threshold (see text for details). The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table A11

RD estimate of PMGSY road on ownership of mechanized farm and irrigation equipment (share of households), by size of landholdings

Panel A. Mechanized Farm Equipment

	Landless	0-1 Acres	1-2 Acres	2-4 Acres	4-10 Acres	10-25 Acres	25+ Acres
Road Priority	0.0015	-0.011	-0.024	-0.030	-0.059	-0.181	-0.032
	(0.0043)	(0.0115)	(0.0156)	(0.0213)	(0.0323)*	(0.0528)***	(0.0817)
Outcome Mean	0.0052	0.0150	0.0293	0.0555	0.1227	0.2573	0.3768
N	19250	17094	18157	18272	17194	12602	6987
R2	0.0477	0.0863	0.1217	0.1576	0.1953	0.1968	0.3228

Panel B. Irrigation Equipment

	Landless	0-1 Acres	1-2 Acres	2-4 Acres	4-10 Acres	10-25 Acres	25+ Acres
Road Priority	-0.019	-0.067	-0.034	-0.096	-0.078	-0.104	-0.164
	(0.0098)**	(0.0299)**	(0.0356)	(0.0422)**	(0.0483)	(0.0588)*	(0.0825)**
Outcome Mean	0.0116	0.0861	0.1624	0.2531	0.3755	0.5231	0.5574
N	19250	17091	18154	18269	17193	12602	6987
R2	0.0561	0.2126	0.3397	0.3862	0.4170	0.3785	0.4077

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

This table presents regression discontinuity estimates from Equation 3 of the effect of PMGSY treatment on the share of households reporting (A) ownership of mechanized farm equipment and (B) ownership of irrigation equipment. The first column reports results for households reporting no agricultural land, followed by six columns for households owning agricultural land. The sample consists of villages that did not have a paved road at baseline, with baseline population within the optimal bandwidth (85) of the threshold (see text for details). The specification includes baseline village-level controls (primary school, medical center, electrification, distance to nearest town, log total acres under cultivation, share of agricultural land irrigated, share of households working in agriculture, and share of population belonging to a scheduled caste) as well as district-cutoff fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table A12

Mean share of individuals reporting place of work classified as urban, by road treatment and distance to town of population 500K+ quartile

Road	Quartile				
Treatment	1	2	3	4	Total
No	.12214182	.0465136	.0790319	.13386714	.09561312
Yes	.17764544	.14430664	.14969775	.07375326	.13643238

This table presents the mean share of rural workers who report their place of work as urban, according to the 68th Round of the National Sample Survey Employment/Unemployment data. As in the rest of the paper, the sample is restricted to villages that did not have a paved road in 2001. Results are presented by PMGSY treatment by 2011 (rows) and the quartile of their distance from cities of at least 500,000 inhabitants in the 2001 Population Census (columns).

Figure A1

Sample page from SECC

SECC ड्राफ्ट सूची - ग्रामीण

राज्य :RAJASTHAN	ज़िला :Ajmer	तहसील :Ajmer	शहर/ग्राम :Ajaysar	वार्ड कोड नंबर (केवल शहर के लिए) :0000	गणन ब्लॉक -उप खंड :0158_0

घरेलू संख्या : 0003	घर के प्रकार : साधारण	ग्राम पंचायत :-AJAYSAR	आदिम जनजाति वर्ग से है : नहीं	वैधानिक रुप से छुडाया गया बंधुवा मजदूर : नहीं	हाथ से मैला साफ़ करने वाले : नहीं

संख्या	नाम	मुखिया से संबंध	लिंग जन्मतिथि	पिता का नाम माता का नाम	वैवाहिक स्थिति#	व्यवसाय/ गतिविधि	अनु. जाति / जनजाति / अन्य	विकलांगता	शिक्षा
001	[redacted]	मुखिया	पुरुष 1953	[redacted]	2	मजदूर	अन्य	कोई निःशक्तता नहीं	निरक्षर
002	[redacted]	पत्नी	स्त्री 1955	[redacted]	2	मजदूर	अन्य	कोई निःशक्तता नहीं	निरक्षर
003	[redacted]	पुत्र	पुरुष 1989	[redacted]	1	मजदूर	अन्य	कोई निःशक्तता नहीं	पूर्व माध्यमिक

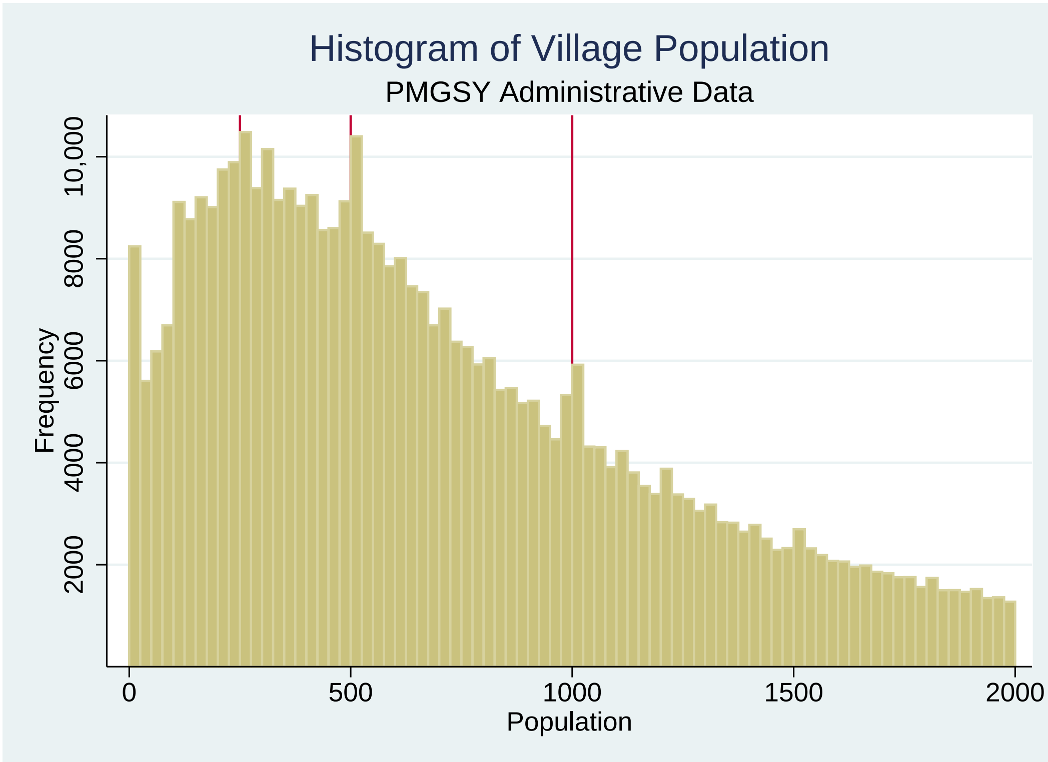
भाग 1 विवरण : आवासीय/निवासीय				भाग 3 रोजगार और आय विशेषताओं					भाग 4 : विवरण सम्पत्तियां			भाग5 अ: भूमि स्वामित्व (एकड में)				भाग 5 ब: अन्य भूमि स्वामित्व		
मकान के दीवार की प्रमुख सामग्री #	मकान की छत की प्रमुख सामग्री #	मकान का मालिकाना हक की स्थिति	निवास के कमरों की संख्या	नियमित वेतन पाने वाला कोई परिवार का सदस्य	आयकर या वृति कर दाता है	स्वयं की /संचलित ऐसी संस्था जो शासन द्वारा पंजीकृत है	परिवार के सबसे अधिक कमाने वाले सदस्य का मासिक आय	परिवार की आय का मुख्य स्त्रोत	रेफ्रिजिरेटर	टेलीफोन / मोबाईल फोन	दो/तीन/चार पहिया या मछली पकडने की नाव पंजीकृत	स्वामित्व की भूमि (वास भूमि को छोडकर)	कुल असिंचित भूमि	2 फसलों वाली सिंचाई भूमि	अन्य सिंचित भूमि	यंत्रीकृत तीन/चार ट्रैक्टर कृषि उपकरण	सिंचाई उपकरण(नलकूप . बोर . डीज़ल/मिट्टी के तेल/विद्युत पंप सेट. फव्वारा/ड्रिप सिंचाई आदि समेत)	किसान क्रेडिट कार्ड की सीमा 50000 रुपए या अधिक है।
6	6	स्वय	4	नहीं	नहीं	नहीं	10,000 या अधिक	1	हां	केवल मोबाइल	दो पहिया	हां	1.0	3.0	1.0	नहीं	हां	नहीं

Ver:4.0.5a DB: 070 09 Mar 2014 06:06:02 PM Signature_____________ Page 5 of 200

This is a sample page taken from a PDF file that was scraped from secc.gov.in. Individual-level variables are name, relationship with head of household, gender, date of birth, parents' names, marital status, occupation, caste category, disability and education. Household-level variables are wall material, roof material, house ownership, dwelling room count, salaried job, payment of income tax, ownership of registered enterprise, monthly income, source of income, asset ownership (refrigerator, telephone, vehicle, mechanized farm equipment, irrigation equipment, Kisan credit card), and land ownership.

Figure A2

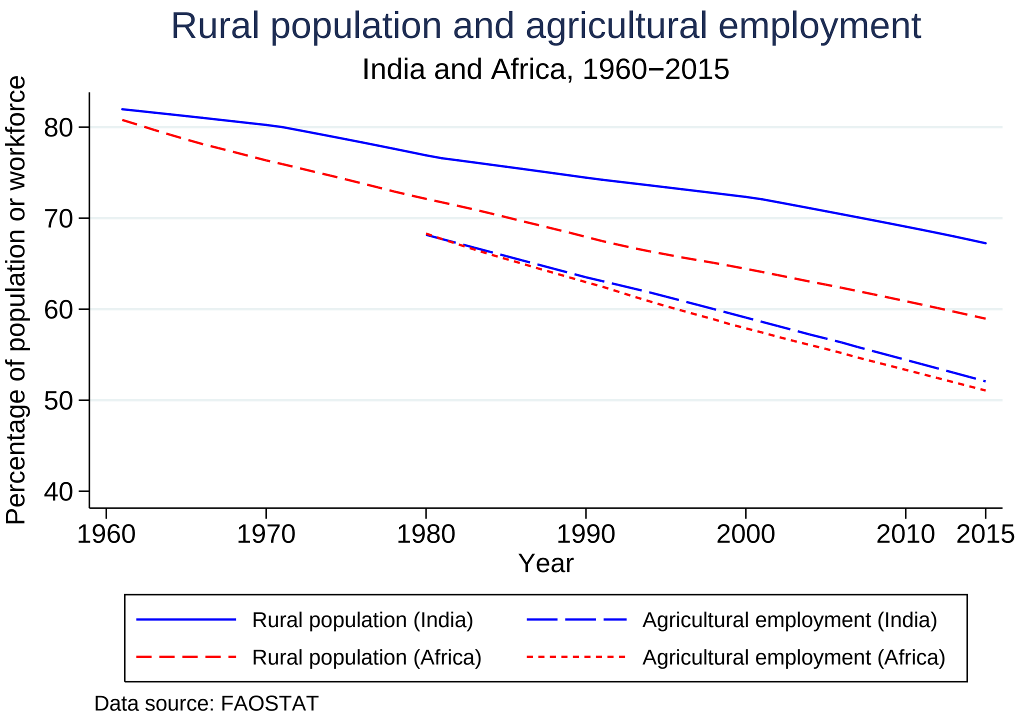
Histogram of habitation populations (PMGSY OMMS)



The figure shows the histogram of village population as reported in the PMGSY Online Monitoring and Management System. The vertical lines show the program eligibility cutoffs at 500 and 1000.

Figure A3

Comparison of India and Africa's rates of urbanization and structural transformation



This figure displays the rural population and agricultural employment percentages for India (blue) and Africa (red).