American Economic Review 2020, 110(3): 797–823
https://doi.org/10.1257/aer.20180268

Rural Roads and Local Economic Development[†]

By Sam Asher and Paul Novosad*

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* J43, O13, O18, R23, R42)

Nearly one billion people worldwide live more than 2 kilometers from a paved road, with one-third living in India (Roberts, Shyam, and Rastogi 2006; 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 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.

*Asher: Johns Hopkins University SAIS (email: sasher2@jhu.edu); Novosad: Dartmouth College (email: paul.novosad@dartmouth.edu). Esther Duflo was the coeditor for this article. Asher was an employee of the World Bank when this manuscript was accepted. We wish to thank the 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.

[†]Go to https://doi.org/10.1257/aer.20180268 to visit the article page for additional materials and author disclosure statements.

Existing research is largely supportive of policymakers' claims: rural road construction is associated with increases in farm and nonfarm 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, Harding, and Wantchekon 2013; Brueckner 2014; Burgess et al. 2015; Lehne, Shapiro, and Vanden Eynde 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, Lanjouw, and Lanjouw (2003) to predict consumption from the set of asset and income variables in the individual microdata. We can rule out a 10 percent increase in predicted consumption with 95 percent 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 nonfarm employment (4 workers per village), which can explain only 23 percent 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), which 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 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, Mobarak, and Barham 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.

[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. 2017). Results on growth have proven somewhat mixed: there is evidence that reducing transportation costs can increase (Ghani, Goswami, and Kerr 2016; Storeygard 2016), decrease (Faber 2014), or leave unchanged (Banerjee, Duflo, and Qian 2012) growth rates in local economic activity. Atkin and Donaldson (2017) shows that intracountry 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, Bakht, and Koolwal 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, Asher, and Novosad 2020). Other papers also suggest that the lack of rural transport infrastructure may be a significant contributor to rural underdevelopment. Wantchekon, Klašnja, and Novta (2015) provides evidence that transport costs are a strong predictor of poverty across sub-Saharan Africa. Fafchamps and Shilpi (2005) offers 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, Khandker, and Rosenzweig 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, Glennerster, and Suri 2013), and agricultural productivity and crop choice (Sotelo 2018). In an urban setting, Gonzalez-Navarro and Quintana-Domeque (2016) finds that paving streets leads to higher property values and consumption.

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, Lagakos, and Waugh 2014; McMillan, Rodrik, and Verduzco-Gallo 2014). We join recent research that finds that high transportation costs are an important barrier to the spatial and sectoral allocation of labor (Bryan, Chowdhury, and Mobarak 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.

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

I. 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.

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

The impacts of roads on nonfarm production in the village are analogous. Lower input prices and higher output prices will increase the production of nonfarm 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 nontradable 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 unwilling 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.

II. 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 preexisting 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

[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

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 meters 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. 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 IV 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 kilometers 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]

III. 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 the online Appendix.

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 come 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 individuals. See the online Appendix for more details on the construction of this dataset. After extracting

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.

[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 et al. (2019) for details on its construction. The dataset containing a limited set of variables, along with keys providing merges to the economic and population censuses from 1990–2013, can be found at https://doi.org/10.7910/DVN/DPESAK.

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 national survey (IHDS-II, 2011–2012) 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, Lanjouw, and Lanjouw (2003), allowing us to test not only for impacts of roads on mean predicted consumption per capita but also for distributional effects.[7] The online Appendix contains additional details of this process, as well as a discussion of the literature on predicting consumption from such data.

Firm data come 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 are used to measure outcomes otherwise unavailable at the village level. Night lights provide a proxy for total village output. As no village-level agricultural production data exist 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 the 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.

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

[8] Online Appendix 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 the online Appendix for additional details.

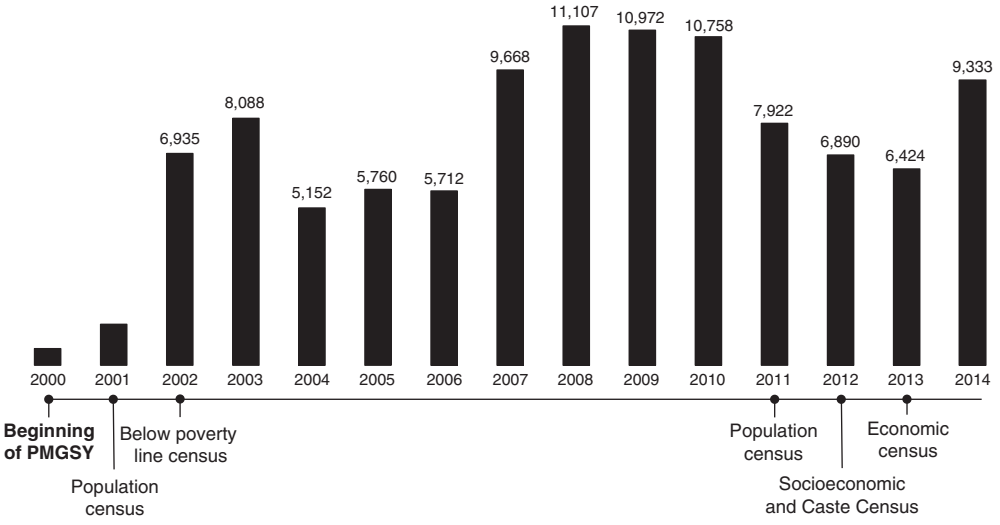

FIGURE 1. TIME LINE OF DATA SOURCES WITH COUNT OF ROADS COMPLETED UNDER PMGSY BY YEAR

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 number of villages receiving new PMGSY roads in each year. Exact counts are also listed.

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 new 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]

IV. 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 naïve estimates in an unknown direction. Second, because roads are costly, road construction programs rarely generate large treatment samples. Sample surveys not directly

[9] Online Appendix 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 one-half of all villages) are less populated (1,513 versus 1,930), have fewer public goods (e.g., 25 percent electrified versus 55 percent), have less 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.

Table 1—Summary Statistics and Balance

	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
Observations	11,432	6,018	5,414				

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 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 ±84 around the population thresholds has been used to define the sample of villages (see text for details).

connected with road construction programs are thus unlikely to have a sufficient number of treated and control observations; 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 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 population threshold prioritization rules as given by the national guidelines of the PMGSY. We worked closely with the National Rural Roads Development Agency to identify the state-specific thresholds that were followed and 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 (2019), 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:

$$(1)\quad Road_{v,j} = \gamma_0 + \gamma_1 \mathbf{1}\{pop_{v,j} \geq T\} + \gamma_2(pop_{v,j} - T) + \gamma_3(pop_{v,j} - T) \times \mathbf{1}\{pop_{v,j} \geq T\} + \nu X_{v,j} + \mu_j + \upsilon_{v,j},$$

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

Here, $Y_{v,j}$ 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 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 kilometers 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 US$4), 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. The variable $Road_{v,j}$ is an indicator that takes the value 1 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 and Kalyanaraman (2012) is 84.[12] We use a triangular kernel which places the most weight on observations close to the

[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.

[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 percent, 5.5 percent, and 6.4 percent, slightly below the 2008–2016 average of 7.1 percent. 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, Cattaneo, and Titiunik (2014) is 78.

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 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 (standard error 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.

V. Results

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 online Appendix. 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 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

[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 (online Appendix 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.

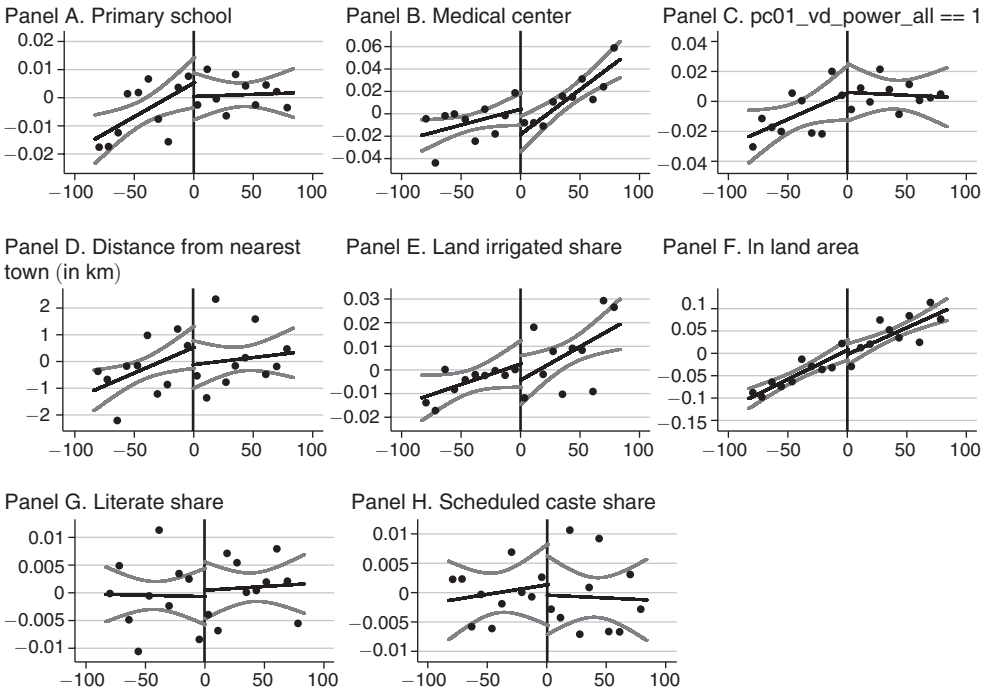

Figure 2. Balance of Baseline Village Characteristics

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 0 are above treatment thresholds, while points to the left of 0 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 percent 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).

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 percent 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.

[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.

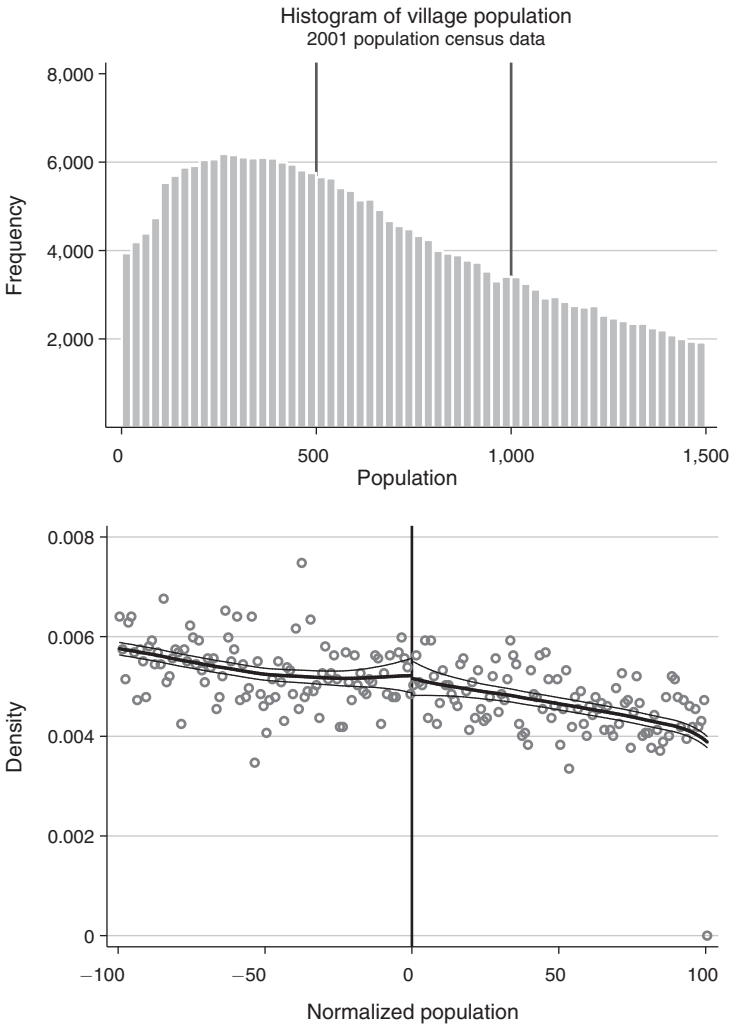

FIGURE 3. DISTRIBUTION OF RUNNING VARIABLE

Notes: The figure shows the distribution of village population around the population thresholds. The top 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 bottom panel uses the normalized village population (reported population minus the threshold, either 500 or 1,000). It plots a nonparametric 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.

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 of transportation, do not experience significant growth. Availability of autorickshaws, the least expensive

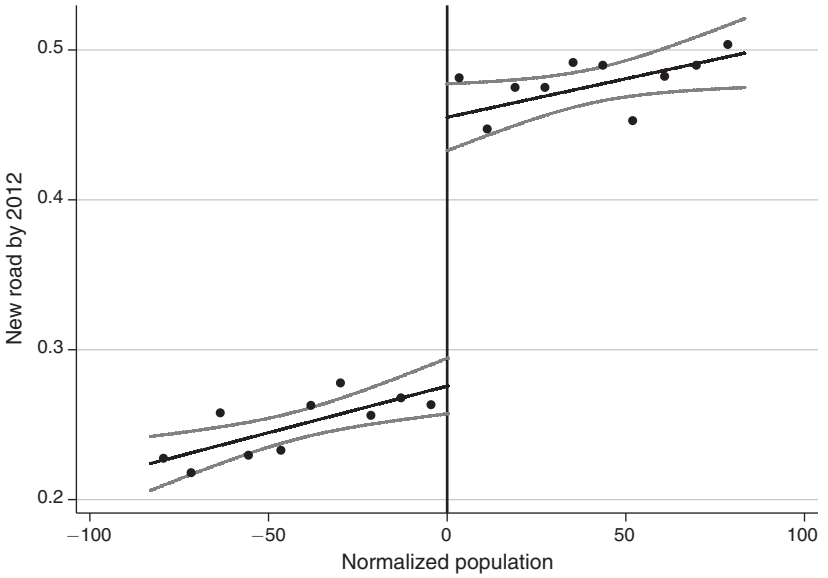

Figure 4. First Stage: Effect of Road Prioritization on Probability of New Road by 2012

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.

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
Observations	8,339	9,720	11,099	12,457	13,871	15,238
R^2	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 1 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 940–1,060 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 IV for details). Heteroskedasticity robust standard errors are reported below point estimates.

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]

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

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
Observations	11,432	11,432	10,678	11,432	11,432
R^2	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 consumption. See the online Appendix 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 IV for details). Heteroskedasticity robust standard errors are reported below point estimates.

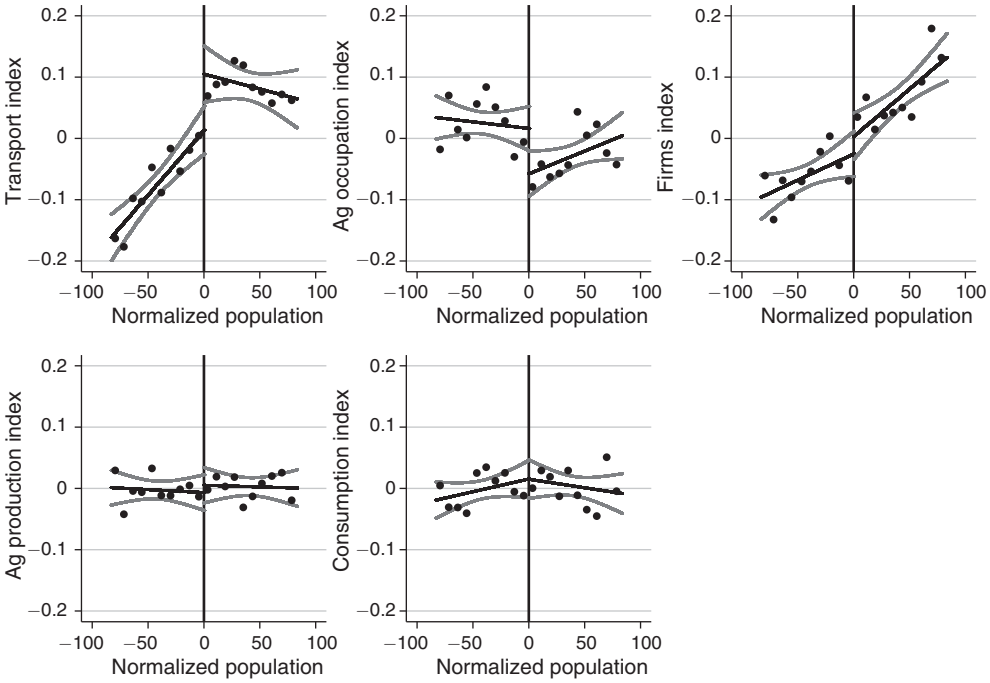

FIGURE 5. REDUCED FORM: EFFECT OF ROAD PRIORITIZATION ON INDICES OF MAJOR OUTCOMES

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 consumption) 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.

Table 5 presents impacts of new roads on occupational choice, the one domain where roads appear to substantially change economic behavior. As 92 percent 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

Table 4—Impact of New Road on Transportation

	Govt. 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
Observations	11,432	11,432	11,432	11,432	11,432
R^2	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 IV 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
Observations	11,432	11,432	11,432	11,432
R^2	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 IV for details). Heteroskedasticity robust standard errors are reported below point estimates.

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 percent decrease from the control group mean) and a 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

[16] The SECC does not report manual labor occupations in more detail. Online Appendix Table A3 breaks down the sectoral distribution of non-agricultural manual laborers using the 68th round of the National Sample Survey (2011–2012). 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.

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.

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 online Appendix 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 online Appendix 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 percent of workers with no land exit agriculture, compared to just 10 percent 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]

We next examine the heterogeneity of the treatment effect as a function of age and gender (online Appendix 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.

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

Table 6—Impact of New Road on Firms

	Total	Livestock	Manufacturing	Education	Retail	Forestry
Panel A. Log employment growth, by sector						
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)
Observations	10,678	10,678	10,678	10,678	10,678	10,678
R^2	0.30	0.42	0.23	0.18	0.23	0.35
Panel B. Level employment growth, by sector						
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
Observations	10,678	10,678	10,678	10,678	10,678	10,678
R^2	0.30	0.46	0.18	0.13	0.17	0.36

Notes: This table presents regression 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 IV for details). Heteroskedasticity robust standard errors are reported below point estimates.

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 nonfarm 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 percent of nonfarm employment. We estimate a 27 percent increase in employment in nonfarm 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 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 percent 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 nonfarm employment growth in the village may be more a function of new consumption opportunities (perhaps due

TABLE 7—IMPACT OF NEW ROAD ON AGRICULTURAL OUTCOMES

	NDVI			EVI		
	Max–June	Cumulative	Max	Max–June	Cumulative	Max
Panel A. Agricultural yields (log)						
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
Observations	11,333	11,332	11,333	11,333	11,332	11,333
R^2	0.71	0.89	0.82	0.72	0.86	0.72

	Mechanized farm equipment	Irrigation equipment	Land ownership	Non-cereal/ pulse crop	Cultivated land (log)
Panel B. Agricultural inputs					
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
Observations	11,431	11,432	11,432	8,272	11,165
R^2	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.

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 III, 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 percent higher agricultural yield (equivalent to 0.044 SD) and can rule out a 6.8 percent or a 0.25 standard deviation increase in yield with 95 percent 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.

Table 8—Impact of New Road on Predicted Consumption, Earnings, and Assets

	Consumption per capita (log)	Poverty rate	Night lights (log)	Share of HH earning ≥ INR 5k	
Panel A. Consumption and earnings					
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	
Observations	11,432	11,432	11,102	11,432	
R^2	0.41	0.30	0.66	0.25	
	Asset index	Solid house	Refrigerator	Vehicle	Phone
Panel B. Asset ownership					
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
Observations	11,432	11,432	11,432	11,432	11,432
R^2	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 online 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 2011–2013, 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 5,000 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 IV 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 online Appendix.

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 percent increase in predicted consumption; we can rule out a 10 percent increase with 95 percent confidence. As explained in the online Appendix, 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); online Appendix 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

[20] Online Appendix 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).

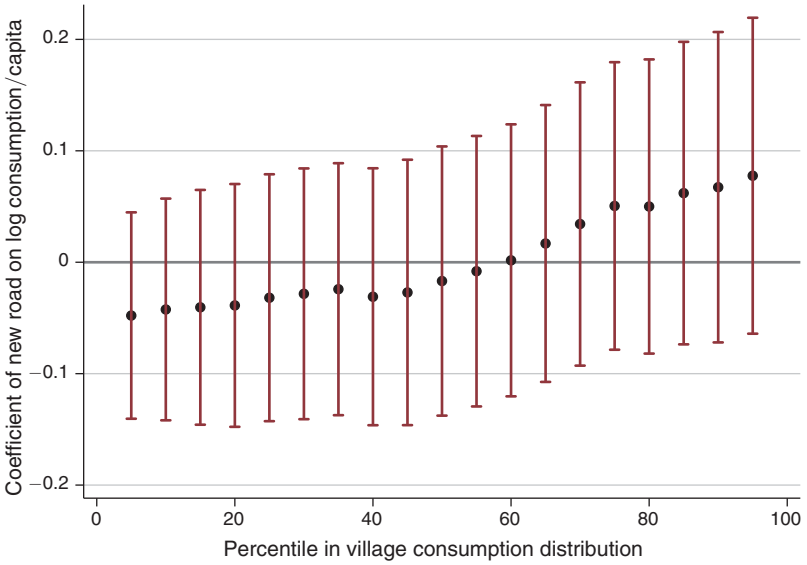

FIGURE 6. DISTRIBUTIONAL IMPACTS OF NEW ROAD ON PREDICTED CONSUMPTION

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 the online Appendix for description of bootstrapping.

positive at the top of the distribution, but very small and insignificant everywhere. Online Appendix 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, Storeygard, and Weil (2011) estimates a robust elasticity of 0.3 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 cause a 10 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

[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.

[22] As noted in the online Appendix, 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.

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.

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 eligibility 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). Online Appendix 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 online Appendix 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 those in our main specification (Table 3).

If migration is correlated with individual or household characteristics, as some studies have found (Bryan, Chowdhury, and Mobarak 2014; Morten and Oliveira 2018), then compositional changes in village population could bias treatment estimates. In online Appendix 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

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

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 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.

Online Appendix 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 online Appendix 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-kilometer radius of villages in the main sample, using the standard regression discontinuity specification. Online Appendix 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 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.

VI. 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.

[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 highway construction.

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 policymakers 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 $3,945 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 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, Asher, and Novosad 2020). 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. 2020. "Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction." *American Economic Journal: Applied Economics* 12 (1): 348–76.

Aggarwal, Shilpa. 2018. "Do Rural Roads Create Pathways Out of Poverty? Evidence from India." *Journal of Development Economics* 133: 375–95.

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

Anderson, Michael L. 2008. "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* 103 (484): 1481–95.

[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.

Asher, Sam, Ryu Matsuura, Tobias Lunt, and Paul Novosad. 2019. "The Socioeconomic High-Resolution Rural-Urban Geographic Dataset on India (SHRUG)." Unpublished.

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

Banerjee, Abhijit, Esther Duflo, and Nancy Qian. 2012. "On the Road: Access to Transportation Infrastructure and Economic Growth in China." NBER Working Paper 17897.

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang. 2017. "Roads, Railways and Decentralization of Chinese Cities." *Review of Economics and Statistics* 99 (3): 435–48.

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

Binswanger, Hans P., Shahidur R. Khandker, and Mark R. Rosenzweig. 1993. "How Infrastructure and Financial Institutions Affect Agricultural Output and Investment in India." *Journal of Development Economics* 41 (2): 337–66.

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

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

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

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

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

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel. 2015. "The Value of Democracy: Evidence from Road-Building in Kenya." *American Economic Review* 105 (6): 1817–51.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli. 2016. "Agricultural Productivity and Structural Transformation. Evidence from Brazil." *American Economic Review* 106 (6): 1320–65.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

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

de Janvry, Alain, Marcel Fafchamps, and Elisabeth Sadoulet. 1991. "Peasant Household Behaviour with Missing Markets: Some Paradoxes Explained." *Economic Journal* 101 (409): 1400–17.

Dell, Melissa. 2015. "Trafficking Networks and the Mexican Drug War." *American Economic Review* 105 (6): 1738–79.

Dercon, Stefan, Daniel O. Gilligan, John Hoddinott, and Tassew Woldehanna. 2009. "The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages." *American Journal of Agricultural Economics* 91 (4): 1007–21.

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

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review* 108 (4–5): 899–934.

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

Duflo, Esther, and Rohini Pande. 2007. "Dams." *Quarterly Journal of Economics* 122 (2): 601–46.

Elbers, Chris, Jean O. Lanjouw, and Peter Lanjouw. 2003. "Micro-Level Estimation of Poverty and Inequality." *Econometrica* 71 (1): 355–64.

Faber, Benjamin. 2014. "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System." *Review of Economic Studies* 81 (3): 1046–70.

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

Fan, Shenggen, and Peter Hazell. 2001. "Returns to Public Investments in the Less-Favored Areas of India and China." *American Journal of Agricultural Economics* 83 (5): 1217–22.

Fernando, A. Nilesh. 2018. "Shackled to the Soil: The Long-Term Effects of Inherited Land on Labor Mobility and Consumption." Unpublished.

Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics* 37 (3): 447–56.

Ghani, Ejaz, Arti Grover Goswami, and William R. Kerr. 2016. "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing." *Economic Journal* 126 (591): 317–57.

Gibson, John, and Susan Olivia. 2010. "The Effect of Infrastructure Access and Quality on Non-Farm Enterprises in Rural Indonesia." *World Development* 38 (5): 717–26.

Gibson, John, and Scott Rozelle. 2003. "Poverty and Access to Roads in Papua New Guinea." *Economic Development and Cultural Change* 52 (1): 159–85.

Goldin, Claudia. 1995. "The U-Shaped Female Labor Force Function in Economic Development and Economic History." In *Investment in Women's Human Capital*, edited by T. Paul Schultz, 61–90. Chicago: University of Chicago Press.

Gollin, Douglas, David Lagakos, and Michael E. Waugh. 2014. "The Agricultural Productivity Gap." *Quarterly Journal of Economics* 129 (2): 939–93.

Gollin, Douglas, and Richard Rogerson. 2014. "Productivity, Transport Costs and Subsistence Agriculture." *Journal of Development Economics* 107: 38–48.

Gonzalez-Navarro, Marco, and Climent Quintana-Domeque. 2016. "Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects." *Review of Economics and Statistics* 98 (2): 254–67.

Henderson, Vernon, Adam Storeygard, and David N. Weil. 2011. "A Bright Idea for Measuring Economic Growth." *American Economic Review* 101 (3): 194–99.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* 79 (3): 933–59.

Imbens, Guido, and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.

Indian Roads Congress. 2002. *Rural Roads Manual. Indian Roads Congress Special Publication.* New Delhi: IRC.

Jacoby, Hanan G. 2000. "Access to Markets and the Benefits of Rural Roads." *Economic Journal* 110 (465): 713–37.

Jacoby, Hanan G., and Bart Minten. 2009. "On Measuring the Benefits of Lower Transport Costs." *Journal of Development Economics* 89 (1): 28–38.

Jayachandran, Seema. 2006. "Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries." *Journal of Political Economy* 114 (3): 538–75.

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

Khandker, Shahidur R., and Gayatri B. Koolwal. 2011. "Estimating the Long-Term Impacts of Rural Roads: A Dynamic Panel Approach." World Bank Policy Research Paper 5867.

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

Lipscomb, Molly, A. Mushfiq Mobarak, and Tania Barham. 2013. "Development Effects of Electrification: Evidence from the Geologic Placement of Hydropower Plants in Brazil." *American Economic Journal: Applied Economics* 5 (2): 200–31.

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

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

Morten, Melanie, and Jaqueline Oliveira. 2018. "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City." NBER Working Paper 22158.

Mu, Ren, and Dominique van de Walle. 2011. "Rural Roads and Local Market Development in Vietnam." *Journal of Development Studies* 47 (5): 709–34.

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

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

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

Raballand, Gael, Rebecca Thornton, Dean Yang, Jessica Goldberg, Niall Keleher, and Annika Muller. 2011. "Are Rural Road Investments Alone Sufficient to Generate Transport Flows? Lessons from a Randomized Experiment in Rural Malawi and Policy Implications." World Bank Policy Research Paper 5535.

Redding, Stephen J., and Matthew A. Turner. 2015. "Transportation Costs and the Spatial Organization of Economic Activity." In *Handbook of Regional and Urban Economics*, Vol. 5B, edited by Gilles Duranton, J. Vernon Henderson, and William Strange, 1339–98. Amsterdam: Elsevier.

Roberts, Peter, K. C. Shyam, and Cordula Rastogi. 2006. "Rural Access Index: A Key Development Indicator." World Bank Technical Report Transport Paper TP-10.

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

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

Sotelo, Sebastian. 2018. "Domestic Trade Frictions and Agriculture." Unpublished.

Spears, Dean. 2015. "Effects of Sanitation on Early-Life Health: Evidence from a Governance Incentive in Rural India." Unpublished.

Storeygard, Adam. 2016. "Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa." *Review of Economic Studies* 83 (3): 1263–95.

Wantchekon, Leonard, Marko Klašnja, and Natalija Novta. 2015. "Education and Human Capital Externalities: Evidence from Colonial Benin." *Quarterly Journal of Economics* 130 (2): 703–57.

World Bank Group. 2016. *Measuring Rural Access Using New Technologies.* Washington, DC: World Bank.

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