Rural Roads and Local Economic Development

Sam Asher and Paul Novosad*

January 2019

Abstract

Nearly one billion people worldwide live in rural areas without access to national paved road networks. We estimate the impacts of India's $40 billion national rural road construction program using a fuzzy regression discontinuity design and comprehensive household and firm census microdata. Four years after road construction, the main effect of new feeder roads is to facilitate the movement of workers out of agriculture. However, there are no major changes in agricultural outcomes, income or assets. Employment in village firms expands only slightly. Even with better market connections, remote areas may continue to lack economic opportunities.

JEL Codes: O12/O18/J43.

*Asher: The World Bank, 1818 H St NW, Washington, DC 20009 (email: sasher@worldbank.org); Novosad: 6106 Rockefeller Hall, Hanover, NH 03755 (email: paul.novosad@dartmouoth.edu). This paper was accepted to the AER under the guidance of Esther Duflo, Coeditor. We wish to thank the editor and anonymous referees, as well as Abhijit Banerjee, Lorenzo Casaburi, Melissa Dell, Eric Edmonds, Ed Glaeser, Doug Gollin, Ricardo Hausmann, Rick Hornbeck, Clement Imbert, Lakshmi Iyer, Radhika Jain, Asim Khwaja, Michael Kremer, Erzo Luttmer, 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 Toby Lunt, Kathryn Nicholson, and Taewan Roh for exemplary research assistance. Dr. IK Pateriya was generous with his time and knowledge of the PMGSY. This paper previously circulated as "Rural Roads and Structural Transformation." This project received financial support from the Center for International Development and the Warburg Fund (Harvard University), the IGC, the IZA GLM-LIC program, and NSF Doctoral Dissertation Improvement Grant 1156205. This document is an output from a project funded by the UK Department for International Development (DFID) and the Institute for the Study of Labor (IZA) for the benefit of developing countries. The views expressed are not necessarily those of DFID or IZA. All errors are our own.
Keywords: Infrastructure, Roads, Rural Development

I Introduction

Nearly one billion people worldwide live more than 2 km from a paved road, with one-third living in India (Roberts et al., 2006; The World Bank Group, 2016). Fully half of India's 600,000 villages lacked a paved road in 2001. To remedy this, the Government of India launched the Pradhan Mantri Gram Sadak Yojana (Prime Minister's Village Road Program, or PMGSY) in 2000. Premised on the idea that "poor road connectivity is the biggest hurdle in faster rural development" (Narayanan, 2001) and promising benefits from poverty reduction to increased employment opportunities in villages (National Rural Roads Development Agency, 2005), by 2015 the PMGSY had funded the construction of all-weather roads to nearly 200,000 villages at a cost of almost $40 billion. Yet rural areas may have other disadvantages that may make it difficult to realize these gains; for example, they lack agglomeration economies and complementary inputs such as human capital. Lowering transport costs may not be enough to transform economic activity and outcomes in rural areas.

Existing research is largely supportive of policymakers' claims: rural road construction is associated with increases in farm and non-farm economic growth as well as poverty reduction. But the causal impacts of rural roads have proven difficult to assess, mainly due to the endogeneity of road placement. The high costs and potentially large benefits of infrastructure investments mean that the placement of new roads is typically correlated with both economic and political characteristics of locations (Blimpo et al., 2013; Brueckner, 2014; Burgess et al., 2015; Lehne et al., 2018). We overcome this challenge by taking advantage of an implementation rule that targeted roads to villages with population exceeding two discrete thresholds (500 and 1,000). This rule causes villages just above the population threshold to be 22 percentage points more likely to receive a road, allowing us to estimate the causal impact of rural roads using a fuzzy regression discontinuity design.

We construct a high spatial resolution dataset that combines administrative microdata covering all households and firms in our regression discontinuity sample of villages with remote sensing data and village aggregates describing amenities, infrastructure, and demographic information. Because variation induced by program rules is across villages rather than across larger administrative units, and because of the possibility of heterogeneous effects by individual characteristics, village-identified microdata are essential for studying the impacts of roads. The limitation of this approach is that the administrative data are based on shorter questionnaires than traditional regional sample surveys. On average, we observe outcomes four years after road completion, meaning that we estimate the short to medium run impact of these roads.

In contrast to the dramatic economic benefits anticipated by policymakers, rural roads do not appear to transform village economies. Roads cause a substantial increase in the availability of transportation services, but we find no evidence for increases in assets or income. Farmers do not own more agricultural equipment, move out of subsistence crops, or increase agricultural production. We follow the methodology of Elbers et al. (2003) to predict consumption from the set of asset and income variables in the individual microdata. We can rule out a 10% increase in predicted consumption with 95% confidence, with no significant or economically meaningful subgroup heterogeneity in terms of occupation, education, or position in the consumption distribution.

We do find that rural roads lead to a large reallocation of workers out of agriculture. A new road causes a 9 percentage point decrease in the share of workers in agriculture and an equivalent increase in wage labor. These impacts are most pronounced among the groups likely to have the lowest costs and highest potential gains from participation in labor markets: households with small landholdings and working age men.

We find suggestive evidence that the growth in non-agricultural workers is due to greater access to jobs outside the village. We estimate a small and insignificant increase in village non-farm employment (4 workers per village), which can explain only 23% of the point estimate on reallocation of workers out of agriculture, although we cannot reject equality of these estimates. We can decisively rule out small changes in permanent migration, implying that the results we find are not the product of compositional changes to the village population.

In short, we find that the primary impact of new roads is to make it easier for workers to gain access to non-agricultural jobs. Our research suggests that rural roads do not meaningfully facilitate growth of village firms or predicted consumption in the short to medium run. Roads alone appear to be insufficient to transform the economic structure of remote villages.

This paper contributes to a wide literature estimating the impacts of investments in transportation infrastructure. New highways and railroads have been shown to have substantial impacts on the allocation of economic activity, land use, and migration.[1] Our finding that rural roads do not lead to major economic changes apart from reallocation of labor from agriculture is consistent with Faber (2014), who finds that Chinese highways actually lead to decreases in local GDP in rural areas newly connected to more productive urban centers. But studies of major transportation corridors have limited applicability to the rural roads that we study, which connect poor, rural villages to regional markets. Existing research on rural roads in developing countries has used difference-in-differences and matching methods, largely finding positive impacts on both agricultural and non-agricultural earnings.[2] These

[1] Trunk transportation infrastructure has been shown to raise the value of agricultural land (Donaldson and Hornbeck, 2016), increase agricultural trade and income (Donaldson, 2018), reduce the risk of famine (Burgess and Donaldson, 2012), increase migration (Morten and Oliveira, 2018) and accelerate urban decentralization (Baum-Snow et al., 2011). Results on growth have proven somewhat mixed: there is evidence that reducing transportation costs can increase (Ghani et al., 2016; Storeygard, 2016), decrease (Faber, 2014), or leave unchanged (Banerjee et al., 2012) growth rates in local economic activity. Atkin and Donaldson (2017) show that intra-country trade costs are very high in developing countries, with remote areas benefiting little from increased integration into world markets. For a recent survey of the economic impacts of transportation costs, see Redding and Turner (2015).

[2] 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). Concurrent research on the PMGSY demonstrates that districts that built more roads experienced improved economic outcomes (Aggarwal, 2018), incidentally treated villages experienced gains in agriculture (Shamdasani, 2018), and PMGSY increased educational outcomes (Mukherjee, 2012; Adukia et al., 2017). Other papers also suggest that the lack of rural transport infrastructure may be a significant contributor to rural underdevelopment. Wantchekon et al. (2015) provide evidence that transport costs are a strong predictor

studies are both limited in sample size (the largest examines just over 100 roads) and in their ability to address the endogeneity of road placement. Our study is the first large-scale study on rural roads with exogenous variation in road placement; in this regard we join recent work that has used instrumental variables to estimate the impacts of major infrastructural investments such as dams (Duflo and Pande, 2007) and electrification (Dinkelman, 2011; Lipscomb et al., 2013). The small treatment effects that we detect, especially when contrasted with a district-level analysis of the same program (Aggarwal, 2018), suggest that new roads are disproportionately built in villages that are growing for other reasons.

We also add to a large literature seeking to understand the barriers to reallocation of labor out of agriculture in developing countries. Much emphasis has been put on the role of agricultural productivity in facilitating structural transformation.[3] Theoretically, there is reason to believe that transport costs could also play an important role: if rural workers are unable to access outside nonfarm jobs, or if rural firms are unable to grow due to high transport costs, roads may accelerate structural transformation in poor countries. There is considerable evidence that across the developing world, labor productivity outside agriculture may be higher than in agriculture (Gollin et al., 2014; McMillan et al., 2014). We join recent research that finds that high transportation costs are an important barrier to the spatial and sectoral allocation of labor (Bryan et al., 2014; Bryan and Morten, 2015). However, we find that reallocation of labor out of agriculture is not necessarily associated with other large changes to the village economy.

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. 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. Later work generally demonstrated that rural roads are associated with large economic benefits by looking at their impact on agricultural land values (Jacoby, 2000; Shrestha, 2017), estimated willingness to pay for agricultural households (Jacoby and Minten, 2009), complementarities with agricultural productivity gains (Gollin and Rogerson, 2014), search and competition among agricultural traders (Casaburi et al., 2013), and agricultural productivity and crop choice (Sotelo, 2018). In an urban setting, Gonzalez-Navarro and Quintana-Domeque (2016) find that paving streets lead to higher property values and consumption.

[3]For a recent example and discussion of the literature, see Bustos et al. (2016).

The rest of the paper proceeds as follows: Section II provides a theoretical discussion of how rural roads may affect local economic activity. Section III provides a description of the rural road program. Sections IV and V describe the data construction and empirical strategy. Section VI presents results and discussion. Section VII concludes.

II Conceptual Framework

In this section, we sketch out a conceptual framework for understanding the impacts of new roads on village economies. Because we are interested in villages' productive structure, we explore impacts on occupational choice, agricultural production, and nonfarm firms. We focus on a set of channels that have received attention in existing research and in policymakers' justification for building rural roads.

The first order effect of a feeder road is to reduce transportation costs between a village and external markets, causing prices and wages to move toward prices outside the village. Given the sample of previously unconnected villages in India, this almost always implies higher wages, lower prices for imported goods, and higher prices for exported goods.

We first consider farm production. A decline in the prices of imported inputs such as fertilizer and seeds can be expected to lead to greater input use and increased agricultural production. Changes in farmgate prices will cause crop choice to move in the direction of crops with the greatest price increases—those where the village has a comparative advantage. If agricultural production increases, it will also increase labor demand in agriculture, though these effects may be small or even reversed if production shifts to less labor intensive crops or if it becomes easier to import labor-substituting technology such as tractors.

The major offsetting effect is the increased access of village workers to external labor markets, which is likely to raise village wages. Higher labor costs will make farm work more expensive and may cause farms to reduce production and shift toward less labor intensive crops or technologies.

The impacts of roads on non-farm production in the village are analogous. Lower input prices and higher output prices will increase the production of non-farm goods, but these will be offset by higher wages. The relative changes in on-farm and off-farm production and labor demand will depend on the magnitude of the relative price changes between these markets.

These are the main channels that typically underlie the argument that rural roads will help grow the rural economy, both on and off the farm. But importantly, note that none of these production increases are unambiguous. The external labor demand effects could dominate the input/output price effects in both sectors, so that the net impact on both agricultural and non-agricultural production is negative—in other words, the village's comparative advantage could be the export of labor. This is especially likely to be the case if labor productivity in the region surrounding the village is high relative to in the village, for example, due to greater agglomeration or human capital externalities. Village production could also fall if effective transportation costs are reduced more for labor than for certain goods.

There are, of course, many other ways a road can affect village production. There may be increases in demand for local non-tradable goods if any of the changes above cause increases in income. Improved access to capital could raise investment in productive activities; alternately, access to better savings options could reduce local investment. Or improved information alone could shift prices and investments.

All of these effects will be mitigated by factors that continue to inhibit factor price equalization. For instance, few people in these villages will own vehicles; they will rely on transportation services offered by the market. But if villages have few exports, they may generate so little demand for transport that vehicle operators would be willing to pay the fixed cost to get to the village. Put differently, rural workers and firms may continue to face high effective transportation costs even after road construction.

III 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 road access to unconnected villages across India. 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 greatest number of locations to the external road network at the lowest possible price, routes terminating in villages were prioritized over routes passing through villages and on to larger roads.

Importantly for this paper, the national program guidelines prioritized larger villages according to arbitrary thresholds based on the 2001 Population Census. The guidelines originally aimed to connect all villages with populations greater than 1,000 by 2003, all villages with population greater than 500 by 2007, and villages with population over 250 after that.[4] The thresholds were lower in desert and tribal areas, as well as hilly states and districts affected by left-wing extremism. These rules were to be applied on a state-by-state basis, meaning that states that had connected all larger villages could proceed to 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 prioritized village. Groups of villages within 500 m of each other could combine their populations. 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 could also influence allocation. Different states used different thresholds; for instance, states with few unconnected villages with over 1,000 people used the 500-person threshold immediately.

[4]The unit of targeting in the PMGSY is the habitation, defined as a cluster of population whose location does not change over time. Revenue villages, which are used by the Economic and Population Censuses, are comprised of one or more habitations (National Rural Roads Development Agency, 2005). In this paper, we aggregate all data to the level of the revenue village.

Some states did not comply with the threshold guidelines at all. We identified complying states based on meetings with officials at the National Rural Roads Development Agency, which was the federal body overseeing the program (see Section V for details).

Although funded and overseen by the federal Ministry of Rural Development, responsibility for program implementation was delegated to state governments. Funding came 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 almost $40 billion.[5]

IV Data

To take advantage of village-level variation in road construction, we combine village-level administrative data from the PMGSY program with multiple external datasets, including data covering every firm and household in rural India. The core dataset combining multiple rounds of the population and economic censuses comes from the Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG), Version 1.0.[6] This section gives an overview of the data sources, collection process and variable definitions; additional details are provided in Appendix B.

Identities of connected villages and completion dates come from the official PMGSY website (http://omms.nic.in), which we scraped in January 2015. Household microdata comes from the Socioeconomic and Caste Census (SECC) of 2012, which describes every household and individual in India. This dataset was collected by the Government of India to determine eligibility for social programs. It was made publicly available on the internet in a combination of formats; we scraped and processed over two million files covering 825 million rural individ-

[5]Source: PMGSY administrative data. This figure describes the total amount disbursed by the end of 2015. The cost of a new road to a previously unconnected village in our sample was approximately $150,000.

[6]See Asher and Novosad (2018) for details on its construction. The dataset, along with keys providing merges to the economic and population censuses, can be found at http://www.dartmouth.edu/~novosad/data.html.

uals. After extracting text from the PDF tables, we translated fields from various languages into English, classified occupations into standardized categories and matched locations to the 2011 Population Census based on names. This process yielded a range of variables covering both household characteristics (assets and income) and individual characteristics (age, gender, occupation, caste, etc). Anonymized microdata from the 2002 Below Poverty Line (BPL) Census, an earlier national asset census, was used to construct village level controls.

To generate a proxy measure for consumption, which is not directly surveyed by the SECC, we predict consumption in a survey (IHDS-II, 2011-12) that contains the same asset, income, and land data as the SECC but only contains district-level geographic identifiers. We then impute consumption for each individual in sample villages following the small area estimation methodology of Elbers et al. (2003), allowing us to test not only for impacts of roads on mean predicted consumption per capita but also for distributional effects.[7] Appendix B contains additional details of this process, as well as a discussion of the literature on predicting consumption from such data.

Firm data comes from the Sixth Economic Census (2013). This covers every economic establishment in India, including public and informal establishments, other than those engaged only in crop production, public administration, and defense. It contains detailed information on location (which we match to the 2011 Population Census), employment, industry, and a handful of other firm characteristics, but includes no variables on wages, inputs, or outputs. We trim outliers to eliminate villages where the number of workers in village nonfarm firms is greater than the total number of workers resident in the village according to the 2011 Population Census. Results are not substantively changed by this restriction.

Remote sensing data is used to measure outcomes otherwise unavailable at the village level. Night lights provide a proxy for total village output. As no village-level agricultural

[7] Standard errors for all predicted consumption and poverty regressions are produced using the bootstrapping procedure outlined in Elbers et al. (2003).

production data exists in India, we use two satellite-based vegetative indices (NDVI and EVI) for the primary (kharif) growing season (late May - October) to proxy for village-level agricultural production. To control for differences in non-crop vegetation, our preferred measure is generated by subtracting early cropping season value from the maximum growing season value.[8] We use village boundary polygons purchased from ML Infomap to map gridded remote sensing data to villages and to determine treatment spillover catchment areas.

The 2001 and 2011 Population Censuses (Primary Census Abstract and Village Directory tables) provide village infrastructure, demographics, transportation services and population. The 2011 Population Census also describes the three primary crops grown in each village; we consolidate these into an indicator for whether one out of the three is something other than a cereal (rice, wheat, etc.) or pulse (lentils, chickpeas). Finally, the Population Censuses provide the basis for linking all other datasets together at the village level.

Figure 1 provides a visual representation of the timing 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 study. Road construction is negligible before baseline data collection in 2001, then slowly ramps up to a peak of over 11,000 villages receiving roads annually in 2008 before slowing down slightly.

The analysis sample is restricted to the 11,432 villages that (i) did not have a paved road in 2001; (ii) were matched across all primary datasets; and (iii) had populations within the optimal bandwidth from a treatment threshold. Column 1 of Table 1 reports the average characteristics of the villages in the sample; they are very similar to the average unconnected village in India.[9]

[8] Table A1 shows that this measure is highly correlated with two other proxies for agricultural productivity and per capita predicted consumption at the village level, as well as annual agricultural output at the district level, for both NDVI and EVI. We find similar results when using alternate functional forms. See Appendix B for additional details.

[9] Table A2 shows village-level summary statistics for all villages in the 2001 Population Census, separated into those with and without roads. Villages without paved roads (which comprise nearly half of of all villages) are less populated (1,513 vs 1,930), have fewer public goods (e.g. 25% electrified vs 55%), have less

V Empirical Strategy

The impacts of infrastructure investments are challenging for economists to measure for several reasons. First, the high cost and large potential returns of such investments mean that few policymakers are willing to allow random allocation. Political favoritism, economic potential, and pro-poor targeting would lead infrastructure to be correlated with other government programs and economic growth, biasing naive estimates in an unknown direction. Second, because roads are costly, road construction programs rarely generate large treatment samples. Sample surveys not directly connected with road construction programs are thus unlikely to have a sufficient number of treated and control groups; in contrast, analysis at more aggregate levels is underpowered and faces greater identification concerns. We address these challenges by combining quasirandom variation from program rules with administrative census data georeferenced to the village level.

We obtain causal identification from the guidelines by which villages were prioritized to receive new roads. As previously described, new roads were targeted first to villages with population greater than 1,000, then those greater than 500, and finally greater than 250. While selection into road treatment may have been partly determined by political or economic factors, these factors do not change discontinuously at these population thresholds. As long as these rules were followed to any degree, the likelihood of treatment will discontinuously increase at these population thresholds, making it possible to estimate the effect of new roads using a fuzzy regression discontinuity design.

We pool villages according to the population thresholds that were applied in each state, so the running variable is village population minus the treatment threshold. Very few villages around the 250-person threshold received roads by 2012, so we limit the sample to villages with populations close to 500 and 1,000. Further, only certain states followed the popula-

irrigated agricultural land, and are farther from the nearest urban center than villages with paved roads. The extent to which differences like these are endogenous or causal is the central question of this paper.

tion 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 we define our sample accordingly. Our sample is comprised of villages from the following states, with the population thresholds used in parentheses: Chhattisgarh (500, 1,000), Gujarat (500), Madhya Pradesh (500, 1,000), Maharashtra (500), Orissa (500), and Rajasthan (500).[10]

Under the assumption of continuity of all other village characteristics other than road treatment at the treatment threshold, the fuzzy RD estimator calculates the local average treatment effect (LATE) of receiving a new road for a village with population equal to the threshold. Following the recommendations of Imbens and Lemieux (2008) and Gelman and Imbens (2018), our primary specification uses local linear regression within a given bandwidth of the treatment threshold, and controls for the running variable (village population) on either side of the threshold. We use the following two stage instrumental variables specification:

$$\begin{aligned} Road_{v,j} = \gamma_0 + \gamma_1 1\{pop_{v,j} \geq T\} + \gamma_2(pop_{v,j} - T) + \\ \gamma_3(pop_{v,j} - T) * 1\{pop_{v,j} \geq T\} + \nu X_{v,j} + \mu_j + \upsilon_{v,j} \end{aligned} \tag{1}$$

$$\begin{aligned} Y_{v,j} = \beta_0 + \beta_1 Road_{v,j} + \beta_2(pop_{v,j} - T) + \\ \beta_3(pop_{v,j} - T) * 1\{pop_{v,j} \geq T\} + \zeta X_{v,j} + \eta_j + \epsilon_{v,j}. \end{aligned} \tag{2}$$

$Y_{v,j}$ is the outcome of interest in village v and district-threshold group j, T is the population threshold, $pop_{v,j}$ is baseline village population, $X_{v,j}$ is a vector of village controls measured at

[10] 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.

baseline, and η_j and μ_j are district-threshold fixed effects. Village-level controls include indicators for presence of 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, the share of workers in agriculture, the literacy rate, the share of inhabitants that belong to a scheduled caste, the share of households owning agricultural land, the share of households who are subsistence farmers, and the share of households earning over 250 INR cash per month (approximately 4 USD), all measured at baseline. District-threshold fixed effects are district fixed effects interacted with an indicator variable for whether the village is in the 1,000-person threshold group. $Road_{v,j}$ is an indicator that takes the value one if the village received a new road before the year in which Y is measured, which is 2011, 2012, or 2013 (depending on the data source).[11] Village controls and fixed effects are not necessary for identification but improve the efficiency of the estimation. The coefficient β_1 captures the effect of a new road on the outcome variable. The optimal bandwidth according to the method of Imbens (2018) is 84.[12] We use a triangular kernel which places the most weight on observations close to the threshold, as in Dell (2015). Results are highly similar with different fixed effects or controls, a rectangular kernel, or alternate bandwidths.

Regression discontinuity estimates can be interpreted causally if baseline covariates and the density of the running variable are balanced across the treatment threshold. Table 1 presents the mean values for various village baseline characteristics, including the set of controls that we use in all regressions. While there are average differences between villages above and below the population threshold (Columns 2 and 3), in part because many village characteristics are correlated with size, we find no significant differences when we use the RD specification to

[11] Our primary outcomes are measured in 2011 (Population Census), 2012 (SECC), and 2013 (Economic Census). These were not particularly unusual years for the Indian economy. GDP growth these years was 6.6%, 5.5% and 6.4%, slightly below the 2008-16 average of 7.1%. Rainfall for the main growing season (June-September) was neither particularly high or low: 901, 824 and 937 mm, compared to the 2000-2014 average of 848 mm.

[12] The optimal bandwidth according to the method of Calonico et al. (2014) is 78.

test for discontinuous changes at the threshold. Figure 2 shows the graphical version of the balance test, plotting means of baseline variables in population bins, residual of fixed effects and controls. Baseline village characteristics are continuous at the treatment threshold. Figure 3 shows that the density of the village population distribution is also continuous across the treatment threshold; the McCrary test statistic is -0.01 (s.e. 0.05) (McCrary, 2008).[13]

Figure 4 shows the share of villages that received new roads before 2012 in each population band relative to the treatment threshold; there is a substantial discontinuous increase in the probability of treatment at the threshold. Table 2 presents first stage estimates using the main estimating equation at various bandwidths. Crossing the treatment threshold raises the probability of treatment by 21-22 percentage points; as suggested by the figure, the estimates are very robust to different bandwidth choices.

VI Results

VI.A Main results

We begin by presenting treatment estimates on five indices of the major families of outcomes: (i) transportation services; (ii) sectoral allocation of labor; (iii) employment in nonfarm village firms; (iv) agricultural investment and yields; and (v) income, assets and predicted consumption. We generate these indices to have a mean of 0 and a standard deviation of 1, following Anderson (2008); the variables that make up each index are described in the Data Appendix (Section B7). Table 3 presents the RD estimate of the impact of roads on each outcome, along with unadjusted p-values. The first column shows a large positive effect on the availability of transportation services, and the second shows that roads cause a significant reallocation of labor out of agriculture. We find a smaller positive effect on employment

[13]Note that the density function of habitation population as reported in the internal PMGSY records exhibits notable discontinuities above the treatment thresholds, indicating that some habitation were able to misreport population to gain eligibility (Figure A1). For this reason, we use village population from the 2001 Population Census as the running variable. The Population Census was collected before PMGSY implementation began to scale up, and was done so by a government agency considered to be apolitical and impartial.

growth in village firms (Column 3, p = 0.09), but very small and insignificant positive effects on agricultural yields/investments and on the asset/consumption index (Columns 4 and 5). These indices address concerns about multiple hypothesis testing within families of outcomes. To correct for cross-family multiple hypothesis testing, we follow the step-down procedure of Benjamini and Hochberg (1995), which allows us to reject the null hypothesis of zero effect on both transportation and agricultural labor share with a false discovery rate (adjusted p-value) of 0.075.

Figure 5 presents graphical representations of each regression discontinuity estimate, showing the average of each index as a function of distance from the treatment threshold. The plots show residuals from controls and fixed effects, along with linear estimations on each side of the threshold and 95% confidence intervals. The graphs corroborate the tables, showing significant treatment effects for transportation and labor exit from agriculture, but little clear impact on the firms, agricultural production, and asset/consumption indices.[14] These results broadly summarize the findings of this paper: rural roads lead to increases in transportation services and reallocation of labor out of agriculture, but not to major changes to village firms, agricultural production, or predicted consumption. The rest of this section examines the components of each of these indices to explain the impacts of roads in more detail, and presents results on heterogeneity.

Table 4 shows regression discontinuity estimates of the impact of a new road on an indicator variable for the regular availability at the village level of the five motorized transportation services that are recorded in the 2011 Population Census. A new road causes a statistically significant 12.9 percentage point increase in the availability of public bus services, more than doubling the control group mean of 11.8 percent. The impact on private buses is nearly as large but measured with less precision. Taxis and vans, which are more expensive forms

[14]The table point estimates are larger than the jumps observed in the figures because the tables present fuzzy RD (IV) estimates, while the figures show the reduced form difference at the threshold.

of transportation, do not experience significant growth. Availability of auto-rickshaws, the least expensive private form of motorized transport, increases as well. Given that we are unable to observe transportation costs directly, we interpret these results as evidence that the new roads studied in this paper do meaningfully affect connections between treated villages and outside markets.[15]

Table 5 presents impacts of new roads on occupational choice, the one domain where roads appear to substantially change economic behavior. As 92% of workers in sample villages report their occupation to be either in agriculture or in manual labor, we focus our investigation on these categories. The first two columns show the impact of new roads on the share of workers (aged 21-60) who work in agriculture, and the share who work as manual laborers. New roads cause a 9.2 percentage point reduction in workers in agriculture (representing a 19% decrease from the control group mean) and an 7.2 percentage point increase in workers in (non-agricultural) manual labor.[16] Columns 3 and 4 report estimates on the share of households deriving their primary source of income from cultivation (any crop production) and from manual labor (which includes agricultural labor, non-agricultural labor, and wages from labor on public works projects such as the National Rural Employment Guarantee scheme). We find no significant changes in these measures. While these results may indicate that the workers who respond to a new road by moving out of agriculture are not the primary earners in the household, it is also possible that households may associate primary income source with their identity and thus continue to identify themselves as farmers. Alternately, the inclusion of agricultural labor in the manual labor category for primary income source may help to explain the difference with the occupational results.

[15] This finding is not a given; Raballand et al. (2011) argue that in remote areas of Malawi, willingness to pay for transportation services may be so low that roads may not appreciably improve transportation options.

[16] The SECC does not report manual labor occupations in more detail. Table A3 breaks down the sectoral distribution of non-agricultural manual laborers using the 68th round of the National Sample Survey (2011-12). By far the most common category of manual labor in India is construction, making it a likely sector for many of these former agricultural workers.

Theoretically, we should expect those who exit agriculture in favor of nonfarm labor market opportunities will be those for whom the losses of agricultural income are smallest and the labor market gains are largest. By using individual-level census data, we can examine the distribution of treatment effects across subgroups with different factor endowments. As land is the major input into agricultural production, land endowments may play a major role in determining which workers respond most to a rural road. We first examine the impact of road construction on the landholding distribution in Table A4. We find that a new road does not significantly change the share of households that are landless, own less than 2 acres, own between 2 and 4 acres, or own more than 4 acres of agricultural land. We thus both reject major consolidation of landholdings and treat ex post observed landholdings as a baseline variable upon which to conduct heterogeneity analysis.

Panel A of Table A5 presents our main specification, estimating the effect on agricultural occupation share separately by size of landholdings. We find that movement out of agriculture is strongest for workers in households without land, and that this treatment effect is monotonically decreasing in landholding size.[17] The decrease in agriculture for those with no land (11.7 percentage points) is even larger as a percentage of the control group mean: our estimates suggest that 33% of workers with no land exit agriculture, compared to just 10% in households with more than four acres of land.[18] These results are consistent with recent work finding that land ownership in India can significantly reduce rates of migration and participation in non-agricultural occupations (Fernando, 2018), supporting earlier work by Jayachandran (2006).[19]

[17]We cannot statistically reject equality between any of these estimates. It is also possible that the observed heterogeneity may be affected by the small shift in the distribution of landholdings.

[18]It is important to note that productivity in agriculture 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).

[19]These effects also suggest that new roads may be a progressive investment in that those with the least agricultural wealth (as proxied by landholding) show the largest labor market effects. Jayachandran (2006) shows theoretically that an inelastic agricultural labor supply harms the poor (landless) and acts as

We next examine the heterogeneity of the treatment effect as a function of age and gender (Table A5, Panel B). There are no differential results by age: the point estimate for workers aged 21-40 (a 8.5 percentage point decrease in the share in agriculture) is almost identical to the effect for workers aged 41-60 (a 9.3 percentage point decrease). While the differences are not significantly different, we do find that men are more likely to exit agriculture as compared to women, particularly in the younger cohort (-8.5 percentage point effect for men compared to -2.0 percentage points for women). These estimates could be the result of a male physical advantage in non-agricultural work or attitudes against women's working far away from home (Goldin, 1995). However, as a percentage of the control group mean, the estimates for male and female workers are much closer.

Table 6 presents results on employment in village firms; Panel A shows estimates in logs and Panel B in levels. Because the data source is the Economic Census, these counts include all work in the village, formal and informal, excluding crop production. These results capture economic activity that takes places in the village, in contrast to Table 5, which describes economic activities for village residents even if they take place outside the village. We present estimates for total non-farm village employment (Column 1), as well as employment in the five largest sectors in the sample (livestock, manufacturing, education, retail and forestry), which together account for 79% of non-farm employment. We estimate a 27 percent increase in employment in non-farm firms ($p = 0.09$). While the two largest village sectors (livestock and manufacturing) show similar growth to total employment, the only statistically significant estimate we find is for retail, which we estimate grows 33 percent in response to a new road. In levels, we find no significant results overall or in any sector, with estimates ranging from 2.0 jobs lost in livestock to 2.8 jobs gained in manufacturing.

While the log changes in employment are quite large, the level changes are small because the typical 500- or 1,000-person village has few people engaged in economic activities other

insurance for rich (landed) households, and that landless households will be more likely to migrate.

than crop production. We estimate that a new road on average creates 4.2 new jobs in a village. In contrast, the estimate from Table 5 suggests that 18.5 workers are exiting agriculture in the average village. Taking these point estimates seriously, only 23% of these workers appear to be finding this non-agricultural work in the village, although the standard errors on these estimates are large enough that we cannot reject that all workers leaving agriculture are finding work in village firms. We view this as suggestive evidence that roads are facilitating more access to external labor markets than growth of jobs in village firms. The proportional changes are the largest in the retail sector, suggesting that non-farm employment growth in the village may be more a function of new consumption opportunities (perhaps due to cheaper imports) rather than new productive opportunities. Unfortunately we are aware of no village-level data that would make it possible to directly test for changes in the availability or prices of consumption goods, nor do any village-level censuses ask workers about location of employment.

In Table 7, we examine whether new roads increase investments in agriculture or agricultural yields. Panel A presents the impact of roads on the three different remotely sensed proxies of yield, described in Section IV, generated from two different vegetative indices (NDVI and EVI). Point estimates are very close to zero and the standard errors are tight. In our preferred measure, we estimate an impact of 1.7% higher agricultural yield (equivalent to 0.044 SD) and can rule out a 6.8% or a 0.25 standard deviation increase in yield with 95% confidence.

In Panel B, we examine agricultural input usage. We find no evidence for increases in ownership of mechanized farm or irrigation equipment. There is also no indication of a movement away from subsistence crops, of land extensification, or of changes in the distribution of land ownership. In short, we find no evidence of substantial changes in agricultural production in villages after they receive new roads. Our measures are admittedly incomplete and we are not able to directly measure agricultural output or earnings, but the zero effects for all these different correlates of agricultural production suggest that the structure of agricultural production is not dramatically affected by these new roads.

Finally, in Table 8, we examine the impact of roads on predicted consumption, earnings and assets, which are the best available measures of whether these roads make people appreciably better off in villages. Panel A reports impacts on various measures of predicted consumption and income. We estimate that roads cause a statistically insignificant 2% increase in predicted consumption; we can rule out a 10% increase with 95% confidence. As explained in Appendix B, our predicted consumption measure is a weighted sum of various assets and other measures of economic well-being.[20] To verify that our null result is not the outcome of offsetting positive and negative results, we estimate impacts on each measure (aggregated to village-level shares); Table A7 shows that all are close to zero and there is only one estimate with a p-value below 0.05 (plastic roof, p = 0.02). Given that we run these regressions for 28 variables, this is likely to be spurious. Because we can calculate the consumption measure for every individual in every village, we can further estimate changes at any percentile of the village predicted consumption distribution. Figure 6 shows RD estimates at every ventile of the within-village predicted consumption distribution; effects are weakly more positive at the top of the distribution, but very small and insignificant everywhere. Table A8 separates predicted consumption estimates by education and occupation of the household head; there are no significant gains in any of the categories.[21]

Log night light intensity at the village level (Column 3) provides an alternative proxy for GDP per capita; we again find a point estimate very close to zero. Henderson et al. (2011) estimate a robust elasticity of .3 when regressing log GDP per capita on log night lights per area. Taking this seriously, we would need an estimate of 0.33 to conclude that rural roads

[20]Table A6 presents the "first stage" weights given to each measure, taken from regressions of consumption on these variables in the IHDS. These look very reasonable, with most expensive items having the largest coefficients, such as four-wheeled vehicle (85,686 INR) and refrigerator (29,477 INR).

[21]Note that we measure occupation of the household head in 2012, so some share of the household heads working for wages may be doing so as a result of the treatment.

cause a ten percent increase in GDP per capita—our point estimate is one tenth of that. Finally, Column 4 shows estimates on the share of households in the village whose primary earner makes more than 5,000 rupees (approximately $100) per month.[22] Once again, we find no statistically or economically significant effect; the coefficient even smaller than that for predicted consumption.

Panel B of Table 8 estimates the impact of new roads on asset ownership. The normalized asset index suggests a small and statistically insignificant 0.11 standard deviation increase in assets. The remaining columns show small and insignificant estimates on ownership of the assets that make up the index. The evidence suggests that rural roads do not greatly increase earnings, assets, or consumption, even for relatively inexpensive assets such as mobile phones.

To summarize, new roads do not appear to substantially change either the aggregate economy or predicted consumption in connected villages. We do observe a large shift of workers out of agricultural work and into wage work, but this occupational change does not lead to economically meaningful changes in income or predicted consumption. The average treated village has had a road for 4 years at the time of measurement in 2012, and a quarter for 6 years or more. Given the small positive point estimates on the asset/consumption and agricultural investment indices, it is possible that long-run effects are larger. But the results do not paint a picture of villages poised to reap large benefits from improved transportation infrastructure in the short run.

VI.B Robustness

In this section we examine the robustness of our results to alternative specifications and explanations.

First, as a placebo exercise, we estimate the first stage and reduced form estimation on the family indices for the set of states that did not follow guidelines regarding the population el-

[22] As noted in Section B, the SECC reports income only in three bins and only for the highest earner of the household, so we do not have a more granular measure.

igibility threshold. If villages above the PMGSY thresholds are changing in ways other than through eligibility for roads, we would expect to find similar reduced form effects in these placebo villages as well. Specifically, we include villages close to the two population thresholds in states that built many roads but did not follow the rules at all (Andhra Pradesh, Assam, Bihar, Jharkhand, Karnataka, Uttar Pradesh and Uttarakhand), and villages close to the 1,000 threshold in states that used only the 500-person threshold (Gujarat, Maharashtra, Orissa and Rajasthan). Table A9 presents the estimates. There is no evidence of either a first stage or reduced form effect on any outcomes in the placebo sample, suggesting that our primary estimates can indeed be interpreted as resulting from new roads.

In Table A10, we present the five family index results for bandwidths from 60 to 100, for both triangular and rectangular kernels. The results are consistent with the those in our main specification (Table 3).

If migration is correlated with individual or household characteristics, as some studies have found (Bryan et al., 2014; Morten and Oliveira, 2018), then compositional changes in village population could bias treatment estimates. In Table A11, we examine three proxies for permanent migration.[23] First we test for impacts on village population in 2011 (Panel A). We find no evidence for significant impacts on total population, either in logs or levels. The limitation of population growth as an outcome is that any impacts on net migration could be offset by changes to fertility and mortality. But such offsetting effects would cause changes in village demographics, which we can estimate in the comprehensive census data. In Panels B and C, we show that roads cause no changes to the age distribution or gender ratios in any age cohort. Taken together, these three pieces of evidence suggest that new roads do not lead to major changes in out-migration.[24] The absence of an impact on migration also

[23]Short-term migrants and commuters are considered resident in the village, and thus covered in both the Population Censuses and the SECC.

[24]This difference with Morten and Oliveira (2018) may be due to the differences between rural feeder roads and highways. The construction of a paved rural road is unlikely to significantly change the one-time cost of permanent migration relative to the lifetime benefits, in contrast to the major changes induced by

allows us to interpret the observed sectoral reallocation of labor as the result of changes in occupational choice rather than compositional effects due to selective migration.

Table A12 addresses the possibility that the workforce has changed, which would make it difficult to interpret changes in the share of workers in agriculture or non-agricultural wage work. The table shows that roads do not affect the share of adults who are either not working or who are in occupations that we are unable to classify, suggesting that this potential bias is not important.

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 population 1,000: the Total Sanitation Campaign (Spears, 2015), which attempted to reduce open defecation through toilet construction and advocacy. It is unlikely that this program is spuriously driving our results for two reasons. First, there is little theoretical reason to believe that investments in sanitation could drive large increases in transportation services or reallocation of labor away from agriculture. Second, in Table A13 we present regression discontinuity estimates of the impact of road prioritization on four measures of sanitation. We find no evidence that being above the population threshold is associated either with open defecation or any measure of access to toilets, suggesting that there is no discontinuity in the implementation of the program that might affect our results.

Finally, we consider the possibility that roads have spillover effects on nearby villages; if so, our estimates of direct effects could be biased either upwards or downwards relative to the total effects of new road provision. To do so, we examine outcomes in villages within a 5 km radius of villages in the main sample, using the standard regression discontinuity specification. Table A14 presents results of these regressions for the five outcome indices. We also test for an impact on unemployment in order to test the hypothesis that the reallocation of labor out of agriculture may be coming at the expense of jobs held by those living nearby. We find

highway construction.

no evidence of spillovers, and can reject equality with the main point estimates on the transportation and agricultural occupation measures. It is an open question whether rural road provision has spillover effects in nearby urban labor markets, but our identification strategy does not allow us to answer this question, as every town is surrounded by many villages, few of which are near our population thresholds. Further, PMGSY villages tend to be small and relatively remote, making spillovers onto regional labor markets even harder to detect.

VII Conclusion

Many of the world's poorest live in places that are not well connected to outside markets. The resulting high transportation costs potentially inhibit gains from the division of labor, specialization, and economies of scale.

In this paper we estimate the economic impacts of the Pradhan Mantri Gram Sadak Yojana, a large-scale program in India that has aimed to provide universal access to paved "all-weather" roads in rural India. We find that the effects of this program on village economies are smaller than those anticipated by policy-makers or suggested by the existing body of research on roads. Four years after road completion, we find few impacts on assets, agricultural investments, or predicted consumption, and only small changes to employment in village firms. We do find that new paved roads lead to increased transportation services and a large reallocation of labor out of agriculture.

Roads are costly investments: the cost of connecting each additional village to the paved road network is approximately \$150,000. A back of the envelope calculation from our estimates suggests that the average village (with 696 residents) gains an additional \$5.67 of consumption per year on a base of \$267, or \$3945 per village per year.[25] Even if we use the upper bound of the confidence interval, we find small effects relative to the cost of roads. Worse yet, the villages in India still lacking paved roads are less populated and more remote

[25] Maintenance costs of paved rural roads are very similar to gravel roads in India (Indian Roads Congress, 2002), and so do not affect our calculation.

than those in our sample, suggesting that impacts for future rural road investments are likely to be even smaller.

Our estimates admittedly do not capture every dimension of welfare. The long run effects of roads may be larger than the short to medium term estimates here. Access to employment outside the village may play an important insurance role, and improved access to external health and education services may be valuable; indeed, we find elsewhere that rural roads cause increases in educational attainment (Adukia et al., 2017). We also do not estimate the impact of spillovers into larger regional markets. Additional research into these other potential impacts would be valuable, as would analysis of how market access interacts with complementary policies and investments.

Both researchers and policymakers have claimed that roads have the potential to revolutionize economic opportunities in remote, rural areas. This paper suggests that even in a fast growing economy such as India in the 2000s, rural growth is constrained by more than the poor state of transportation infrastructure. Instead of facilitating growth on village farms and firms, the main economic benefit of rural transportation infrastructure may be the connection of rural workers to new employment opportunities.

References

Adukia, Anjali, Sam Asher, and Paul Novosad, "Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction," 2017. Working paper.

Aggarwal, Shilpa, "Do Rural Roads Create Pathways Out of Poverty? Evidence from India," *Journal of Development Economics*, 2018, *133*, 375–395.

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

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

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: a Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2008, *103* (484), 1481–1495.

Asher, Sam and Paul Novosad, "The Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG)," 2018. Working paper.

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs," 2017. Working paper.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," 2012. NBER Working Paper No. 17897.

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang, "Roads , Railways and Decentralization of Chinese Cities," *Review of Economics and Statistics*, 2011, pp. 1–42.

Bedi, Tara, Aline Coudouel, and Kenneth Simler, eds, *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*, Washington, D.C.: The World Bank, 2007.

Benjamini, Yoav and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, *57* (1), 289–300.

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*, 2013, *22* (AERC Supplement 2).

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

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

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

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

——, **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), 1817–1851.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli, "Agricultural Productivity and Structural Transformation. Evidence from Brazil," *The American Economic Review*, 2016, *106* (6), 1320–1365.

Calonico, Sebastian, Matias D. 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. Working Paper.

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.

Dell, Melissa, "Trafficking Networks and the Mexican Drug War," *American Economic Review*, 2015, *105* (6), 1738–1779.

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.

Donaldson, Dave, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

—— **and Richard Hornbeck**, "Railroads and American Economic Growth: A "Market Access" Approach," *Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

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

Elbers, Chris, Jean Lanjouw, and Peter Lanjouw, "Micro-level Estimation of Poverty and Inequality," *Econometrica*, 2003, *71* (1), 355–364.

Faber, Benjamin, "Trade Integration, Market Size, and Inudstrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

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," 2018. Working paper.

Gelman, Andrew and Guido Imbens, "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 2018, pp. 1–10. NBER Working Paper No. 20405.

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*, 2016, *126* (591), 317–357.

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.

Gonzalez-Navarro, Marco and Climent Quintana-Domeque, "Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects," *Review of Economics and Statistics*, 2016, *98* (2), 254–267.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, "A Bright Idea for Mesuring Economic Growth," *American Economic Review*, 2011, *101* (3), 194–199.

Hentschel, Jesko, Jean Olson Lanjouw, Peter Lanjouw, and Javier Poggi, "Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador," *The World Bank Economic Review*, 2000, *14* (1), 147–165.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, "Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices," *Remote Sensing of Environment*, 2002, *83* (1-2), 195–213.

Imbens, Guido, "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 2018, *79* (July), 933–959.

—— **and Thomas Lemieux**, "Regression Discontinuity Designs: a Guide to Practice," *Journal of Econometrics*, 2008, *142* (2), 615–635.

Indian Roads Congress, "Rural Roads Manual," Technical Report, New Delhi 2002.

Jacoby, Hanan and Bart Minten, "On Measuring the Benefits of Lower Transport Costs," *Journal of Development Economics*, 2009, *89*, 28–38.

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

Jayachandran, Seema, "Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries," *Journal of Political Economy*, 2006, *114* (3), 538–575.

Khandker, Shaidur R. and Gayatri B. Koolwal, "Estimating the Long-term Impacts of Rural Roads: A Dynamic Panel Approach," 2011. World Bank Policy Research Paper No. 5867.

——, **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.

Kouadio, Louis, Nathaniel K. Newlands, Andrew Davidson, Yinsuo Zhang, and Aston Chipanshi, "Assessing the Performance of MODIS NDVI and EVI for Seasonal Crop Yield Forecasting at the Ecodistrict Scale," *Remote Sensing*, 2014, *6* (10), 10193–10214.

Labus, M. P., G. A. Nielsen, R. L. Lawrence, R. Engel, and D. S. Long, "Wheat Yield Estimates Using Multi-temporal NDVI Satellite Imagery," *International Journal of Remote Sensing*, 2002, *23* (20), 4169–4180.

Lehne, Jonathan, Jacob Shapiro, and Oliver Vanden Eynde, "Building Connections: Political Corruption and Road Construction in India," *Journal of Development Economics*, 2018, *131*, 62–78.

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.

McCrary, Justin, "Manipulation of the Running Variable in the Regression Discontinuity Design: a Density Test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

McKenzie, David J., "Measuring Inequality with Asset Indicators," *Journal of Population Economics*, 2005, *18* (2), 229–260.

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.

Mkhabela, Manasah S., Milton S. Mkhabela, and Nkosazana N. Mashinini, "Early Maize Yield Forecasting in the Four Agro-ecological Regions of Swaziland Using NDVI Data Derived From NOAA's-AVHRR," *Agricultural and Forest Meteorology*, 2005, *129* (1-2), 1–9.

Morten, Melanie and Jaqueline Oliveira, "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," 2018. NBER Working Paper No. 22158.

Mu, Ren and Dominique van de Walle, "Rural roads and local market development in Vietnam," *Journal of Development Studies*, 2011, *47* (5), 709–734.

Mukherjee, Mukta, "Do Better Roads Increase School Enrollment? Evidence from a Unique Road Policy in India," 2012. Working paper.

Narayanan, K. R., "Address by President of India to Parliament," 2001.

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.

Rasmussen, M. S., "Operational Yield Forecast Using AVHRR NDVI Data: Reduction of Environmental and Inter-annual Variability," *International Journal of Remote Sensing*, 1997, *18* (5), 1059–1077.

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.

Roberts, Peter, K. C. Shyam, and Cordula Rastogi, "Rural Access Index: A Key Development Indicator," Technical Report, The World Bank 2006.

Rojas, O., "Operational Maize Yield Model Development and Validation Based on Remote Sensing and Agro-meteorological Data in Kenya," *International Journal of Remote Sensing*, sep 2007, *28* (17), 3775–3793.

Selvaraju, R., "Impact of El Niño-southern Oscillation on Indian Foodgrain Production," *International Journal of Climatology*, 2003, *23* (2), 187–206.

Shamdasani, Yogita, "Rural Road Infrastructure & Agricultural Production: Evidence from India," 2018. Working paper.

Shrestha, Slesh A., "Roads, Participation in Markets, and Benefits to Agricultural Households: Evidence from the Topography-based Highway Network in Nepal," 2017. Working paper.

Skinner, Jonathan, "A Superior Measure of Consumption from the Panel Study of Income Dynamics," *Economics Letters*, 1987, *23* (2), 213–216.

Son, N. T., C. F. Chen, C. R. Chen, V. Q. Minh, and N. H. Trung, "A Comparative Analysis of Multitemporal MODIS EVI and NDVI Data for Large-Scale Rice Yield

Estimation," *Agricultural and Forest Meteorology*, 2014, *197*, 52–64.

Sotelo, Sebastian, "Domestic Trade Frictions and Agriculture," 2018. Working paper.

Spears, Dean, "Effects of Sanitation on Early-life Health: Evidence From a Governance Incentive in Rural India," 2015. Working paper.

Storeygard, Adam, "Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa," *The Review of Economic Studies*, 2016, *83* (3), 1263–1295.

Tange, Ole, "GNU Parallel: The Command-line Power Tool," *The USENIX Magazine*, 2011, *36* (1), 42–47.

The World Bank Group, "Measuring Rural Access Using New Technologies," Technical Report 2016.

Wantchekon, Leonard, Marko Klašnja, and Natalija Novta, "Education and Human Capital Externalities: Evidence from Benin," *The Quarterly Journal of Economics*, 2015, *130* (2), 703–757.

Wardlow, Brian D. and Stephen L. Egbert, "A Comparison of MODIS 250-m EVI and NDVI Data For Crop Mapping: A Case Study for Southwest Kansas," *International Journal of Remote Sensing*, 2010, *31* (3), 805–830.

Young, Alwyn, "The African Growth Miracle," *Journal of Political Economy*, 2012, *120* (4), 696–739.

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.

Table 1: Summary statistics and balance

Variable	Full sample	Below threshold	Over threshold	Difference of means	p-value on difference	RD estimate	p-value on RD estimate
Primary school	0.956	0.951	0.961	0.01	0.01	-0.017	0.62
Medical center	0.163	0.153	0.175	0.02	0.00	-0.093	0.14
Electrified	0.425	0.408	0.443	0.04	0.00	-0.012	0.88
Distance from nearest town (km)	26.782	26.811	26.749	-0.06	0.88	-3.956	0.26
Land irrigated (share)	0.281	0.275	0.288	0.01	0.01	-0.017	0.71
Ln land area	5.160	5.107	5.220	0.11	0.00	-0.091	0.39
Literate (share)	0.456	0.452	0.460	0.01	0.01	-0.013	0.58
Scheduled caste (share)	0.142	0.141	0.144	0.00	0.28	-0.025	0.42
Land ownership (share)	0.736	0.737	0.734	-0.00	0.48	0.006	0.87
Subsistence ag (share)	0.440	0.443	0.436	-0.01	0.18	0.025	0.56
HH income > INR 250 (share)	0.757	0.755	0.759	0.00	0.43	-0.027	0.55
N	11432	6018	5414				

Notes: The table presents mean values for village characteristics, measured in the baseline period. The first eight variables come from the 2001 Population Census, while the final three (below the line) come from the 2002 BPL Census. Columns 1-3 show the unconditional means for all villages, villages below the treatment threshold, and villages above the treatment threshold, respectively. Column 4 shows the difference of means across Columns 2 and 3, and Column 5 shows the p-value for the difference of means. Column 6 shows the regression discontinuity estimate, following the main estimating equation, of the effect of being above the treatment threshold on the baseline variable (with the outcome variable omitted from the set of controls), and Column 7 is the p-value for this estimate, using heteroskedasticity robust standard errors. An optimal bandwidth of $\pm$ 84 around the population thresholds has been used to define the sample of villages (see text for details).

Table 2: First stage: effect of road prioritization on road treatment

	±60	±70	±80	±90	±100	±110
Road priority	0.224	0.221	0.217	0.214	0.213	0.215
	(0.019)	(0.018)	(0.017)	(0.016)	(0.015)	(0.014)
F Statistic	132.8	150.9	167.2	181.4	200.4	223.6
N	8339	9720	11099	12457	13871	15238
R2	0.30	0.30	0.30	0.29	0.29	0.29

Notes: This table presents first stage estimates of the effect of being above the treatment threshold on a village's probability of treatment. The dependent variable is a indicator 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 60 of the population threshold (440-560 for the low threshold and 940-1060 for the high threshold). The second through sixth columns expand the sample to include villages within 70, 80, 90, 100 and 110 of the population thresholds. The specification includes baseline village-level controls for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates.

Table 3: Impact of new road on indices of major outcomes

	Transportation	Ag occupation	Firms	Ag production	Consumption
New road	0.410	-0.341	0.269	0.082	0.033
	(0.187)	(0.160)	(0.157)	(0.124)	(0.137)
p-value	0.03	0.03	0.09	0.51	0.81
N	11432	11432	10678	11432	11432
R2	0.18	0.28	0.30	0.53	0.50

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of a new road on indices of the major outcomes in each of the five families of outcomes: transportation, occupation, firms, agriculture, and welfare. See Section B for details of index construction. The specification includes baseline village-level controls for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates.

Table 4: Impact of new road on transportation

	Gov Bus	Private Bus	Taxi	Van	Autorickshaw
New road	0.129	0.114	0.007	-0.021	0.073
	(0.055)	(0.074)	(0.048)	(0.055)	(0.043)
Control group mean	0.118	0.205	0.069	0.156	0.055
N	11432	11432	11432	11432	11432
R2	0.30	0.10	0.09	0.44	0.26

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of new road construction on regularly available transportation services. Columns 1-5 estimate the impact on the five categories of motorized transport recorded in the 2011 Population Census: government buses, private buses, taxis, vans and autorickshaws. 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 for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates.

Table 5: Impact of new road on occupation and income source

	Occupation		Household Income Source	
	Agriculture	Manual Labor	Agriculture	Manual Labor
New road	-0.092	0.072	-0.030	-0.011
	(0.043)	(0.043)	(0.044)	(0.044)
Control group mean	0.476	0.448	0.418	0.507
N	11432	11432	11432	11432
R2	0.28	0.26	0.31	0.28

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of new road construction on occupational choice and household source of income. Column 1 estimates the impact on the share of workers in agriculture. Column 2 estimates the effect on the share of workers in manual labor (excluding agriculture). Columns 3 and 4 provide estimates of the impact of a new road on the share of households reporting cultivation and manual labor as the primary source of income. 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 for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates.

Table 6: Impact of new road on firms

Panel A. Log employment growth (by sector)

	Total	Livestock	Manufacturing	Education	Retail	Forestry
New road	0.273	0.252	0.260	0.198	0.333	-0.107
	(0.159)	(0.188)	(0.193)	(0.143)	(0.154)	(0.107)
N	10678	10678	10678	10678	10678	10678
R2	0.30	0.42	0.23	0.18	0.23	0.35

Panel B. Level employment growth (by sector)

	Total	Livestock	Manufacturing	Education	Retail	Forestry
New road	4.219	-1.962	2.802	0.686	1.831	2.381
	(7.596)	(3.364)	(3.794)	(0.973)	(1.534)	(4.002)
Mean employment (level)	32.1	6.9	5.8	5.1	4.5	2.8
N	10678	10678	10678	10678	10678	10678
R2	0.30	0.46	0.18	0.13	0.17	0.36

Notes: This table presents IV discontinuity estimates from the main estimating equation of the effect of new road construction on employment in in-village nonfarm firms. Panel A examines the impact on log employment in all nonfarm firms (Column 1) and in the five largest sectors in our sample: livestock, manufacturing, education, retail, and forestry. Panel B presents estimates for the same regressions, instead specifying the level of employment as the dependent variable. The specification includes baseline village-level controls for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates.

Table 7: Impact of new road on agricultural outcomes

Panel A. Agricultural yields (log)

	NDVI			EVI		
	Max - June	Cumulative	Max	Max - June	Cumulative	Max
New road	0.017	0.000	0.011	0.035	-0.001	0.022
	(0.026)	(0.013)	(0.014)	(0.033)	(0.015)	(0.019)
Control group mean	8.236	10.507	8.801	7.957	10.159	8.470
Control group SD	0.273	0.218	0.181	0.336	0.222	0.195
N	11333	11332	11333	11333	11332	11333
R2	0.71	0.89	0.82	0.72	0.86	0.72

Panel B. Agricultural inputs

	Mechanized Farm Equipment	Irrigation Equipment	Land Ownership	Non-cereal/pulse crop	Cultivated land (log)
New road	-0.004	0.002	0.006	0.030	0.040
	(0.012)	(0.028)	(0.036)	(0.073)	(0.081)
Control group mean	0.040	0.141	0.570	0.393	5.046
N	11431	11432	11432	8272	11165
R2	0.26	0.43	0.39	0.45	0.73

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of new road construction on village-level measures of agricultural activity. Panel A examines whether roads have an impact on agricultural production, presenting results for three different NDVI-based proxies for agricultural yields. For each regression, the outcome mean and SD for the control group (villages with population below the threshold) is also shown. Panel B examines the impact of roads on agricultural inputs. Column 1 estimates the impact on the share of households owning mechanized farm equipment, Column 2 the share of households owning irrigation equipment, Column 3 the share of households owning agricultural land, Column 4 an indicator for whether a village lists a non-cereal and non-pulse crop as one of its three major crops, and Column 5 the log total cultivated land (sample restricted to villages reporting non-zero values). For each regression, the outcome mean for the control group (villages with population below the threshold) is also shown. Heteroskedasticity robust standard errors are reported below point estimates.

Table 8: Impact of new road on predicted consumption, earnings and assets

Panel A. Consumption and earnings

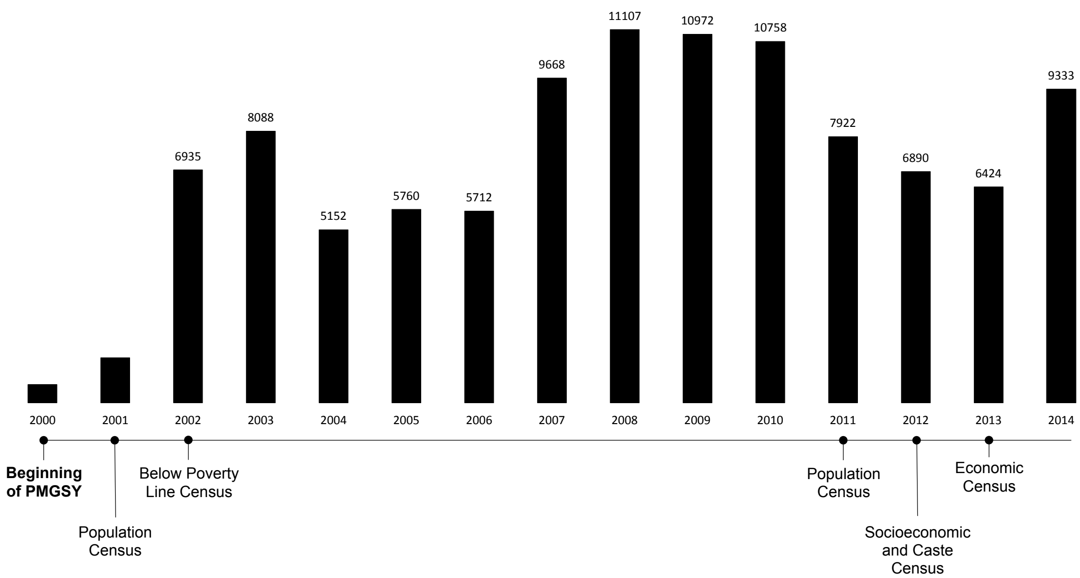
	Consumption per capita (log)	Poverty rate	Night lights (log)	Share of HH earning $\geq$ INR 5k
New road	0.022	-0.010	0.033	-0.001
	(0.038)	(0.042)	(0.165)	(0.032)
Control group mean	9.571	0.282	1.444	0.147
N	11432	11432	11102	11432
R2	0.41	0.30	0.66	0.25

Panel B. Asset ownership

	Asset index	Solid house	Refrigerator	Vehicle	Phone
New road	0.107	0.033	0.005	-0.001	0.033
	(0.132)	(0.029)	(0.013)	(0.023)	(0.041)
Control group mean	-0.015	0.222	0.036	0.140	0.443
N	11432	11432	11432	11432	11432
R2	0.52	0.67	0.26	0.38	0.48

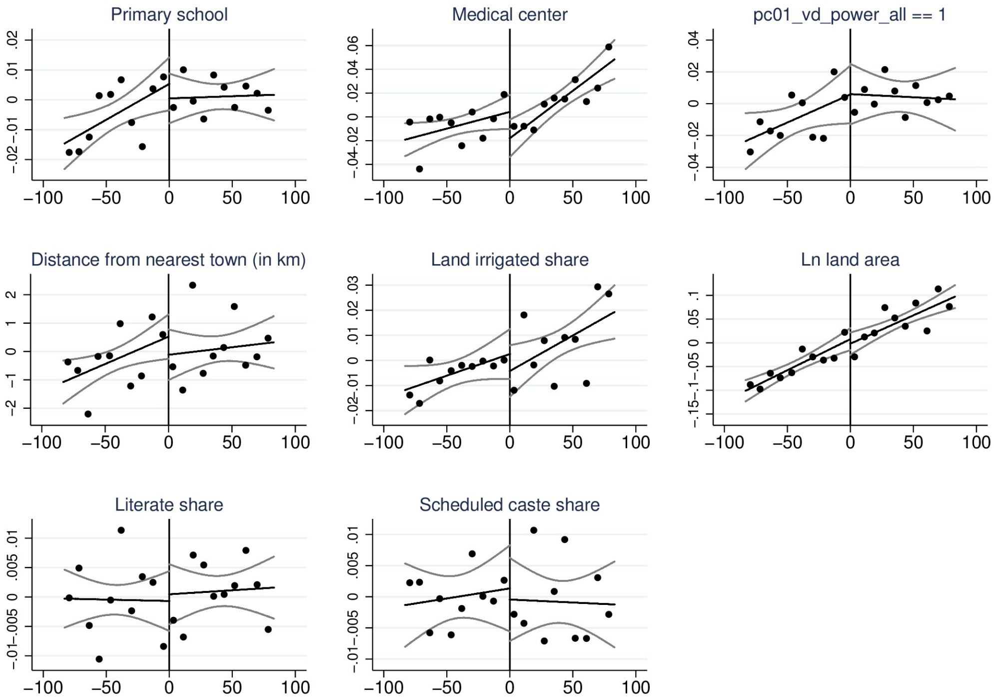
Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of new road construction on various measures of welfare. Panel A examines the impact on measures of predicted consumption and earnings. We use imputed log consumption per capita (outcome for Column 1, see Data Appendix for details of variable construction) and share of the population below the poverty line (Column 2). The dependent variable for Column 3 is the log of mean total night light luminosity in 2011-13, with an extra control for log light at baseline in 2001. The dependent variable for Column 4 is the share of households whose highest earning member earns more than INR 5000 per month. Panel B examines the impact on asset ownership as measured in the 2012 SECC. The dependent variable for Column 1 is the village-level average of the primary component of indicator variables for all household assets measured in the SECC. The remaining four columns present estimates for the impact on the share of households in the village that own each of these assets. The specification includes baseline village-level controls for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates for all estimates except for consumption and poverty, which report bootstrapped standard errors as described in the data appendix.

Figure 1: Timeline of data sources, with count of roads completed under PMGSY by year



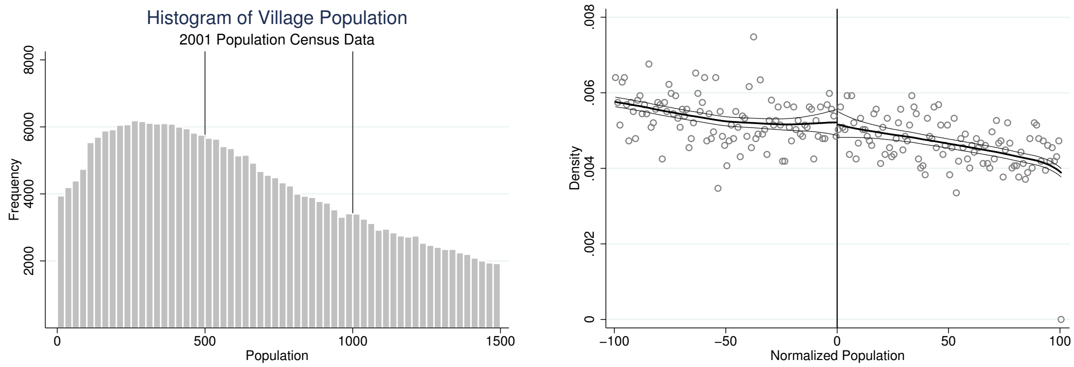
Notes: The figure shows the timing of the population, economic and poverty censuses of India used as principal data sources. Note that while the Socioeconomic and Caste Census (SECC) was intended to be conducted exclusively in 2011, and it is often referred to with this year, it was conducted primarily in 2012. The bar graph above represents the roads completed under PMGSY roads in each year. Exact counts are also listed.

Figure 2: Balance of baseline village characteristics



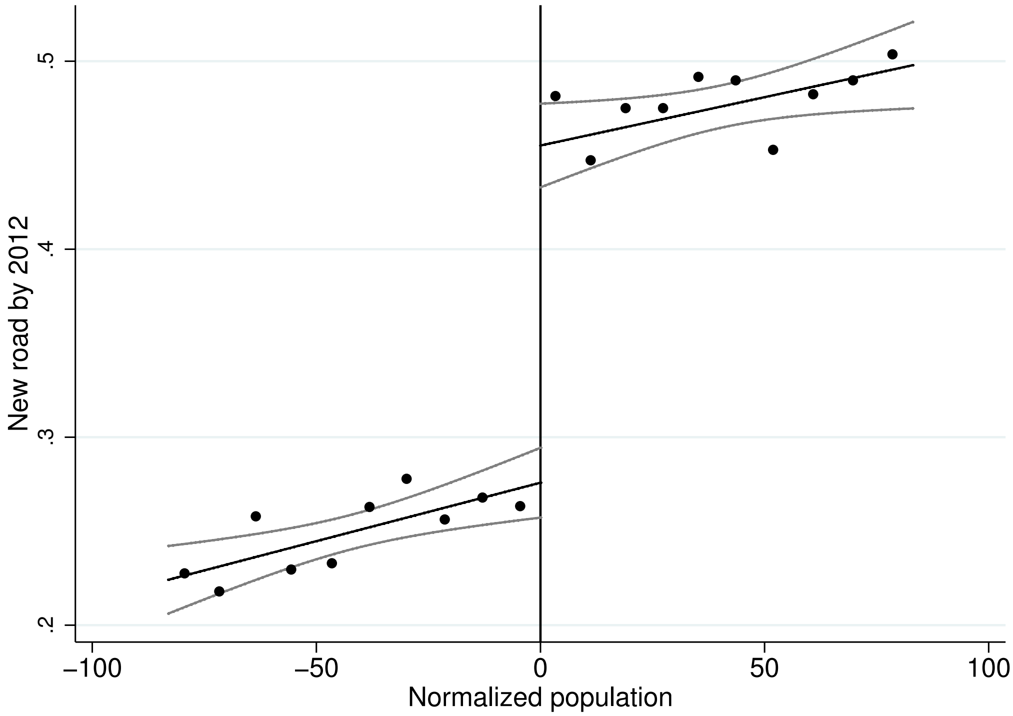
Notes: The figure plots residualized baseline village characteristics (after controlling for all variables in the main specification other than population) over normalized village population in the 2001 Population Census. Points to the right of zero are above treatment thresholds, while points to the left of zero are below treatment thresholds. Each point represents approximately 570 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 (84) of the threshold (see text for details).

Figure 3: Distribution of running variable



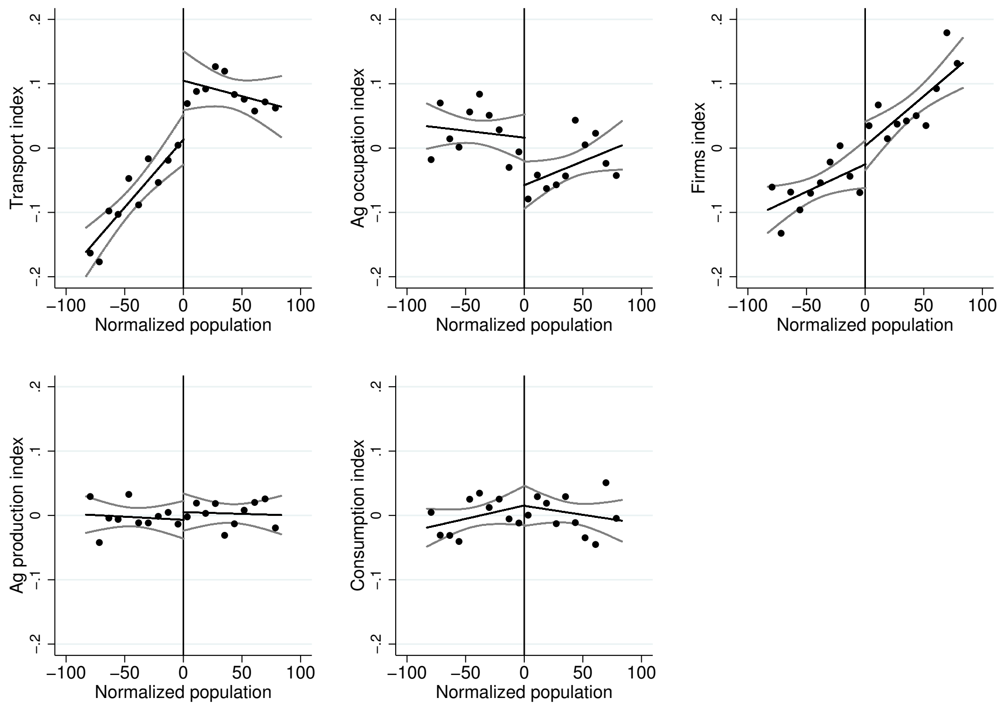
Notes: The figure shows the distribution of village population around the population thresholds. The left panel is a histogram of village population as recorded in the 2001 Population Census. The vertical lines show the program eligibility thresholds used in this paper, at 500 and 1,000. The right panel uses the normalized village population (reported population minus the threshold, either 500 or 1,000). 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 -0.01, with a standard error of 0.05.

Figure 4: First stage: effect of road prioritization on probability of new road by 2012



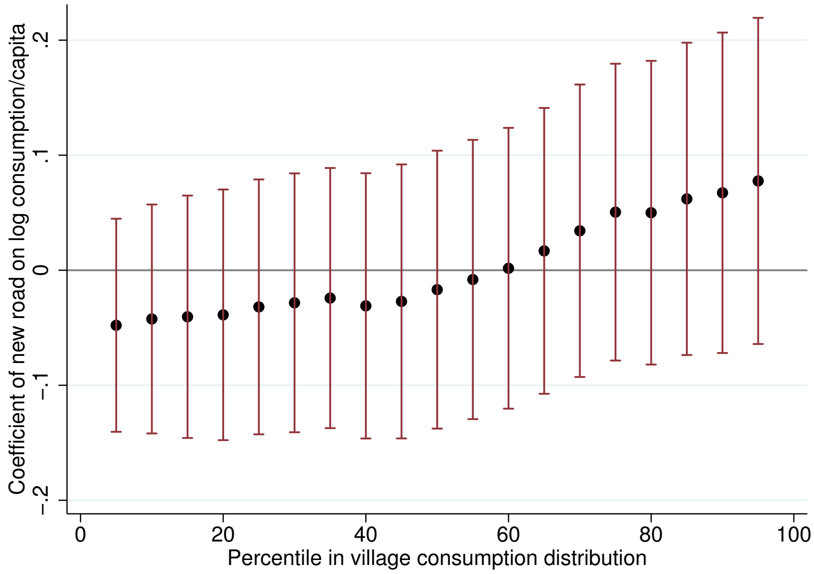
Notes: The figure plots the probability of getting a new road under PMGSY by 2012 against village population in the 2001 Population Census. The sample consists of villages that did not have a paved road at baseline, with baseline population within an optimal bandwidth (84) of the population thresholds. Populations are normalized by subtracting the threshold population.

Figure 5: Reduced form: effect of road prioritization on indices of major outcomes



Notes: The figure plots the residualized values (after controlling for all variables in the main specification other than population) of the indices of the major outcomes in each of the five families of outcomes (transportation, occupation, firms, agriculture, and welfare) 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 (84) of the population thresholds (see text for details). Population is normalized by subtracting the threshold.

Figure 6: Distributional impacts of new road on predicted consumption



Notes: Each point in the figure shows a regression discontinuity estimate and bootstrapped confidence interval of the impact of a new road on log predicted consumption per capita for individuals at a given percentile in the within-village consumption distribution given on the X axis. For example, the point at $X = 5$ represents the impact of a new road on predicted consumption per capita at the fifth percentile of the village distribution. See Data Appendix for description of bootstrapping.

For Online Publication - Appendix

A Additional figures and tables

Table A1: Correlates of NDVI and EVI proxies for agricultural production

Panel A. NDVI/EVI on village proxies of agricultural productivity

	NDVI				EVI			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Crop suitability (log)	0.017			0.017	0.017			0.017
	(0.002)			(0.002)	(0.002)			(0.002)
Irrigation (share)		0.014		0.009		0.038		0.032
		(0.002)		(0.002)		(0.003)		(0.003)
Consumption (log)			0.028	0.026			0.043	0.036
			(0.002)	(0.002)			(0.003)	(0.003)
N	137336	137336	137336	137336	137336	137336	137336	137336
R2	0.49	0.49	0.49	0.49	0.51	0.51	0.51	0.51

Panel B. NDVI/EVI on district agricultural output

	NDVI				EVI			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Agricultural output	0.056	0.035	0.331	0.233	0.398	0.399	0.235	0.197
	(0.018)	(0.017)	(0.042)	(0.040)	(0.026)	(0.026)	(0.046)	(0.041)
Fixed effects	State	State-Year	District	District, Year	State	State-Year	District	District, Year
N	2124	2124	2124	2124	2124	2124	2124	2124
R2	0.39	0.55	0.74	0.78	0.43	0.51	0.85	0.89

Notes: For validation purposes, our favored log-differenced NDVI and EVI agricultural production proxies are regressed on other likely correlates of yields. Panel A presents village level estimates of these proxies regressed on log crop suitability, share of village land irrigated, and log predicted consumption per capita, all with district fixed effects. Panel B presents district-level regressions of these proxies on the value of agricultural output (log) for the years 2000-2006. See Data Appendix for details. The sample has been restricted to states from the primary specification, where states follow PMGSY population guidelines. Heteroskedasticity robust standard errors are reported below point estimates.

Table A2: Summary statistics, by paved road at baseline

	No Road	Paved Road	Total
Primary school	0.692	0.864	0.783
	(0.462)	(0.342)	(0.412)
Medical center	0.183	0.434	0.316
	(0.387)	(0.496)	(0.465)
Electrified	0.249	0.549	0.405
	(0.432)	(0.498)	(0.491)
Crop land irrigated share	0.344	0.456	0.404
	(0.360)	(0.382)	(0.376)
Literate share	0.431	0.499	0.466
	(0.186)	(0.153)	(0.173)
Scheduled caste share	0.157	0.185	0.171
	(0.213)	(0.193)	(0.203)
Distance from nearest town	28.3	20.0	23.9
(in km)	(29.4)	(20.7)	(25.5)
Population	1513.2	1930.5	1730.8
	(30628.4)	(36167.6)	(33631.6)
Number of villages	282864	308263	591127

Notes: This table presents means and standard deviations of baseline variables and outcomes for all villages in India. 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 A3: Sectoral distribution of non-agricultural manual laborers

	Share of non-agricultural manual laborers in sector
Construction	0.60
Transport	0.07
Retail	0.05
Domestic work	0.05
Building materials	0.04
Other	0.17

Notes: This table shows the share of non-agricultural manual laborers in the five largest industries. The sample is the full rural population in the 68th round of the National Sample Survey (2011-12).

Table A4: Impact of new road on distribution of landholdings

	Landless	0-2 Acres	2-4 Acres	4+ Acres
New road	-0.009	-0.012	-0.007	0.028
	(0.036)	(0.033)	(0.016)	(0.024)
Control group mean	0.434	0.287	0.120	0.159
N	11394	11394	11394	11394
R2	0.39	0.41	0.23	0.47

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of new 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 three columns for households owning agricultural land. 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 for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates.

Table A5: Impact of new road on agricultural labor share by land, age, and gender

Panel A. Impact by household landholding

	Landless	0-2 Acres	2-4 Acres	4+ Acres
New road	-0.117	-0.100	-0.075	-0.063
	(0.047)	(0.052)	(0.054)	(0.053)
Control group mean	0.352	0.514	0.590	0.653
N	11101	10698	10380	9945
R2	0.22	0.18	0.19	0.22

Panel B. Impact by age and gender

	All		Male		Female	
	21-40	41-60	21-40	41-60	21-40	41-60
New road	-0.085	-0.093	-0.085	-0.094	-0.020	-0.044
	(0.045)	(0.045)	(0.045)	(0.044)	(0.056)	(0.061)
Control group mean	0.430	0.578	0.450	0.612	0.268	0.330
N	11421	11379	11410	11369	10781	10184
R2	0.28	0.29	0.28	0.29	0.21	0.24

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of new road construction on occupational choice. The dependent variable in each regression is the share of workers in agriculture, for that specific category. Panel A examines whether treatment effects vary by the size of the household landholding. Column 1 estimates the impact for workers in households without agricultural land, Column 2 for workers in households with greater than 0 acres but but weakly less than two acres, Column 3 for workers in households with more than 2 acres but weakly less than 4 acres, and Column 4 for households with 4 or more acres of land. Panel B examines whether treatment effects vary by age and gender. The first two columns present results for workers aged 21-40 and 41-60. The next two present the same results for males workers only, while the final two present the same results for female workers. 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 for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates.

Table A6: Consumption prediction first stage

	Coefficient	(SE)	p-value
Owns land	9657	(1239)	0.000
Two-wheeled vehicle	34253	(2874)	0.000
Four-wheeled vehicle	85686	(14868)	0.000
Landline phone	24639	(8154)	0.003
Mobile phone	23997	(995)	0.000
Both landline and mobile	31479	(6895)	0.000
HH income 5000 - 10000 INR	10076	(1878)	0.000
HH income 10000+ INR	38933	(4779)	0.000
Refrigerator	29477	(2868)	0.000
Number of rooms in home	3429	(599)	0.000
Grass wall	12808	(3551)	0.000
Mud wall	13372	(3269)	0.000
Plastic wall	19748	(6754)	0.003
Wood wall	9217	(3745)	0.014
Brick wall	23030	(3451)	0.000
GI wall	14184	(4505)	0.002
Stone wall	17065	(4492)	0.000
Concrete wall	22316	(3515)	0.000
Grass roof	-2920	(1770)	0.099
Tile roof	-6508	(1772)	0.000
Slate roof	2316	(3018)	0.443
Plastic roof	6474	(8259)	0.433
GI roof	-3359	(1889)	0.075
Brick roof	-9605	(2387)	0.000
Stone roof	11637	(5121)	0.023
Concrete roof	1432	(2519)	0.570
Owns home	-1334	(5550)	0.810
Kisan credit card	12441	(4584)	0.007
Constant	24538	(6572)	0.000
N = 25279			
R2 = 0.359			

Notes: This table presents estimates from the regression of total household consumption on all economic well-being measures that are used to predict consumption. The sample is all rural households in the IHDS-II, with observations weighted according to sampling weights. No other controls are used.

Table A7: Impact of new road on all predictors of consumption

	Coefficient	(SE)	p-value	N	R2
Owns land	0.006	(0.036)	0.87	11432	0.39
Two-wheeled vehicle	-0.003	(0.021)	0.89	11432	0.35
Four-wheeled vehicle	0.001	(0.007)	0.85	11432	0.22
Landline phone	-0.003	(0.004)	0.41	11432	0.08
Mobile phone	0.045	(0.041)	0.26	11432	0.47
Both landline and mobile	-0.009	(0.005)	0.09	11432	0.06
HH income from 5000 - 10000	-0.007	(0.024)	0.76	11432	0.19
HH income over 10000	0.006	(0.015)	0.68	11432	0.20
Refrigerator	0.005	(0.013)	0.70	11432	0.26
Mean number of rooms in home	0.063	(0.086)	0.46	11432	0.36
Grass wall	0.040	(0.028)	0.16	11432	0.25
Mud wall	-0.054	(0.052)	0.30	11432	0.40
Plastic wall	-0.002	(0.005)	0.63	11432	0.07
Wood wall	0.000	(0.012)	0.98	11432	0.12
Brick wall	0.004	(0.035)	0.91	11432	0.41
GI wall	0.001	(0.004)	0.76	11432	0.05
Stone wall	0.003	(0.030)	0.93	11432	0.14
Concrete wall	-0.005	(0.011)	0.69	11432	0.09
Grass roof	-0.003	(0.041)	0.95	11432	0.43
Tile roof	0.013	(0.045)	0.78	11432	0.60
Slate roof	0.016	(0.024)	0.52	11432	0.28
Plastic roof	-0.024	(0.010)	0.02	11432	0.18
GI roof	0.001	(0.021)	0.97	11432	0.51
Brick roof	-0.001	(0.008)	0.93	11432	0.28
Stone roof	0.015	(0.025)	0.56	11432	0.50
Concrete roof	-0.004	(0.018)	0.81	11432	0.43
Owns home	0.007	(0.008)	0.36	11432	0.11
Kisan credit card	-0.007	(0.017)	0.65	11432	0.35

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of new road construction on village shares of all dummy variables used in the consumption prediction exercise (except for number of rooms, which is the village mean). The specification includes baseline village-level controls for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates for all estimates except for consumption and poverty, which report bootstrapped standard errors as described in the data appendix.

Table A8: Impact of new road on log predicted consumption, by education and occupation

Panel A. Consumption by education level

	No education	Primary or below	Middle school+
New road	-0.017	0.013	0.007
	(0.039)	(0.042)	(0.045)
Control group mean	9.39	9.54	9.75
N	11306	11340	11272
R2	0.27	0.31	0.33

Panel B. Consumption by occupation

	Agriculture	Non-ag manual labor	Other
New road	-0.055	-0.002	0.030
	(0.081)	(0.086)	(0.040)
Control group mean	9.40	9.62	9.59
N	8534	8583	11350
R2	0.26	0.40	0.39

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of a new road on log predicted consumption. In Panel A, which divides households by education, Columns 1, 2, and 3 show results for households where the primary earner is illiterate, has primary education or below, and has middle school education or above, respectively. Panel B divides households by the occupation of the primary earner: agriculture, non-agricultural manual labor, and other. 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 for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Bootstrapped standard errors are reported below point estimates; see Data Appendix for details.

Table A9: First stage and reduced form estimates, main and placebo samples

Panel A. Main sample first stage and reduced form effects

	First stage	Reduced form				
	Road by 2012	Transport	Occupation (ag share)	Firms	Ag production	Consumption
Road priority	0.215	0.088	-0.073	0.060	0.018	0.007
	(0.017)	(0.040)	(0.034)	(0.035)	(0.027)	(0.030)
Control group mean	0.25	0.00	0.00	-0.00	-0.00	0.00
N	11432	11432	11432	10678	11432	11432
R2	0.30	0.20	0.30	0.31	0.54	0.50

Panel B. Placebo sample first stage and reduced form effects

	First stage	Reduced form				
	Road by 2012	Transport	Occupation (ag share)	Firms	Ag production	Consumption
Road priority	-0.002	-0.002	-0.016	0.010	-0.047	-0.013
	(0.017)	(0.060)	(0.039)	(0.040)	(0.032)	(0.035)
Control group mean	0.26	0.44	-0.22	0.23	-0.26	0.33
N	9142	9138	9081	8457	9142	9142
R2	0.35	0.29	0.41	0.49	0.51	0.47

Notes: This table presents a comparison of estimates of the effect of PMGSY prioritization on a village's probability of treatment (first stage) and reduced form estimates of the effect of PMGSY prioritization on indices of the five major families of outcomes, for both the main sample (Panel A) and a placebo sample of villages close to the thresholds that were not followed (Panel B). 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 for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates.

Table A10: Impact of new road on indices of major outcomes, by kernel and bandwidth

	Triangular			Rectangular		
	60	80	100	60	80	100
Transport	0.404	0.411	0.401	0.419	0.430	0.307
	(0.208)	(0.188)	(0.172)	(0.205)	(0.182)	(0.154)
	[0.05]	[0.03]	[0.02]	[0.04]	[0.02]	[0.05]
Ag occupation	-0.290	-0.337	-0.332	-0.343	-0.362	-0.260
	(0.181)	(0.162)	(0.148)	(0.176)	(0.157)	(0.133)
	[0.11]	[0.04]	[0.02]	[0.05]	[0.02]	[0.05]
Firms	0.394	0.281	0.235	0.275	0.159	0.172
	(0.177)	(0.158)	(0.144)	(0.172)	(0.153)	(0.131)
	[0.03]	[0.07]	[0.10]	[0.11]	[0.30]	[0.19]
Ag production	0.145	0.093	0.071	0.102	0.080	0.050
	(0.139)	(0.125)	(0.114)	(0.137)	(0.121)	(0.104)
	[0.30]	[0.46]	[0.54]	[0.46]	[0.51]	[0.63]
Consumption	0.112	0.063	0.035	0.098	0.030	-0.023
	(0.154)	(0.138)	(0.126)	(0.149)	(0.133)	(0.112)
	[0.47]	[0.65]	[0.78]	[0.51]	[0.82]	[0.84]
N	[8339]	[11099]	[13871]	[8339]	[11099]	[13871]

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of a new road on indices of the major outcomes in each of the five families of outcomes: transportation, occupation, firms, agriculture and welfare. We show robustness to three different bandwidth choices (60, 80, 100) and two different kernel weighting choices (rectangular and triangular). See Section B for details of index construction. The specification includes baseline village-level controls for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Coefficients are presented for each regression with standard errors in parentheses and p-values in brackets.

Table A11: Impact of new road on population growth, age distribution and gender ratios

Panel A. Population growth (2001-2011)

	Log	Level
New road	-0.024	-9.662
	(0.029)	(20.275)
Control group mean	6.43	653.06
N	11432	11432
R2	0.79	0.83

Panel B. Age group share

	11-20	21-30	31-40	41-50	51-60
New road	-0.004	-0.003	0.002	-0.002	0.002
	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)
Control group mean	0.24	0.19	0.15	0.11	0.07
N	11432	11432	11432	11432	11432
R2	0.22	0.19	0.26	0.38	0.40

Panel C. Male share by age group

	11-20	21-30	31-40	41-50	51-60
New road	-0.010	0.003	0.004	-0.006	0.017
	(0.009)	(0.008)	(0.008)	(0.010)	(0.013)
Control group mean	0.52	0.52	0.51	0.52	0.51
N	11432	11432	11432	11432	11432
R2	0.13	0.19	0.10	0.07	0.05

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of PMGSY treatment on village demographics. Panel A presents results on 2011 village population, both in log and level. Panel B presents results on the share of the village population in ten-year age bins. Panel C presents results on the share of the population in each age bin that is male. Dependent variables in Panels B and C are generated from the SECC microdata. 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 for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates.

Table A12: Impact of new road on unemployment

	Unemployed	Unclassifiable
New road	0.010	-0.009
	(0.024)	(0.010)
Control group mean	0.430	0.018
N	11432	11432
R2	0.30	0.17

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of new road construction on the occupational choice. In the first column, the dependent variable is the share of working age adults (18-60) who do not work outside of the house (household work, student, unemployed, etc), while in the second column the dependent variable is the share of working age adults whose occupation does not make clear whether or not they work. 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 for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates.

Table A13: Impact of new road on sanitation

	Open Defecation	Latrine (on premises)	Pit Latrine (with slab)	Pit Latrine (without slab)
New road	0.006	-0.003	0.019	-0.010
	(0.038)	(0.036)	(0.017)	(0.012)
Control group mean	0.891	0.105	0.019	0.011
N	1776	1776	1776	1776
R2	0.25	0.27	0.09	0.08

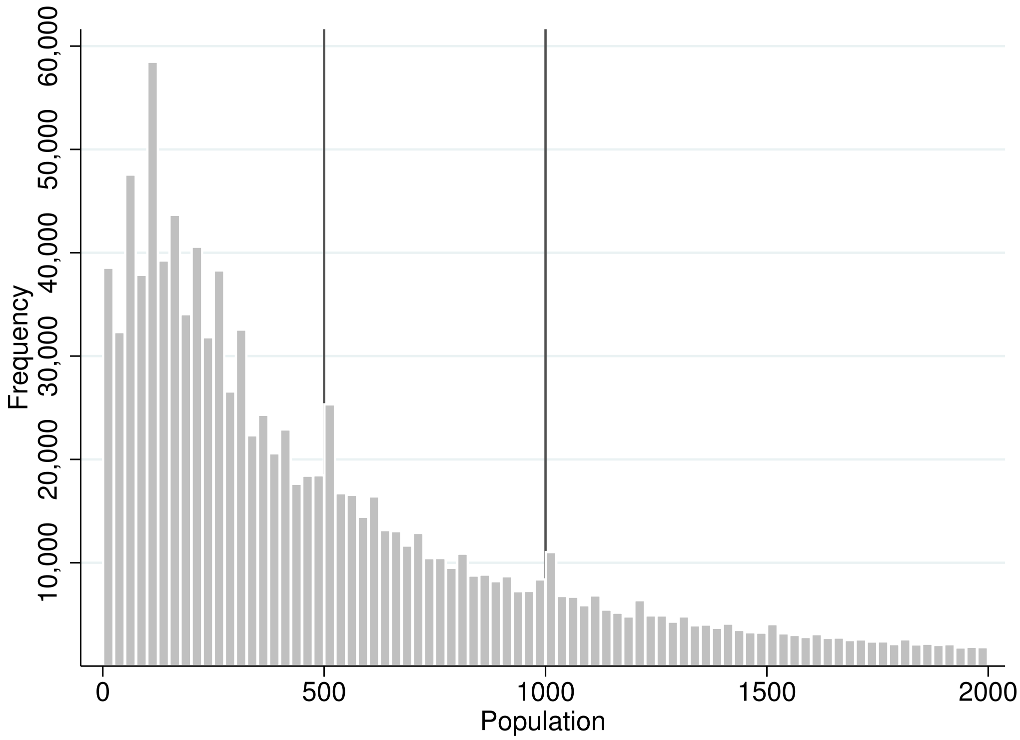
Notes: The Total Sanitation Campaign (TSC) is stated to have "aimed to transition rural households from open defecation to use of on-site pit latrines" (Spears, 2015). The program began construction of latrines in 2001. The outcomes considered here are 2011 Population Census measures of (in order) percentages 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 (84) of 1,000, the threshold used by the TSC. The sample of states here come from our main PMGSY specification. The specification includes baseline village-level controls for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates.

Table A14: Spillovers: impact of new road on nearby villages

	Transportation	Ag occupation	Firms	Ag production	Consumption	Unemployment rate
New road	-0.049	-0.001	-0.165	0.036	0.060	-0.007
	(0.135)	(0.132)	(0.141)	(0.100)	(0.114)	(0.009)
p-value	0.72	1.00	0.24	0.72	0.60	0.45
N	11403	11403	11403	11403	11403	11403
R2	0.51	0.52	0.46	0.71	0.65	0.70

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of a new road on outcomes in nearby villages. Dependent variables are indices of the five families of outcomes (transportation, occupation, firms, agriculture, and welfare), plus a sixth column for the unemployment rate. A catchment area for a PMGSY sample village is defined as other villages within 5 km. Outcomes are aggregated across spillover villages. Otherwise the specification is identical to the main regression specification for estimating direct effects. See Section B for details of index construction. The specification includes baseline village-level controls for amenities and economic indicators, as well as district-cutoff fixed effects (see Section V for details). Heteroskedasticity robust standard errors are reported below point estimates.

Figure A1: Histogram of habitation populations (PMGSY OMMS)



Notes: The figure shows the histogram of the habitation populations as reported in the PMGSY Online Monitoring and Management System. The vertical lines show the program eligibility thresholds at 500 and 1,000. Due to evidence of manipulation in the PMGSY administrative data, the running variable used in the analysis is population from the 2001 Population Census.

Figure A2: 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 वृत्ति 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	रेफ्रिजरेटर	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 व्हीलर कृषि उपकरण	सिंचाई उपकरण(नलकूप . बोर . डीजल/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

Notes: 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.

B Data Appendix

Section IV gives an overview of the data used in this paper. This data appendix provides more detail on the data sources and construction of the main variables.

B1 Administrative Data on Road Construction

Data on road construction come from the administrative software designed for the management of the program. The data include road sanctioning and completion dates, cost and time overruns, contractor names, and quality monitoring reports.

PMGSY data are posted online (http://omms.nic.in) at either the habitation or the road level; the data for this paper were all scraped in January 2015. 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 road program if at least one habitation in the village received a completed road by the year before outcome data were collected.

We matched the administrative road data to economic, population and poverty census data at the village level. In order to generate a village correspondence across multiple datasets, we conducted a fuzzy matching of location names, along with manual cleaning and quality verification.[26] We successfully match over 85% of habitations listed in the PMGSY to their corresponding population census villages.

B2 Socioeconomic censuses

Data on occupation, earnings and assets 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 individual demographic variables such as age, gender, and caste group, as well as various measures of household economic activity and assets, which we use to construct baseline control variables.

[26]For fuzzy matching, we used a combination of the reclink program in Stata, and a custom fuzzy matching script based on the Levenshtein algorithm but modified for the languages used in India. The fuzzy matching algorithm can be downloaded from the corresponding author's web site.

The fourth such census, the Socioeconomic and Caste Census (SECC), was launched in 2011 but primarily conducted in 2012.[27] To increase the likelihood of collecting data on all individuals and households, it was based on the National Population Register (NPR) from the 2011 Population Census. To increase transparency, the Government of India made the SECC publicly available at http://secc.gov.in in a mix of PDF and Excel formats; currently only aggregated data is available on the website. See Figure A2 for a de-identified sample page for a single household. We scraped over two million files, parsed the files into text data, and translated these from twelve different Indian languages into English. At the individual level, these data contain variables describing age, gender, occupation, caste group, disability and marital status. Data on occupations are written free-form in the SECC; after translation, we cleaned and matched these descriptions to the 2004 National Classification of Occupations (NCO). Our main occupational variables (share of workers in agriculture and share of workers in non-agricultural manual labor) are based on this classification: agricultural workers are those with NCO single digit code 6 (skilled agricultural workers) or NCO 2 digit 92 (agricultural laborers), while non-agricultural manual laborers are those with NCO single digit code 9 (elementary occupations) excluding those in agriculture (code 92).

At the household level, this dataset contains variables describing housing, landholdings, agricultural assets, household assets and sources of income.

We geocoded and matched 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.

B3 Economic and population censuses

The Indian Ministry of Statistics and Programme Implementation (MoSPI) conducted the 6th Economic Census in 2013. The Economic Census is a complete enumeration of all economic establishments except those engaged in crop production, defense and government administration. Establishments are any location, commercial or residential, where an economic activity is carried out. There is no minimum firm size, and both formal and informal establishments are enumerated, including people working out of their houses. We obtained the location directory for the Economic Census, and then used a series of fuzzy matching algorithms to match villages and towns by name to the population census of 2011. Employment is defined as the number of workers at the firm on the work day prior to the enumerator's visit, including casual wage laborers. We aggregate the microdata to the village level to obtain a measure of employment in village nonfarm firms. We use the sum of employment in all firms reported in the 2013 Economic Census to pro-

[27] 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.

duce an endline measure of nonfarm employment. The Economic Census also reports the sector of the firm, which we use to test for heterogeneous effects across the five largest sectors in our sample (livestock, forestry, manufacturing, retail and education), which together account for 79% of employment in in-village nonfarm firms. For all regressions using this data, we define the outcome variable as $log(employment_{i,v} + 1)$, where employment is the sum of employment in all firms in sector i in village v. To ensure that outliers do not drive our results, we restrict our sample in regressions using outcomes from the Economic Census to villages where total employment is less than total inhabitants in the village.

The Primary Census Abstract (PCA) and Village Directory (VD) give us village-level data in the Population Censuses of 2001 and 2011. The 2001 data provides control variables for the main regressions and is used to establish baseline balance for the regression discontinuity, while 2011 data is used to measure endline outcomes. The PCA is the source for demographic information (such as total population) and the VD for village characteristics and amenities (such as roads, electricity, schools, regular availability of transportation, etc.).

We also test for outcomes from two new measures of agricultural inputs from the 2011 Population Census Village Directory. The first is crop choice. The census records the three major crops for each village—from this we generate an indicator variable for whether the village grows any non-subsistence crops, which we define as anything other than cereals (rice, wheat, etc) and pulses (lentils, chickpeas, etc). The second is total agricultural land, which we transform into logs.

These censuses also provide the basis for linking the various other datasets. We use a key provided by the 2011 Population Census to link data from 2011 to 2001. GIS data of village boundaries in 2011, procured from ML Infomap and based on official census maps, is used for the aggregation of gridded remote sensing to the village level.

We have combined multiple rounds of the economic and population censuses into a single dataset, referred to as the Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG), Version 1.0. Asher and Novosad (2018) provides details of its construction and guidance on its use. The dataset can be found at http://www.dartmouth.edu/~novosad/data.html.

B4 Agricultural production

As no comprehensive village-level data is collected on agricultural production in India, we use two commonly-used and closely related vegetative indices (VIs) to proxy for agricultural production in baseline and endline survey periods: the normalized difference in vegetation index (NDVI) and the enhanced vegetation index (EVI), which is very similar but uses additional information from the blue part of the electromagnetic spectrum.

NDVI and EVI are chlorophyll-sensitive measures of plant matter, generated at global coverage and 250 m resolution by the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASA's Earth Observing System-Terra satellite. NDVI is built using near infrared and red bands, while EVI uses additional information from the blue band to reduce atmospheric interference and the influence of background vegetation (Son et al., 2014). NDVI and EVI have shown to be equivalently effective for crop classification tasks (Wardlow and Egbert, 2010), and have also been shown to be equally successful at predicting wheat yields in Canada when combined with agroclimate data (Kouadio et al., 2014). Each image represents a 16-day composite where each pixel value is optimized considering cloud cover obstruction, image quality, and viewing geometry via the MODIS VI algorithm (Huete et al., 2002). Composite images were downloaded from the Columbia University IRI Data Library for the years 2000-2014 for nine 16-day periods from late May through mid-October, covering the major (kharif) cropping season in India (Selvaraju, 2003).

For each composite image, pixels were spatially averaged to village polygons. After village aggregation within each 16-day composite, three proxies for agricultural production were calculated for each year's growing season: the difference between early-season VI (the mean of the first three 16-day composites) and the max VI value observed at the village level (Labus et al., 2002; Rasmussen, 1997), mean VI (Mkhabela et al., 2005), and cumulative NDVI (Rojas, 2007) (the sum of NDVI from each of the nine composites during the growing season).[28] All VI measures are then log transformed for the regressions to allow for an interpretable effect. We prefer the differenced measure because it effectively controls for non-crop vegetation (such as forest cover) by measuring the change in vegetation from the planting period (when land is fallow) to the moment of peak vegetation.

We use additional likely correlates of agricultural production to validate the use of these growing-season VI measures as a proxy for agricultural output at the village level (Table A1). Cross-sectional regressions with state fixed effects were run using log endline year (2011-2013 average) growing season change in NDVI (as described above) as the dependent variable. At the village level, these correlates are: cereal crop potential production measure (low input usage) from the FAO Global Agro-Ecological Zones (GAEZ) aggregated to the village level (log); share of village agricultural land area under any type of irrigation; and per capita annual predicted consumption (described below). Additionally, panel NDVI data was regressed at the district level on agricultural output from the Planning Commission's series of district domestic product data, across a consistent sample of districts. While these remotely sensed measures of agricultural production do not capture other determinants of agricultural earnings such as quality or price changes, their strong correlation with both agricultural productivity measures and real measures of pro-

[28] To reduce noise, we define our endline measure as the average of the measures for 2011, 2012 and 2013, and our baseline measures as the average of the measure for 2000, 2001 and 2002.

duction supports using them to estimate impacts of roads on village agricultural production.

B5 Consumption

We combine data from 2012 SECC and the concurrent IHDS-II (2011-12) to predict village-level consumption measures following the methodology in Elbers et al. (2003). To do this, using IHDS data, we regress total household consumption on dummy variables that are equivalent to all asset and earnings information contained in the SECC.[29] The results of this regression are given in Table A6. We then use the coefficients to predict household-level consumption in the SECC microdata. This is used to generate consumption per capita at the individual level, which is in turn used to produce village level statistics for mean predicted consumption per capita, per capita predicted consumption at different village percentiles, and share of the population below the poverty line.[30] For the purpose of regressions, consumption variables are winsorized at the 1st and 99th percentiles, and log transformed. As outlined in Elbers et al. (2003), in order to get correct standard errors and p-values, we perform a double bootstrap, first in the IHDS regressions to generate 1,000 different asset coefficient vectors, and then over villages in our main sample.[31]

This method is supported by a large literature on predicting consumption and proxying welfare using asset and related data. Early work showed that in the United States, up to 78% of the variation in total consumption could be predicted by a linear regression on food consumption, housing expenditures and valuation, vehicle ownership, size of the family, and age (Skinner, 1987). Hentschel et al. (2000) show that this method yields unbiased estimates of poverty and performs well except when sample sizes are very small. McKenzie (2005) evaluates the ability of this method to generate accurate measures of inequality and poverty, finding that it better predicts non-durable consumption than other methods considered; he also validates this measure by finding that predicted consumption and directed measured consumption generate highly similar conclusions on the relationship between inequality and schooling in Mexico. Both McKenzie (2005) and Young (2012) make the further point that assets have the advantage of likely capturing real, permanent income better than consumption measured at any moment in time. Predicted consumption using this method has also been widely used (most notably by the World Bank) to generate poverty estimates using census data for areas not covered by detailed (and expensive)

[29]These variables are roof material (grass, tile, slate, plastic, GI metal, brick, stone, and concrete), wall material (grass, mud, plastic, wood, brick, GI sheets, stone, and concrete), number of rooms, phone ownership (landline only, mobile only, and both landline and mobile), house ownership (owned), vehicle ownership (two wheeler and four wheeler), land ownership, kisan credit card, refrigerator, and highest individual income in household (between 5,000 and 10,000 rupees and more than 10,000 rupees).

[30]We use the official rural poverty line of INR 27/day from the Tendulkar Committee Report (Government of India, 2014).

[31]To speed up the computation of the bootstrapped estimations, we modify GNU Parallel code (Tange, 2011).

household consumption surveys (Bedi et al., eds, 2007). While we are undoubtedly missing some of the variation in consumption not explained by these assets and income variables, our large sample sizes (median village has 152 households in the SECC) and wide range of assets covered (from housing materials to vehicles to mobile phones) give us confidence that our measure of predicted consumption is sufficiently precise to pick up major changes in consumption.

For an alternative way of aggregating information across assets, we create an index at the village level by taking the primary component of the indicator variables described above in the SECC microdata, normalized to have a mean of 0 and standard deviation of 1 within our sample.

The only earnings variable available at the village level comes from the SECC. It records monthly earnings of the highest earning member of the household, censored into three bins: 0 to 4,999 rupees, 5,000 to 9,999 rupees and 10,000+ rupees. As 85% of households report being in the lowest bin, we define our earnings variable to be the share of households in the top two bins (with the highest earner earning 5,000 rupees or more).

We generate another consumption proxy using lights at night, as measured by satellites. Night lights are a proxy for consumption that have the advantage of high resolution and objective measurement over a 20+ year period (Henderson et al., 2011). We match gridded data to village polygons, sum over all pixels in the village and then take the log of the value plus 1 in order to not drop observations that take the value 0. To increase precision, we define our dependent variable as the log of the mean value from 2011, 2012 and 2013 (plus 1), and include a control for log mean baseline light (plus 1) in 2000-2002.

B6 Spillovers

Spillover effects of PMGSY road construction on nearby villages are assessed using 2001 Population Census GIS data purchased from ML InfoMap. Catchment areas with radii of 5 km were constructed by measuring distances from the centroids of villages in the sample to the centroids of all other villages. Outcomes were then aggregated across all villages within these catchment areas, constructed in the same manner as for the non-spillover regressions. On average, there are 15 villages per 5 km catchment area. 55 percent of non-sample villages within a catchment appear in more than one catchment at 5km. These villages are double counted, but should not bias the estimates due to the exogeneity of road construction in our regression discontinuity sample.

B7 Family-wise indices

In order to address concerns of multiple hypothesis testing, we follow Anderson (2008) in generating five indices for our main families of outcomes: transportation, labor market, firms, agriculture and assets/consumption. Each of these is generated by demeaning its component outcomes and converting to effect sizes through dividing by control group standard deviation; demeaned values are then combined by weighting according to the inverse of the covariance matrix. The transportation index is comprised of five indicator variables for availability of motorized transit: public buses, private buses, vans, taxis and auto-rickshaws. The labor market index is comprised of the share of workers in agriculture and the opposite of the share of workers in manual labor (so that their covariance is positive). The firms index is comprised of log of employment plus 1 in all nonfarm firms; it does not include the other firm outcomes as they are simply disaggregations of total employment by sector. The agriculture index is comprised of our favored measure of agricultural yields (differenced NDVI, described above) and each of the measures of agricultural inputs: share of households owning mechanized farm equipment, share of households owning irrigation equipment, share of households owning land, log total cultivated acres and an indicator for non-cereal/pulse (subsistence) crops among the primary three crops in the village. Finally, the asset/consumption index is comprised of log predicted consumption per capita, the primary component asset index, log night light luminosity and the share of households with the primary earner making more than 5,000 INR per month.