Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction *

Anjali Adukia†

Sam Asher‡

Paul Novosad§

January 2018

Abstract

The rural poor in developing countries, once economically isolated, are increasingly being connected to outside markets. Whether these new connections crowd out or encourage educational investment is a central question. We examine the impacts on educational choices of 115,000 new roads built under India's flagship road construction program. We find that children stay in school longer and perform better on standardized exams. Treatment heterogeneity supports the predictions of a standard human capital investment model: enrollment increases are largest where nearby labor markets offer the highest returns to education and lowest where they imply high opportunity costs of schooling.

JEL Codes: I25; O18; J24.

*For helpful comments and guidance, we thank Martha Bailey, Chris Blattman, Liz Cascio, Eric Edmonds, Rick Hornbeck, Ruixue Jia, Ofer Malamud, Mushfiq Mobarak, Doug Staiger, Bryce Steinberg, and participants of seminars at AEFP, APPAM, Boston University, CIES, Columbia University, CSWEP, Dartmouth, DePaul, EEA, the Federal Reserve Banks of Chicago (CHERP) and New York, Georgetown University, NEUDC, the NBER Education Program Meetings, the OECD, PAA, PacDev, UC Berkeley, University of Chicago, University of Connecticut, University of Illinois at Chicago, University of Michigan, University of Missouri, the Urban Institute and Yale. We thank Srinivas Balasubramanian, Anwita Mahajan, Olga Namen, and Taewan Roh for excellent research assistance. We thank Arun Mehta and Aparna Mookerjee for help in data acquisition.

†University of Chicago, 1155 East 60th Street, Chicago, IL 60637, adukia@uchicago.edu

‡World Bank, 1818 H Street, NW, Washington, DC 20433, sasher@worldbank.org

§Dartmouth College, Economics Department, 6106 Rockefeller Center, Room 301, Hanover, NH 03755, paul.novosad@dartmouth.edu

I Introduction

Increased access to international markets has important influences on schooling decisions, which are central to supporting long-run economic growth.[1] A large share of the world's rural poor are not well-connected to international markets, however, and depend instead on domestic linkages to nearby towns and cities.[2] The impacts of domestic market integration are less studied than the impacts of connection to international markets. The key individual tradeoff is between long-run investment in human capital and immediate economic opportunities that might discourage increased schooling. Connections to new markets should encourage educational attainment if they increase returns to education, or otherwise raise household income or liquidity. However, immediate earnings opportunities for the young could motivate an earlier exit from schooling. If the schooling response to market integration is important, it could be central to the long run impacts of policies that affect market integration.

We examine the human capital investment response when a paved road is built to a previously unconnected village, effectively connecting it to a wider market. The source of variation is the rollout of India's national rural road construction program (PMGSY, or Prime Minister's Road Construction Program), under which the government built high quality roads to over 115,000 villages across the country between 2001 and 2015, connecting over 30 million rural households to nearby towns. We focus on new rural feeder roads, which provide terminal connections between the broader transportation network and previously unconnected villages. Ex ante, the impact of new road connections on schooling is theoretically ambiguous, because they may raise the returns to education, raise the opportunity cost of schooling, or have important income or liquidity effects.

The major challenge in estimating causal effects of new roads is the endogeneity of road placement. If, for example, roads are targeted either to wealthy or poor regions, then comparisons of villages with and without roads will be biased. To overcome this bias, we exploit

[1] See, for example, Edmonds and Pavcnik (2006), Edmonds et al. (2010) and Shastry (2012).

[2] See, for example, Atkin and Donaldson (2015), who show that domestic trade costs in developing countries can be considerably higher than international.

the timing of road completion in each village, estimating a panel regression with village and state-time fixed effects. Village fixed effects control for unobserved village-specific factors that may have influenced the timing of road construction. State-time fixed effects control for time-variant state-specific shocks and policies. Essentially, we compare educational outcomes in villages before and after a road is built, flexibly controlling for time-variant regional shocks and static differences between early-treated and late-treated villages.

We use village-level enrollment data from India's national annual school enrollment census for the first through the eighth grades, the District Information System for Education (DISE, 2002-2015). Through a combination of human and machine-based fuzzy matching, we linked DISE data to administrative data from the national rural road construction program. The result is a panel of over 300,000 villages across all of India. The use of census data is essential to our analysis, since variation in the road program is at the village-level. It also gives us power to precisely estimate impacts in subsamples of the population. Our sample spans the range of conditions in India today and across many places worldwide that remain unreached by paved roads.

We find that road construction significantly increases enrollment among middle-school children, who are most at risk of leaving school. We estimate that connecting a village with a new paved road causes a seven percent increase in middle-school enrollment over the following three years. The estimates are precise and statistically significant. We also estimate increases in the number of students taking and scoring highly on middle-school completion exams, indicating that educational performance is also improving.[3] The results are robust to a range of specifications and sample definitions, as well as a regression discontinuity specification that exploits a program rule that caused villages above specific village population thresholds to be targeted for road construction.

Next, we explore variation in treatment effects, guided by a standard human capital in-

[3] In many cases, interventions that improve attendance and enrollment do not improve student test scores (e.g., Miguel and Kremer (2004), Behrman et al. (2008), Adukia (2017)), perhaps due to congestion. Congestion effects in our study may be counterbalanced by already-enrolled children working harder.

vestment model. The model predicts four primary mechanisms. Under an assumption that roads lead to factor price equalization, a new road is likely to: (i) raise the unskilled wage and thereby increase the opportunity cost of schooling; (ii) raise the skill premium and thus increase the returns to education; (iii) increase lifetime household earnings (an income effect); and (iv) ease a liquidity constraint.[4] To test the importance of these mechanisms, we generate regional measures that would be expected to predict the size of each effect. We predict the size of opportunity cost and returns to education effects using regional urban-rural gaps in unskilled wages and Mincerian returns to education. To predict expected income and liquidity effects, we use a measure of asset poverty.[5] The variation in treatment effects across all three measures supports the standard human capital investment model. The opportunity cost effect is the most important predictor of treatment effects. Partitioning our data according to these measures, we find that market integration has negative effects on enrollment only in the 9% of villages where we expect opportunity cost effects to be high, and returns to education and income/liquidity effects to be low—exactly where the theory predicts treatment effects would be most negative. Even these negative effects are very small and statistically insignificant. Effects are positive and statistically significant in 39% of villages and positive but insignificant in the remaining 50% of villages.

We explore and rule out several other channels: (i) migration effects; (ii) supply-side improvements in school infrastructure; (iii) displacement effects among nearby villages; and (iv) improved access for children on the outskirts of villages. Consistent with earlier literature, we find no enrollment effects on primary-school children, for whom there is less scope for increased school enrollment and fewer opportunities for productive work.[6]

Our findings suggest that integrating the rural poor with regional markets has the poten-

[4]Because roads may cause simultaneous factor price changes in many markets, one can imagine other or opposite effects as well. We focus on the first order effects that are predicted either from the literature or from existing rural-urban price gaps.

[5]Income and liquidity effects are theoretically distinct but difficult to disentangle without detailed household-level data (Edmonds, 2006), so we consider them together.

[6]However, we do find small increases in primary-school performance, suggesting that students may be increasing school effort on the intensive margin.

tial to drive further long-run growth through increased educational attainment. Despite the low quality of schools in rural India (see ASER Centre (2014) for a summary), enrollment and exam performance respond positively to increased economic opportunities. Our results also provide context for the strong correlation around the world between education, growth, and trade.

This study is related to a growing literature on the impact of labor demand shocks on schooling decisions, which finds both positive and negative schooling impacts from new economic opportunities.[7] The heterogeneity in existing studies is well explained by the human capital model, because new opportunities can change the opportunity costs, long-run benefits of schooling and demand for schooling through income and liquidity effects. We show how treatment heterogeneity of a single intervention can be partitioned according to these mechanisms. Our work is also related to studies of the impact of India's national public works program (the National Rural Employment Guarantee Scheme, or NREGS) on human capital accumulation.[8] These studies are concerned with exogenous shocks to local labor markets, but none of them are directly informative about the effect of improving village access to already-existing nearby markets. Impacts of road connections are particularly policy-relevant, as the degree of market integration between villages and their nearby towns is a direct consequence of infrastructure investment policy.

Our paper also contributes to the literature on the development impacts of transport infrastructure.[9] Relative to earlier work on roads and schooling, our large village-level sample

[7] The opening of new outsourcing facilities in India and garment factories in Bangladesh have driven increases in schooling (Jensen, 2012; Oster and Steinberg, 2013; Heath and Mobarak, 2015). Positive agricultural demand shocks in India, expansion of natural gas fracking in the United States, and expanded export manufacturing in Mexico have increased dropout rates, especially for middle-school children and older children (Shah and Steinberg, 2017; Cascio and Narayan, 2015; Atkin, 2016).

[8] Studies on Andhra Pradesh find that access to the workfare program increases children's enrollment (Afridi et al., 2013) and test scores (Mani et al., 2014). All-India studies find increased enrollment for primary-school-aged children, but decreased enrollment for middle and high school children (Islam and Sivasankaran, 2014; Das and Singh, 2013; Li and Sekhri, 2015; Shah and Steinberg, 2015). NREGS increases demand for unskilled labor, and thus raises the opportunity cost of schooling; it is unlikely to increase returns to education, though it could have important income and liquidity effects.

[9] Some examples include Jacoby (2000); Jacoby and Minten (2009); Donaldson (n.d.); Gibson and Olivia (2010); Mu and van de Walle (2011); Donaldson and Hornbeck (2016); Casaburi et al. (2013). For a detailed review, including studies on the impacts or highways and regional roads, see Hine et al. (2016). In a working

and research design allow a more precise estimation of the causal effects of road construction. Finally, we contribute to a wide body of research on improving educational attainment in developing countries (see Glewwe and Muralidharan (2016) and Evans and Popova (2016) for reviews of this literature). Our results highlight that investments outside the education sector can have first order effects on schooling decisions.

This paper is organized as follows. Section II presents a conceptual framework describing human capital investment decisions and the role of market integration. Section III provides background on road construction and education in India. We describe the data in Section IV and the empirical strategy in Section V. Section VI presents basic results, Section VII explores the mechanisms suggested by the human capital model, and Section VIII concludes.

II Conceptual Framework: Schooling Decisions and Economic Opportunity

We outline a standard conceptual framework to help explain how human capital investment decisions respond to changes in labor market opportunities (Becker, 1954). This framework helps to reconcile why the impacts of labor demand shocks on schooling vary across the empirical literature, and motivates our later analysis of how roads' impacts on rural schooling decisions are affected by characteristics both of villages and of local labor market conditions outside the village.

The key decision point in the framework is the individual's tradeoff between the long-run benefits of human capital accumulation and the short-run return to labor. A two-period model is sufficient to highlight the essential comparative statics. In the first period, an agent chooses between working for a low-skill wage and obtaining schooling. In the second period, the agent works and receives either a high or a low wage, depending upon her schooling choice

paper, Mukherjee (2012) uses a regression discontinuity approach around population thresholds and finds that PMGSY increases school enrollment. We present comparable regression discontinuity estimates on middle-school enrollment in the robustness section, but we favor the panel estimates: they are an order of magnitude more precise and allow for analysis of treatment heterogeneity. Asher and Novosad (2017) show that PMGSY road construction leads to significant occupational change but has little effect on village assets, incomes or consumption, also using regression discontinuity. Using district-level data from India, Aggarwal (2017) finds an association between road construction and school enrollment. Khandker et al. (2009) and Khandker and Koolwal (2011) show that small-scale road construction in Bangladesh is associated with increased school enrollment.

in the first period. The agent consumes in both periods, drawing from an initial endowment and wages earned in each period that the agent works. The agent can save at some interest rate, but may be restricted in borrowing. The agent's initial endowment can reflect household wealth or wages of household adults who have completed their schooling. Education may also be a normal good, which households value independently of its impact on future wages.[10]

When a village is connected to an external market via a new road, the parameters underlying this tradeoff change. The first order effect of reduced transport costs is likely to be a change in prices due to factor price equalization. We focus first on the direct effect on wages. In equilibrium, urban areas have higher wages than rural areas for both unskilled and skilled workers, and higher Mincerian returns to education (see Appendix Table A1). Connecting a village to its external market is therefore likely to: (i) increase the unskilled wage; and (ii) increase the returns to education. We can think of all of these as changes in real wages, such that any changes in local goods prices due to new roads are subsumed in the above effects.[11] Wage convergence could come from permanent migration, temporary migration (*e.g.*, daily commuting to larger markets along new roads), or changes in factor prices due to goods market integration.[12]

An increase in the low-skill wage raises the opportunity cost of schooling and motivates agents to reduce human capital investments—we call this the opportunity cost effect. An increase in the high-skill wage raises the returns to education and motivates increased human capital accumulation—the returns to education effect. Changes in prices and access to opportunities could also cause income or liquidity effects. Income effects will increase the demand for schooling if schooling is a normal good. Higher liquidity may increase schooling

[10]This framework underlies much of the theoretical literature on child labor and human capital investment decisions. See, for example, Ranjan (1999) or Baland and Robinson (2000). We abstract away from intra-household bargaining, because it does not change our key predictions.

[11]It is possible that these static price differentials reflect unobserved differences in skills of workers in different locations, even controlling for education. For example, the quality of education in rural areas is probably lower than in urban areas. However, it is doubtful that unobserved education quality differences drive the entire differential, given the presence of higher skilled jobs in cities and towns, and the high returns to rural-to-urban migration documented in other studies, e.g. Bryan et al. (2014).

[12]Asher and Novosad (2017) show that the first order effect of new PMGSY roads on village economic structure is an increase in the number of people working for wages outside of villages.

if the return to schooling is high but families are credit constrained and cannot afford to either pay school fees or require children to perform household or market work. In principle, these effects could go in either direction, but based on urban-rural wage and skill gaps and existing empirical evidence from other studies, we expect the opportunity cost effect to reduce schooling, and the returns to education, income and liquidity effects to increasing schooling.

The predictions from factor price equalization in goods and capital markets are less straightforward, because many prices can change simultaneously. While the prices of intermediate goods used in village production are likely to fall, the prices of final goods produced in the village could move in either direction. Capital market integration could also lower the borrowing rate, easing liquidity constraints, or it could raise the return on savings, which would lower the relative returns to schooling.[13]

To understand which of these mechanisms are important, we identify places where individual effects are likely to be particularly small or large. We focus on the labor market channels as the existing literature on market integration suggests that they have first order effects on schooling decisions (Edmonds and Pavcnik, 2006; Edmonds et al., 2010; Shastry, 2012). Regional labor market conditions are plausibly good predictors of the sizes of the opportunity cost and returns to education effects, because regional markets are likely to dictate the magnitude of changes in skilled and unskilled wages when a village becomes integrated with that market. If roads lead to factor price equalization, the opportunity cost effect should be particularly large when the unskilled regional wage is much larger than the unskilled wage in the unconnected village. Similarly, the returns to education effect should be larger when regional returns to education are much larger than village returns to education. Income and liquidity effects on schooling would be expected to be larger in villages that are liquidity constrained or have low incomes; of course, the economic opportunities created by new roads may differ in these villages as well. In the absence of shocks that affect liquidity and not income, these last two effects are difficult to disentangle (Edmonds, 2006), so we consider them together.

[13] Roads could also lead to changes in information, marriage markets, and healthcare access, among other domains. We focus here on labor markets and leave these other domains for further research.

III Background and Details of the Road Construction Program

The study period (2002-2015) was a period of substantial education reform in India. Several programs were put into place with the explicit goal of increasing school participation, including a national drive supporting the goal of universal primary education under the flagship program *Sarve Shiksha Abhiyan* (Education for All). School enrollment increased substantially over this period, parallel to a similar global trend.

Both educational attainment and economic growth vary substantially across India. Indian policy-makers have long allocated public goods with an aim to mitigate spatial inequality, but large disparities remain and are at the center of public debate in India (Banerjee et al., 2007; Dreze and Sen, 2013). The high cost and poor durability of roads have constrained the ability of the government to connect every village: in 2001, 49 percent of Indian villages remained inaccessible by all-season roads. These villages were characterized by greater poverty and lower educational attainment.

In 2000, the Government of India launched the Pradhan Mantri Gram Sadak Yojana (Prime Minister's Road Construction Program, or PMGSY), a national program that aimed to eventually build a paved road to every village in India. The federal government issued implementation guidelines, but decisions on village-level allocations of roads were ultimately made at the district level. While one of the guidelines was a population-based eligibility criteria, in practice it was followed in only a subset of states, and even in these states it overlapped with several other eligibility criteria. The population threshold rule makes it possible to conduct a robustness check of the difference-in-difference estimates, and we discuss it in more detail in Section VI.C. Roads were targeted to habitations, which are the smallest rural administrative unit in India; a village is typically comprised of between one and three habitations.[14] We focus on villages as the unit of analysis, because (i) many villages have only one habitation; (ii) many habitations were pooled to the village level for the purposes of the program; and (iii) very little economic data is available at the habitation level. Given the

[14]There are approximately 600,000 villages in India and 1.5 million habitations.

program rules, early-treated villages tended to have larger population, but were not different from late-treated villages on other characteristics.[15]

At the outset, about 170,000 habitations in approximately 80,000 villages were eligible for the program, a number that has grown as the guidelines have been expanded to include smaller villages. By 2015, over 115,000 villages had access roads built or upgraded under the program. Construction projects were most often managed through subcontracts with larger firms, and were built with capital-intensive methods and external labor; the building of the road itself was therefore not a major local labor demand shock. These roads are distinct from new roads being built under the National Rural Employment Guarantee Scheme (NREGS), which were lower quality roads built with labor intensive methods.[16] Figure 1 shows the distribution of road construction by state and across time. The median road length was 4.4 kilometers; given the difficult terrain in many of these villages, a new road easily represents multiple hours saved on a daily trip out of the village.

IV Data

We constructed a village panel dataset, combining data on road construction with village characteristics and educational outcomes. We matched three successive Indian Population Censuses (1991, 2001, 2011) to an annual census of Indian schools, the District Information System for Education (DISE, 2002-2015), as well as the administrative data from the implementation of the road program (2001-2015). All data were merged primarily through fuzzy matching of location names, though in some cases unique identifiers were available for

[15] District fixed effects explain 30% of the variation in year of treatment among treated villages. A population quartic explains another 9% of the variation, after which inclusion of additional control variables has virtually no additional predictive power.

[16] We are aware of no other major rural road construction program in India during this period. Local or district administrators interested in road construction were more likely to lobby for PMGSY roads than allocate other funding to new roads. To the extent that sample villages received roads from NREGS or other sources during the sample period, it would bias our estimates toward zero. Major highway projects during this period, such as the Golden Quadrilateral, were planned and executed independently of PMGSY; there is no evidence of coordination of PMGSY roads with the construction of the Golden Quadrilateral or of other district road improvement projects.

subsets of the match.[17,18]

The DISE is an annual census of primary and middle schools in India. It includes data on student enrollment, exam completion, and school infrastructure. This dataset was created by the Ministry of Human Resource Development of the Government of India and is administered by the National University of Educational Planning and Administration. DISE data are considered to comprehensively cover every registered Indian government primary and middle school beginning in 2005.[19] We also have DISE data for a smaller sample of schools from 2002-2004, a period when the data-collection system was still being rolled out on a district-by-district basis. We are able to replicate national survey-based statistics on enrollment, suggesting that the DISE data are reliable.[20] The fact that DISE data are based on interviews with school headmasters raises a misreporting concern, but because new roads are likely to lower the cost of monitoring enrollment numbers, this would most likely bias treatment effects downward. Misreporting effects would also most likely affect data on all levels of enrollment, as well as on school characteristics (such as availability of computers), rather than only middle-school enrollment. These are thus unlikely to substantially bias our results.

Our primary outcome variable is log middle-school enrollment, which we define as the natural logarithm of one plus the total number of middle-school children enrolled in all schools in a village. As with the previous literature, we focus on outcomes for middle-school children (grades 6-8), both because there is little variation in dropout rates for younger children and

[17]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.

[18]We were able to match 83 percent of villages in the road administrative data to the population censuses, and 65 percent of villages in DISE. The match rate is worse for DISE because of frequent miscoding of census block identifiers in the DISE dataset. We matched 80 percent of census blocks; within census blocks, we matched 81 percent of villages.

[19]We refer to academic years (which begin in June or July) according to the beginning of the school year (i.e., we refer to academic year 2007-08 as 2007).

[20]We dropped enrollment observations from DISE that appeared to be erroneous. Our preferred sample drops all villages that reported total enrollment (first through eighth grades) greater than 60 percent of total population, which was the 99th percentile of this statistic. By comparison, in 2001 only 22.4 percent of the population was of primary- or middle-school age (ages 6-15). Demographic data from the Below Poverty Line Census (2002) suggests that fewer than 40 percent of village residents are between 6 and 15 years of age in 99 percent of villages. Our results are not materially changed by these decisions.

because younger children have fewer labor market opportunities. Further, the transition to middle school is a natural breakpoint in a child's schooling at which educational milestones are often measured. DISE does not report enrollment information for higher grades, nor does it report the total number of school-age children in a village, so we are unable to calculate enrollment rates. However, we can track total village population at 10-year intervals using the Population Census, allowing us to make indirect inferences about enrollment rates.

DISE collects information on examination outcomes in the set of states with terminal primary- and middle-school examinations. These are used for promotion decisions and completion verification. The information collected includes the number of students that appeared for the exam, that passed the exam, and that scored with distinction. Examination data are available for years 2004-2009. Finally, we use DISE data on school infrastructure, which describe the school-level presence of blackboards, electricity, sanitation facilities, water (by source), a playground, a library, a boundary wall, access to regular medical checkups, and access ramps, available from 2002-2011.

For data on road construction, we use the administrative records which are used to track and implement the PMGSY program, which we scraped from the government's public reporting portal for this program.[21] Road data are reported at either the village or habitation level; we aggregate these data to the village level. We define a village as having a paved road at baseline if any habitation in that village had a paved road. We define a village as receiving a new road by a given year if any habitation in the village received a new road before September 30 of the school year, which is the date on which DISE records enrollment numbers. We restrict our sample to villages that did not have a paved road in 2001, and we discard villages where roads were categorized as upgrades rather than as new roads. We further limit the primary analysis sample to villages that received new program roads between September 2003 and September 2015, so that we have enrollment data for at least one pre- and post-treatment year for each village. Appendix Figure A1 shows how we arrive at our

[21] At the time of writing, the Indian government's public reporting portal for PMGSY was hosted at http://omms.nic.in.

final sample of villages. Our main estimates are drawn from the 10,014 villages which built roads between 2003 and 2015. We find similar results when we broaden to an unbalanced sample (n=19,152), or include villages that never received PMGSY roads (n=112,475).

To calculate district-level rural and urban wages, which we use to measure the opportunity cost and returns to education effects, we use all individuals reporting wages from the 55th round of the NSS Employment and Unemployment Survey, undertaken in 1999-2000. Finally, we use data from the 1991, 2001 and 2011 Population Censuses of India, which include village population and other demographic data. We also use the 1998 rural Economic Census to generate village level control variables.

Table 1 shows summary statistics of villages at baseline. The enrollment dropoff at middle school is substantial: the average primary-school cohort has 36 children per year, while the average middle-school cohort has only 13 children.

V Empirical strategy

Our goal is to estimate the causal impact of roads on educational choices. Cross-sectional estimates of the relationship between a village's accessibility and schooling decisions are biased by the fact that villages that do not have access to paved roads are different from connected villages along many dimensions. They are likely to be smaller, have more difficult terrain, and be more politically marginalized. Our primary empirical specification is a panel fixed effect regression that exploits the timing of road construction, within the set of all villages that received new roads under the program by 2015.

The panel estimation exploits variation in the year that a village was connected to the road network. The panel estimator is defined by the following equation:

$$Y_{i,s,t} = \beta \cdot ROAD_{i,s,t} + \boldsymbol{\gamma}_{s,t} + \boldsymbol{\eta}_i + \epsilon_{i,s,t}. \tag{1}$$

$Y_{i,s,t}$ is the outcome variable (such as school enrollment), measured in village i and state s in year t. $ROAD_{i,s,t}$ is an indicator of whether the village has been connected by a paved road by year t. $\boldsymbol{\gamma}_{s,t}$ is a state-year fixed effect, and $\boldsymbol{\eta}_i$ is a village fixed effect. The error term, $\epsilon_{i,s,t}$, is clustered at the village level to account for serial correlation in the dependent variable. β is the coefficient of interest and measures the impact of a new road on village-level enrollment. All villages have $ROAD_{i,2002} = 0$ and $ROAD_{i,2015} = 1$, i.e., all sample villages received a road at some point under the program between 2003 and 2015. We thus avoid making a potentially biased comparison between villages that were and were not eligible for new roads. Unless otherwise specified, the outcome variable is the natural logarithm of one plus enrollment, so impacts can be interpreted as percentage changes.

The state-year fixed effects control flexibly for differential enrollment growth across states. This alleviates any concern that states with more effective governments simultaneously built roads and also provided other government services; it also controls for any broader regional trends in enrollment that might be correlated with road construction. The village fixed effects control for systematic differences between early- and late-treated villages. No additional controls are included, because the village fixed effects account for all static village characteristics, and we do not have annual data on any time varying characteristics of villages other than school enrollment. We also present specifications that control for village time trends and for baseline village characteristics interacted with year fixed effects. The panel estimates can be interpreted as causal effects under the assumption that the only changes that occurred in a village at the time that a road was built were changes caused by that road, after partialling out state-year and village time trends where applicable.

VI Results

VI.A Average Impacts on School Enrollment

Table 2 shows estimates of the effect of road construction on village school enrollment, using Equation 1. Column 1 shows the balanced panel estimate from the main sample of villages, which were unconnected at baseline and received a road between 2003 and 2015. The estimate implies that a new road leads to a seven percent increase in middle-school enrollment.

The estimate is statistically significant, with a p-value less than 0.001. Given the sample mean of 39 students enrolled in middle school, this corresponds to approximately three additional students in middle school, on average 3.7 years after a road is built.[22] In Columns 2 and 3 of Table 2, we split the main result into enrollment of girls and boys respectively, for whom results are nearly identical. Columns 4 through 6 show comparable estimates using the level of middle-school enrollment as the dependent variable rather than log enrollment. Results are consistent with the log estimates.

To fully describe the time series of enrollment before and after a new road is built, we estimate log middle-school enrollment on a set of relative time dummies, which indicate the number of years before or after treatment of a given observation. The estimating equation is:

$$Y_{i,s,t} = \sum_{\tau \in (-5,+5), \tau \neq -1} \boldsymbol{\zeta}_\tau \big(\mathbb{1}\big(t = t_{i,s}^{treatment} + \tau\big)\big) + \boldsymbol{\gamma}_{s,t} + \boldsymbol{\eta}_i + \epsilon_{i,s,t}, \tag{2}$$

where τ indicates the year relative to when a new road was built, i.e., $\tau = -1$ is the year before road construction. State-year and village fixed effects are included as above.

Borusyak and Jaravel (2017) show that event study designs where all groups are eventually treated can be identified only up to a linear trend in relative time. For instance, an upward linear trend in enrollment could either be described by Equation 2 with linearly increasing time fixed effects, or with linearly increasing relative time effects. In this setting, we can therefore identify trend breaks, but cannot test either average trends or pretrends. Effectively, this means the regression above can only be estimated with *two* relative time coefficients omitted.[23] We follow the suggestion of Borusyak and Jaravel (2017) and omit the relative time coefficients indicating the year before treatment and the first year available. We plot the remaining τ coefficients in Figure 2. The graph confirms that the enrollment

[22] The estimate is thus a weighted difference between enrollment in all treated years and enrollment in untreated years. Estimating a weighted linear combination of relative treatment time dummies according to Borusyak and Jaravel (2017) delivers a very similar treatment estimate of 0.06.

[23] McKenzie (2006) makes a similar point by arguing that without normalization, only second differences in relative time effects can be identified. The standard difference-in-differences specification (Equation 1) has an implicit normalization with zero pretrend.

increase corresponds to the timing of the construction of the new road, and appears to be persistent. The F-test of the pre-treatment coefficients (which tests for non-linear pretrends) is insignificant (p=0.94). The timing and persistence of the change in enrollment also makes it unlikely that treatment effects are driven by labor demand on the actual road construction project; if effects were driven by work on the road itself, we would expect to see changes before the road was built and disappearing rapidly thereafter.

VI.B Robustness: Sample Definition and Specification

Table 3 shows that the average estimated enrollment effect is robust to a range of empirical specifications and sample definitions. To test whether treatment effects could be driven by different trends across early- and late-treated villages, we include village-specific linear time trends (Column 1), and interactions between year fixed effects and baseline village population, share of irrigated land, number of schools, log middle- and primary-school enrollment, literacy rate, population share of scheduled castes, and distance to nearest town (Column 2). Results are not substantively changed by these inclusions. Columns 3 expands the sample to an unbalanced panel by including villages with missing data in one or more years, and Column 4 shows the unbalanced panel with village time trends. In Column 5, we restrict the data to years after 2004, when the DISE data have the highest coverage of villages and schools. Column 6 restricts the sample to a set of villages for which we have four observations before and four observations after the completion of road construction; the sample is limited to those observations, thus providing nine observations per village. The estimates are all highly similar in magnitude and statistical significance. Given that all these specifications permit differential functional forms of time-variant village characteristics, the stability of the treatment effect strongly suggests that these estimates are not driven by different types of villages being treated at different times.[24] Appendix Table A2 repeats Table 2 with district-by-year fixed effects; estimates are substantively unchanged.

[24]We use village time trends as a robustness check, rather than in the main specification, because of the possibility that the time trends in part pick up the effects of the new road over time (Wolfers, 2006). This said, all results presented below are similarly unchanged by inclusion of village time trends.

To verify that p-values are estimated correctly, we run a randomization test. In the spirit of the Fisher Randomization Test, we randomly generate a placebo year of road completion for each village, and then estimate Equation 1 as if the placebo year were the treatment year. We run this estimation 1000 times; Appendix Figure A2 shows the distribution of β, the placebo impacts of a new road on log middle-school enrollment growth. This gives us a non-parametric distribution of test statistics under the sharp null hypothesis, with existing data. The placebo estimates are centered around zero and, consistent with Table 2, none of the thousand estimates attains our primary estimate of the effect of a new road on log enrollment (0.07 increase in log enrollment).

Columns 1-3 of Appendix Table A3 show comparable estimates for primary-school enrollment. Consistent with much of the prior work on labor market impacts of schooling, we find no effects on younger children. This is not that surprising, given that India had almost achieved universal primary completion in this period and that children under the age of twelve have few labor market opportunities.

VI.C Robustness: Regression Discontinuity

In this section, we present regression discontinuity estimates of the impact of new roads on schooling (Lee and Lemieux, 2010). Under the program guidelines, states were instructed to first target villages with populations greater than 1000 in the population census, and then villages with population greater than 500. Only some states followed these guidelines, and even then, each followed the rules to different degrees, in part because there were often several conflicting guidelines.[25] In states where there were few unconnected villages with populations over 1000, they tended to use the 500-person threshold immediately. In most states, construction proceeded in villages both above and below the population threshold simultaneously, but there were more villages treated above the threshold, and these were

[25] For example, under certain circumstances, proximate habitations could pool their populations to exceed this cutoff; we do not observe where this took place. We met several times with the National Rural Roads Development Agency, the national coordinating body for the program, to identify the set of states that adhered to program guidelines and which eligibility thresholds were used. The states in the sample are Chhattisgarh, Gujarat, Madhya Pradesh, Maharashtra, Odisha and Rajasthan.

treated sooner. Population above a treatment threshold is therefore an imperfect predictor of program treatment status. Figure 3 shows the relationship between the share of unconnected villages that received new roads before 2011 and the population relative to the treatment threshold. The change in treatment status at the population threshold is clear. There is no discontinuous change in the density of villages on either side of the cutoff, nor in characteristics of villages prior to road construction.[26]

We estimate the impacts of road construction using the following implementation of a local linear estimator:

(3)

$$\ln(Y_{i,s,t}) = \gamma_1 1\{pop_{i,s,2001} - T \geq 0\} + \gamma_2 (pop_{i,s,2001} - T) +$$
$$\gamma_3 (pop_{i,s,2001} - T) * 1\{pop_{i,s,2001} - T \geq 0\} + \gamma_4 \ln(Y_{i,s,2002}) + \boldsymbol{\lambda} X_{i,s,2001} + \eta_s + \upsilon_{i,s}.$$

$Y_{i,s,t}$ is log enrollment in village i, region s at time t, T is the population threshold, $pop_{i,s,2001}$ is baseline village population (the running variable), $X_{i,s,2001}$ is a vector of village controls measured at baseline, and η_s is a region fixed effect.[27] The change in the outcome variable across the population threshold T is captured by γ_1. The population controls allow for different slopes on either side of the treatment threshold. We limit the sample to populations close to the treatment threshold, using an optimal bandwidth calculation (Imbens and Kalyanaraman, 2012).

Panel A of Figure 4 shows first stage and reduced form regression discontinuity estimates for all sample years. The first stage estimates show that the population threshold rule begins to be applied around 2007 and stabilizes in importance from 2011 to 2015, during which years

[26]To test this formally, we fit a non-parametric function to the village population distribution, with allowance for a discontinuity at the treatment threshold (McCrary, 2008); the p-value testing the null of no discontinuity is 0.31. Appendix Figure A3 presents the population histogram and the graphical rendering of the McCrary Test. Appendix Table A4 and Figure A4 present RD estimates and and graphs showing that baseline village covariates do not vary systematically at the treatment threshold.

[27]For control variables, we include baseline log enrollment, the literacy rate, number of primary schools, number of middle schools (all from the 2001 Population Census), and the log number of non-farm jobs in the village (from the 1998 Economic Census).

villages just above the threshold are 20-25 percentage points more likely to have received new roads. The reduced form estimates on log middle school enrollment follow a similar pattern, ramping up in 2007 and stabilizing in 2011. To maximize power, we estimate the regression discontinuity on the pooled set of enrollment estimates from 2011 to 2015, clustering Equation 3 at the village level to account for serial correlation. Panel B of Figure 4 plots log middle school enrollment as a function of population relative to the treatment threshold; the increase in enrollment just above the treatment threshold is clear.

Panel A of Table 4 presents regression discontinuity estimates for the pooled 2011-2015 sample.[28] Column 1 reports the first stage estimate, where the dependent variable is a village-level indicator equal to one if a village received a road. 33% of villages in the sample received new roads; a village just above the population treatment threshold is 24 percentage points more likely to receive a new road. Column 2 reports the reduced form estimate of the impact of crossing the population threshold on village-level log middle-school enrollment; Column 3 presents the IV estimate. The estimates suggest large positive treatment effects, albeit just outside of conventional thresholds for statistical significance (p=0.103).

As a placebo test, we run the same empirical specification on the states that did not follow program guidelines.[29] Panel B of Table 4 shows that there is no substantive first stage in these states (Column 1), and encouragingly, a reduced form treatment effect (Column 2) close to zero. This provides reassurance that there is not some other characteristic of villages above the population threshold that caused their schooling to grow.[30]

The RD estimate is considerably larger than the panel estimate, but it is also substantially less precise. While it is possible that the local average treatment effect of roads on enrollment for RD complier villages is substantially higher than for villages in the diff-in-diff sample, we

[28] Figure A5 shows RD treatment estimates for each year from 2010 to 2015 as well as the pooled 2011-2015 estimate, under the optimal bandwidth and alternate bandwidths that are 25% higher and lower.

[29] Major states that built roads under PMGSY but did not follow program guidelines include Andhra Pradesh, Assam, Bihar, Uttar Pradesh, and Uttarakhand.

[30] Columns 4 and 5 of Appendix Table A3 show analogous RD estimates of log primary-school enrollment growth. Consistent with the differences-in-differences estimates, the RD indicates no change in primary-school enrollment.

note that the 95% confidence interval of the RD estimate includes the panel estimate, and we are hesitant to put a large weight on the specific point estimate.[31]

The regression discontinuity estimates corroborate the results from the main panel specification, indicating higher middle-school enrollment following road construction. The strength of the regression discontinuity approach is its reliance on few assumptions for causal inference, but the power of the test is limited by imperfect compliance, as well as the restriction of the sample to villages close to threshold populations in states that followed the allocation rules. These factors reduce the precision of the estimates and make them less representative of impacts across India. We therefore focus on the panel setting to examine how the impacts of road connections vary in response to local labor market conditions.

VI.D Average Impacts on School Achievement

Increasing school enrollment may not directly translate into increasing human capital, especially if school quality is low or if there are congestion effects. We turn to exam scores as a measure of what students are actually learning. Table 5 presents panel estimates of the impact of new roads on a set of dependent variables describing students' exam-taking decisions and exam performance. We focus on middle-school completion exams, which were required if students were to go on to high school. Column 1 estimates the effect of roads on the log number of students who appear for completion exams plus one. Column 2 estimates effects on the number of students who pass the exam, and Column 3 shows effects on the number who pass with distinction.[32] For exam appearance and passing, we find similar effects to the enrollment effects: six percent more students take and pass exams in villages after new roads have been built. We find a positive but smaller three percent increase in those passing with distinction. While the percentage effects are similar, the number of students achieving these outcomes is smaller than the enrollment effects, because for every ten students enrolled in the 8th grade,

[31] Variation in village size and state in the panel estimates does not explain this difference. However, we find subgroup estimates approaching these numbers in villages where district wage and return gaps suggest treatment effects should be particularly large (see Section VII).

[32] Sample size is smaller for the exam estimates than for enrollment estimates because we were only able to obtain examination results for years 2004-2009. Results are highly similar for the unbalanced panel.

only six appear for the exam, five pass the exam, and two pass the exam with distinction.

The impacts on examinations reflect the net impact on achievement and can be interpreted in two ways. The first possibility is that the students induced to stay in school take and pass exams at the same rate as non-marginal students (but receive slightly fewer top grades), and there are no effects on the exam performance of non-marginal students. Alternately, the marginal students who were induced to stay in middle school could do worse on exams (perhaps because there may be negative selection (in terms of ability) into the group of students on the margin of not dropping out), but students who would have stayed in school independent of road construction are now performing better on exams. The latter could occur if non-marginal students perceive that human capital accumulation is more valuable given increased access to external markets. Without data on individual student performance, it is difficult to disentangle these two scenarios. However, in both cases we can reject the possibility that enrollment is increasing but learning is unchanged. Rather, the exam data show that the total stock of human capital in connected villages is increasing.

Appendix Table A5 shows comparable results for primary-school completion exams. In contrast with the zero enrollment effects in primary school, here we find weakly positive results with estimates between 2 and 3 log points, albeit with marginal statistical significance. The p-values for exam taking, passing, and scoring with distinction, are respectively 0.07, 0.18 and 0.15. Given unchanged enrollment in primary school, this implies improved performance among enrolled children. This could arise directly from future labor market returns to education, or because a number of these students newly anticipate attending middle school.

VII Mechanisms

VII.A Human Capital Investment Incentives

In this section, we examine the mechanisms underlying the estimated impact of new rural roads on human capital accumulation. The conceptual framework outlined in Section II guides our analysis. We are interested in three primary channels: a negative opportunity cost effect, a positive returns to education effect, and an income/liquidity effect. We pool the income and liquidity effects, because it is difficult to distinguish between the effects of higher lifetime income and higher cash in hand, given the available data. Our goal is to identify subsets of the sample where each of these mechanisms is likely to be particularly important.

To predict these mechanisms, we assume that reductions in transportation costs will lead to factor price equalization: when a rural village receives a new road, its wages and returns to education will adjust toward the wages and returns in the broader geographic area. If the unskilled wage gap between the village and surrounding market is high, the village unskilled wage will rise more than if the unskilled wage gap is small. We therefore expect the largest opportunity cost effects in the places with the largest gaps in unskilled wages between the village and its surrounding market. We proxy the expected size of the opportunity cost effect with the district-level urban-rural wage gap, the most granular level at which wages can be calculated. Urban and rural wages for this calculation are drawn from the 55th round of the National Sample Survey (NSS), undertaken in 1999-2000, the last NSS round before any PMGSY roads were built.

To proxy for the size of the returns to education effect, we again aim to identify the difference in returns to education between each village and its regional market. The underlying assumption remains that a new road will shift village returns to education in the direction of equalization with the regional market. We calculate district-level returns to education in the 55th round NSS by running Mincerian regressions at the district level, separately for individuals in rural and urban areas. We call this difference the urban-rural returns gap, or the skill premium gap.[33] We assume the returns to education effect is stronger when the skill premium gap is higher.

Finally, to proxy for the importance of income/liquidity effects, we assume that households with few assets are more likely to be liquidity constrained, and that a given change in wages

[33] Specifically, in each district we regress log wage for working individuals on years of education, age, age squared, and the log of household land owned, separately for urban and rural locations. Mincerian returns are minimally affected by alterations to this specification, such as excluding log land or including state fixed effects. We drop districts with no urban data.

for these households has a larger income effect. We use the 2002 Below Poverty Line Census to measure average baseline assets at the village level. We define a village as having low assets (and hence high potential income/liquidity effects) if the share of households reporting zero durable assets on the survey is above the sample median.[34] We similarly convert the opportunity cost and returns to education measures into binary indicators.

We then estimate the panel regression, interacting the treatment indicator with these binary mechanism proxies. If the interaction term is important in magnitude, it provides suggestive evidence that the relevant mechanism is an important channel through which new roads affect schooling decisions. Table 6 shows the results. Column 1 repeats the main specification without interaction terms in the set of data for which the interaction terms are non-missing.[35] Columns 2 through 4 include the interaction terms separately, while Column 5 includes all three. The direction of the interaction estimates is consistent with the predictions from a standard human capital model. Roads have the smallest effects on schooling in districts where they would be expected to raise the opportunity cost the most and the largest effects in districts where they would be expected to raise the skill premium the most and to have the largest income and liquidity effects. The opportunity cost effect is strongly statistically significant ($p < 0.01$); the p-values for the returns to education and income/liquidity effects are respectively $p = 0.08$ and $p = 0.37$ in the joint specification.[36] The greater magnitude of the opportunity cost effect could be in part because the urban-rural wage gap is much larger than the urban-rural skill premium gap (see Appendix Table A1). While the estimated interaction effects are consistent with a standard human capital investment model, note that there could be other unobserved district-level characteristics that influence the size of treatment effects, which could be correlated with the proxies that we use. Therefore, we see these estimates not as definitive but as suggestive indications of the mechanisms underlying the main estimates. For instance, high rural-urban wage gaps

[34]The surveyed assets are a radio, a television, a telephone, and a motorcycle.

[35]We drop all districts without NSS data for both urban and rural areas in 1999-2000.

[36]For completeness, Appendix Table A6 shows results by quartile of each mechanism proxy.

are correlated with greater remoteness, worse infrastructure, lower returns to education and tend to be in the North. However, the interaction effects described here are robust to the inclusion of interactions with these other variables.

If we fully interact the three binary mechanism variables, we can obtain treatment effects in eight theory-based partitions of the sample. Table 7 shows linear combinations of treatment coefficients that describe the treatment effect in each subgroup, from the fully interacted regression. The point estimate is negative (but small and statistically insignificant) only in the partition with a high opportunity cost, low returns to education, and low potential income/liquidity effects—which is precisely the group where theory predicts roads may have adverse effects on education. Treatment effects are positive and significant only in partitions where at least two of the mechanisms are favorable. Overall, point estimates of treatment effects are negative in 9% of villages, and above 0.05 (and statistically significant) in 39% of villages.

The treatment heterogeneity is consistent with the heterogeneity in results from earlier work on impacts of labor demand shocks on school enrollment. Jensen (2012) and Oster and Steinberg (2013) find that increasing availability of call center jobs lead to increased schooling. The first order mechanism in these studies is likely an increase in the return to schooling, since spoken English is a requirement for these jobs. Conversely, Shah and Steinberg (2017) find that children are more likely to attend school in drought years, when agricultural labor market opportunities are few. The opportunity cost effect is likely to be first order: agricultural labor opportunities (or children's substitution into home production while parents are working) do not require a high level of schooling, thus the effective low skill wage is rising. The negative effects on schooling of India's national workfare program (NREGS) (Islam and Sivasankaran, 2014; Das and Singh, 2013; Li and Sekhri, 2015; Shah and Steinberg, 2015) are plausibly driven by a similar effect. Because NREGS hires people for labor intensive public works, it increases the return to low skill work without affecting the return to education. The heterogeneity in impacts of labor demand shocks outside of

India (e.g., fracking jobs in the United States (Cascio and Narayan, 2015), export manufacturing jobs in Mexico (Atkin, 2016), and garment manufacturing Bangladesh (Heath and Mobarak, 2015)) support the same model: individual schooling choices appear to respond to the schooling requirements of immediate labor market opportunities.

The heterogeneity of economic opportunites across India allow us to identify both large positive effects in the places where the relative return to *high* skill work goes up the most, and neutral to weakly negative effects on schooling in places where the relative return to *low* skill work rises the most. But the fact that treatment effects are negative (and small) in only a small share of villages is a striking result given the number of recent studies finding adverse impacts of new labor market opportunities.

VII.B Other Mechanisms: Migration, School Quality, Regional Displacement, and School Accessibility

Our results indicate that new roads cause increases in middle-school enrollment. The treatment heterogeneity suggests that the skill profile of labor demand outside the village may be a primary factor explaining these impacts. In this section, we explore several alternate mechanisms.

Migration. First, we explore whether net migration into treated villages (or reduced out-migration) can explain increases in middle-school enrollment. We present two indications that new roads did not substantially affect net migration in treated villages. First, we use the regression discontinuity specification to show that village population is not affected by road construction.[37] Panel B of Appendix Figure A6 presents the regression discontinuity graph, which shows no discontinuity at the treatment threshold and the point estimate is close to zero. We can rule out with 95% confidence the net entry or exit of more than four people from a treated village. Second, migration effects would be expected to equally affect families with primary-school-aged children. As discussed above, Appendix Table A3 and Panel B of

[37]Village population is measured only in the decennial censuses, so we cannot use the panel approach to measure impacts on migration.

Appendix Figure A6 show zero estimates on changes in primary-school enrollment. There is thus little evidence that net migration explains the effects of roads on school enrollment.[38]

School Quality. The estimated impacts on schooling could also be influenced by changes in school quality or in the number of schools available. Appendix Table A7 shows estimated effects of road completion on school quality, as proxied by a series of school infrastructure measures included in the DISE data, as well as the number of schools reported in DISE. While a minority of specifications show statistically significant effects on school infrastructure, none approach the size of the enrollment effects presented above. Adjusting these estimates for multiple hypothesis testing would further weaken the case for detectable impacts of roads on school infrastructure or school quantity. Thus, it does not appear that school enrollment effects are driven by changes in school quantity or quality.[39]

Other Government Programs. To alleviate any concern that other government programs could have been using the same eligibility criteria as the road program, or simultaneously implemented other programs along with roads, we use the regression discontinuity approach to test for appearance of other public goods in treated villages.[40] Appendix Table A8 shows RD estimates of changes in other village-level public goods. The estimates are all close to zero and none are statistically significantly different from zero. It does not appear that schools, electricity, health centers, or banks were delivered simultaneously to new roads.

Regional Displacement. We next explore whether our results could be driven by displacement effects, in which enrollment increases would be counterbalanced by declines in enrollment in nearby villages. This mechanism seems less plausible, as villages with no road by 2001 tended to be poor and remote, and thus unlikely to have more desirable schools

[38]Since the dependent variable is gross enrollment (rather than an enrollment rate), outmigration of students with a high propensity to drop out could not drive our estimates. Whether students not in school remain in the village or not does not directly affect enrollment figures.

[39]We find similar estimates if we weight the school infrastructure variables by the number of students attending the school, to reflect the share of children in a village who benefit from a particular kind of infrastructure investment.

[40]The differences-in-differences specification is not available here, because we observe village public goods only in the decennial population census.

than their neighbors. Nevertheless, we calculate total annual middle-school enrollment for all other villages within a small radius of each village that received a new road. Columns 1 and 2 of Appendix Table A9 report panel estimates of the impact of roads on log middle-school enrollment in surrounding villages, respectively within a 3 km and a 5 km radius.[41] We find precise zero impacts on these nearby villages, indicating that displacement effects are unlikely to explain the main findings.

School Accessibility. Finally, we examine the possibility that a new road impacts schooling by increasing accessibility to the school itself. For example, Muralidharan and Prakash (2017) find that the provision of bicycles made girls more likely to attend middle and high school. Children usually walk to village schools, and paved roads could make them easier to access, especially during the rainy season. We explore this hypothesis in two ways. First, we estimate the impact of roads on schooling in villages that are more or less dispersed. Children living in dispersed villages have further to walk to school, and thus might be expected to benefit more from a new road. We proxy village dispersion with surface area, and divide the sample into villages with above- and below-median surface area. Columns 3 and 4 of Appendix Table A9 show that treatment effects are similar in dispersed and dense villages.[42] Second, if treatment effects are driven by ease of access to village schools, we might see larger effects in places where there are nearby villages *without* middle schools—the road could make it easier for children from a different village to access middle school. To test this, we counted the number of school-age children within a 5 km radius of sample villages, who were living in villages without middle schools.[43] Columns 5 and 6 of Appendix Table A9 show that treatment effects are similar across villages close to more or less under-served children. The evidence does not support children's improved ability to walk to school as a primary mechanism for the impact of new roads.

[41] The average Indian village has a diameter of 2.1 km, and the average road built through this program had a length of 4.4 km.

[42] Results are similar if we use area per capita.

[43] We proxied the number of middle-school-aged children with the number of children aged 0-6 in 2001, the closest estimate available from the Population Census. We find similar results if we use total village population in villages without middle schools.

VIII Conclusion

High local transportation costs are a central feature of the lives of the rural poor around the world, leaving them isolated from external markets. Connecting remote villages to high quality transportation networks is a major goal of both governments of developing countries and development agencies. These roads can bring access to new opportunities; however, a concern may be that access to opportunity can paradoxically cause decreased investment in the human capital accumulation that is central to long-run growth.

We shed light on this question by studying the impact of India's flagship rural road program, which has built feeder roads to 115,000 villages in India between 2001 and 2015. We show that the building of these roads had large positive effects on adolescent school enrollment and performance. Our results suggest that the standard human capital investment model remains a powerful predictor of schooling decisions in developing countries. Predicted changes in the opportunity cost of schooling, the returns to education and the size of income and liquidity effects are all predictors of the local impacts of roads on schooling. But we find negative treatment effects only in the 9% of villages where treatment effects are predicted to be the lowest, and the negative effects here are small and statistically significant. We find strong positive treatment effects in 39% of villages, and positive but statistically insignificant effects of new roads in the remaining 52% of villages. The broad conclusion that local market integration substantially promoted schooling across much of rural India.

This paper highlights an understudied but important impact of rural infrastructure. Road investments are usually premised on their potential to bring economic opportunities and growth to rural areas. If road construction leads to increased human capital accumulation, then its long-run economic impact is likely much larger than short-run estimates suggest.

References

Adukia, Anjali, "Sanitation and education," *American Economic Journal: Applied Economics*, 2017, *9* (2).

Afridi, Farzana, Abhiroop Mukhopadhyay, and Soham Sahoo, "Female Labour-Force Participation and Child Education in India: The Effect of the National Rural Employment Guarantee Scheme," 2013. IZA Discussion Paper No. 6593.

Aggarwal, Shilpa, "Do Rural Roads Create Pathways out of Poverty? Evidence from India," 2017. Working Paper.

ASER Centre, *Annual Status of Education Report (Rural)* 2014.

Asher, Sam and Paul Novosad, "Rural Roads and Structural Transformation," 2017. Working paper.

Atkin, David, "Endogenous Skill Acquisition and Export Manufacturing in Mexico," *American Economic Review*, 2016, *106* (8), 2046–2085.

—— **and Dave Donaldson**, "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs," 2015. NBER Working Paper No.21439.

Baland, Jean-Marie and James A. Robinson, "Is Child Labor Inefficient?," *Journal of Political Economy*, 2000, *108* (4), 663–679.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden, "Remedying Education: Evidence from Two Randomized Experiments in India," *The Quarterly Journal of Economics*, aug 2007, *122* (3), 1235–1264.

Becker, Gary, *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, New York: Columbia University Press, 1954.

Behrman, Jere R, Susan W Parker, and Petra E Todd, "Medium-Term Impacts of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico," in "Poverty, Inequality, and Policy in Latin America" 2008, pp. 219–270.

Borusyak, Kirill and Xavier Jaravel, "Revisiting Event Study Designs," 2017. Working Paper.

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

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

Cascio, EU and A Narayan, "Who needs a fracking education? the educational response to low-skill biased technological change," 2015. NBER Working Paper No.21359.

Das, Shreyasee and Abhilasha Singh, "The Impact of Temporary Work Guarantee Programs on Children's Education: Evidence from the Mahatma Gandhi National Rural Employment Guarantee Act from India," 2013. Working Paper.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review (forthcoming).*

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

Dreze, Jean and Amartya Sen, *An Uncertain Glory: India and its Contradictions*, Oxford University Press, 2013.

Edmonds, Eric, Nina Pavcnik, and Petia Topalova, "Trade Adjustment and Human Capital Investments: Evidence from Indian Tariff Reform," *American Economic Journal: Applied Economics*, 2010, *2* (4).

Edmonds, Eric V., "Child labor and schooling responses to anticipated income in South Africa," *Journal of Development Economics*, 2006, *81* (2), 386–414.

——— **and Nina Pavcnik**, "International Trade and Child Labor: Cross-country Evidence," *Journal of International Economics*, 2006, *68* (1), 115–140.

Evans, David and Anna Popova, "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews," *The World Bank Research Observer*, 2016, *31* (2).

Gibson, John and Susan Olivia, "The Effect of Infrastructure Access and Quality on Non-farm Enterprises in Rural Indonesia," *World Development*, 2010, *38* (5), 717–726.

Glewwe, Paul and Karthik Muralidharan, "Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Handbook of the Economics of Education*, Vol. 5, Elsevier, 2016.

Heath, Rachel and A. Mushfiq Mobarak, "Manufacturing Growth and the Lives of Bangladeshi Women," *Journal of Development Economics*, 2015, *115*, 1–15.

Hine, John, Masam Abedin, Richard Stevens, Tony Airey, and Tamala Anderson, *Does the extension of the rural road network have a positive impact on poverty reduction and resilience for the rural areas served? If so how, and if not why not? A Systematic Review*, London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London, 2016.

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

Islam, Mahnaz and Anitha Sivasankaran, "How does child labor respond to changes in adult work opportunities? Evidence from NREGA," 2014. Working Paper.

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.

Jensen, Robert, "Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India," *The Quarterly Journal of Economics*, 2012, *127* (2), 753–792.

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.

Lee, David and Thomas Lemieux, "Regression discontinuity designs in economics," *Journal of Economic Literature*, 2010, *48* (2), 281–355.

Li, Tianshi and Sheetal Sekhri, "The Unintended Consequences of Employment-Based Safety Net Programs," 2015. Working Paper.

Mani, Subha, Jere R Behrman, Shaikh Galab, and Prudhvikar Reddy, "Impact of the NREGS on Schooling and Intellectual Human Capital," 2014. Population Studies Center Working Paper.

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., "Disentangling Age, Cohort, and Time Eï¬€*ectsintheAdditiveModel,″ OxfordBulletinofEconomicsandStatistics*, 2006, *68* (4).

Miguel, Edward and Michael Kremer, "Worms: Identifying Impacts on Education and

Health in the Presence of Treatment Externalities," *Econometrica*, 2004, *72* (1).

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

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

Muralidharan, Karthik and Nishith Prakash, *American Economic Journal: Applied Economics*, 2017, *9* (3).

Oster, Emily and Bryce Millet Steinberg, "Do IT Service Centers Promote Enrollment? Evidence from India," *Journal of Development Economics*, 2013, pp. 1–36.

Ranjan, Priya, "An economic analysis of child labor," *Economics Letters*, 1999, *64* (1), 99–105.

Shah, Manisha and Bryce Millett Steinberg, "Workfare and Human Capital Investment: Evidence from India," 2015. NBER Working Papers Series No. 21543.

____ **and** ____ , "Drought of Opportunities: Contemporaneous and Long Term Impacts of Rainfall Shocks on Human Capital," *Journal of Political Economy*, 2017, *125* (2).

Shastry, Gauri Kartini, "Human Capital Response to Globalization: Education and Information Technology in India," *Journal of Human Resources*, 2012, *47* (2), 287–330.

Wolfers, Justin, "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results," *American Economic Review*, 2006, *96* (5), 1802–1820.

Table 1

Summary Statistics at Baseline

	Mean (SD)
Population (2001 Census)	1291.4
	(998.3)
Non-farm Employment (1998 Economic Census)	60.1
	(173.8)
Number of Primary and Middle Schools	1.7
	(2.0)
Total Enrollment (grades 1-8)	217.1
	(389.0)
Total Primary Enrollment (grades 1-5)	178.0
	(286.8)
Total Middle Enrollment (grades 6-8)	39.1
	(125.6)
Middle School Exam Passers (2005)	7.3
	(15.4)
Exam Passers with Distinction (2005)	1.5
	(5.4)

The table shows means and standard deviations (in parentheses) of key variables at baseline, in the sample of villages that were matched across all analysis datasets. Unless otherwise indicated, the data source is the District Information System for Education (DISE), 2002.

Table 2

Impact of New Roads on Middle-School Enrollment

Dependent Variable	All, log	Girls, log	Boys, log	All, levels	Girls, levels	Boys, levels
	(1)	(2)	(3)	(4)	(5)	(6)
New Road	0.070***	0.060***	0.056***	2.558***	1.331***	1.227***
	(0.015)	(0.012)	(0.013)	(0.537)	(0.287)	(0.284)
N	146678	146678	146678	146678	146678	146678
r2	0.80	0.81	0.80	0.79	0.77	0.78

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

The table reports panel estimates of the effect of new road construction on village-level log middle school enrollment, estimated with Equation 1. Column 1 presents the primary balanced panel specification. The dependent variable in Columns 2 and 3 is log middle-school enrollment for girls and boys respectively. Columns 4-6 repeat these three specifications, using the level of middle-school enrollment as the dependent variable. All specifications have state-year fixed effects and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table 3

Impact of New Roads on Middle-School Enrollment:
Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
New Road	0.058***	0.058***	0.086***	0.078***	0.053***	0.041***
	(0.012)	(0.014)	(0.013)	(0.013)	(0.013)	(0.009)
State-Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Village F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Village Time Trends	Yes	No	No	Yes	No	No
Baseline Vars * Year Dummies	No	Yes	No	No	No	No
Panel Sample	Balanced	Balanced	Unbalanced	Unbalanced	Balanced Post-2004	4 Years Pre/Post
N	146678	142748	237281	237281	115247	148910
r2	0.91	0.83	0.76	0.88	0.87	0.84

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

The table reports panel estimates of the effect of new road construction on village log middle-school enrollment, estimated with Equation 1. Estimates are analogous to those in Table 2, with the following modifications. Column 1 adds a separate linear time trend for each village. Column 2 adds interactions between year fixed effects and each of the following continuous village-level variables measured at baseline: population, number of schools, log middle- and primary-school enrollment, literacy rate, population share of scheduled castes, irrigated land share, and distance to nearest town. Column 3 uses an unbalanced panel, adding additional villages that do not have data in all years. Column 4 adds a village time trend to the unbalanced panel specification. Column 5 restricts the sample to years 2005 or later. Column 6 includes data only for three years before each road is built and three years after. Different years are thus included for different villages, but each village has seven observations. Due to data availability, the Column 6 sample only includes villages with roads built between 2006 and 2012. All specifications have state-year fixed effects and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table 4

Impact of New Roads on Middle-School Enrollment Growth: Regression Discontinuity Estimates

Panel A: RD Estimates

	First Stage	Reduced Form	IV
	(1)	(2)	(3)
Above Population Threshold	0.239***	0.108	
	(0.015)	(0.066)	
New Road by 2011			0.450
			(0.276)
N	55271	55271	55271
r2	0.26	0.28	0.28

Panel B: Placebo RD Estimates

	First Stage	Reduced Form
	(1)	(2)
Above Population Threshold	0.014	0.009
	(0.011)	(0.059)
N	56219	56219
r2	0.27	0.25

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

Panel A shows regression discontinuity estimates of the impact of new road construction on log village middle-school enrollment, estimated with Equation 3. The sample includes all villages and enrollment data from 2011 to 2015. Standard errors are clustered at the village level to account for serial correlation. Column 1 reports first stage estimates of the effect of being above the state-specific population threshold (that defines road program eligibility) on the probability of receiving a new road before 2011. Column 2 shows a reduced form regression discontinuity estimate of the impact of being above the population eligibility threshold on log middle-school enrollment. Column 3 shows the instrumental variable estimate of the impact of a new road on village log middle-school enrollment. Panel B shows a placebo test consisting of the same specification in Columns 1 and 2 of Panel A, but in the set of states that did not adhere to PMGSY rules regarding the population eligibility threshold, and for whom there should thus be no treatment effect. All specifications control for baseline log middle-school enrollment, literacy rate, number of primary schools, number of middle schools, and the log number of non-farm jobs in the village.

Table 5

Impact of New Roads on
Middle-School Completion Examinations

	Exam Taken	Exam Passed	High Exam Score
	(1)	(2)	(3)
New Road	0.060***	0.058***	0.035***
	(0.019)	(0.019)	(0.014)
State-Year F.E.	Yes	Yes	Yes
Village F.E.	Yes	Yes	Yes
Panel Sample	Balanced	Balanced	Balanced
N	32239	32239	32239
r2	0.73	0.72	0.61

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

The table reports panel estimates of the effect of new road construction on village-level school examination performance, estimated with Equation 1. All columns use a balanced panel specification, analogous to Column 1 in Table 2. The dependent variable in Columns 1 through 3 is, respectively: (1) the log number of students sitting for the middle-school completion examination; (2) the log number of students who pass this exam; (3) the log number of students who pass this exam with distinction. All specifications have state-year fixed effects and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table 6

Impact of New Roads on Middle-School Enrollment:
Treatment Heterogeneity

	(1)	(2)	(3)	(4)	(5)
New Road	0.074***	0.115***	0.049**	0.061**	0.073**
	(0.017)	(0.024)	(0.023)	(0.024)	(0.035)
New Road * High Opportunity Cost Effect		-0.085**			-0.088**
		(0.034)			(0.034)
New Road * High Returns to Education Effect			0.053		0.061*
			(0.034)		(0.034)
New Road * High Income / Liquidity Effect				0.026	0.031
				(0.034)	(0.034)
State-Year F.E.	Yes	Yes	Yes	Yes	Yes
Village F.E.	Yes	Yes	Yes	Yes	Yes
Panel Sample	Balanced	Balanced	Balanced	Balanced	Balanced
N	111580	111580	111580	111580	111580
r2	0.81	0.81	0.81	0.81	0.81

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

The table reports panel estimates of the effect of new road construction on village log middle-school enrollment, interacted with binary district-level predictors of different treatment mechanisms. The size of the opportunity cost effect is proxied by the district-level mean unskilled urban wage minus the mean unskilled rural wage. The size of the returns to education effect is proxied by the difference between the urban and rural Mincerian returns to one additional year of education. The size of income and liquidity effects are proxied by the share of households in a village reporting zero assets in 2002. All these interactions take the value of one if the underlying variable is above the value of the median village. The specifications use Equation 1. All columns use a balanced panel specification, analogous to Column 1 in Table 2. Column 1 repeats the main specification without interactions in the sample with non-missing interaction variables. Columns 2 through 4 show the effects of the individual interaction terms, while Column 5 jointly estimates all interaction terms. Wage and education data comes from the 55th round of the NSS Employment and Unemployment Survey (1999-2000), and assets are from the Below Poverty Line survey (2002). All specifications have state-year fixed effects and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table 7

Treatment Heterogeneity in Estimated Road Impacts: Subgroup Estimates

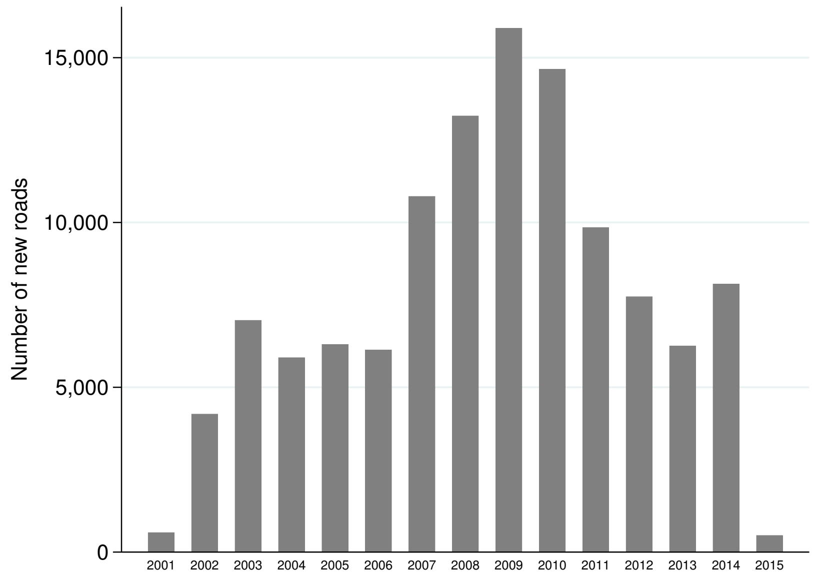
Opportunity Cost Effect	Returns to Education Effect	Income/Liquidity Effects	Treatment Estimate	Number of Villages
Low	Low	Low	0.032	2527
			(0.050)	
Low	Low	High	0.138***	1029
			(0.043)	
Low	High	Low	0.189***	987
			(0.050)	
Low	High	High	0.094*	523
			(0.051)	
High	Low	Low	-0.018	751
			(0.049)	
High	Low	High	0.014	844
			(0.045)	
High	High	Low	0.035	751
			(0.045)	
High	High	High	0.093*	558
			(0.054)	

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

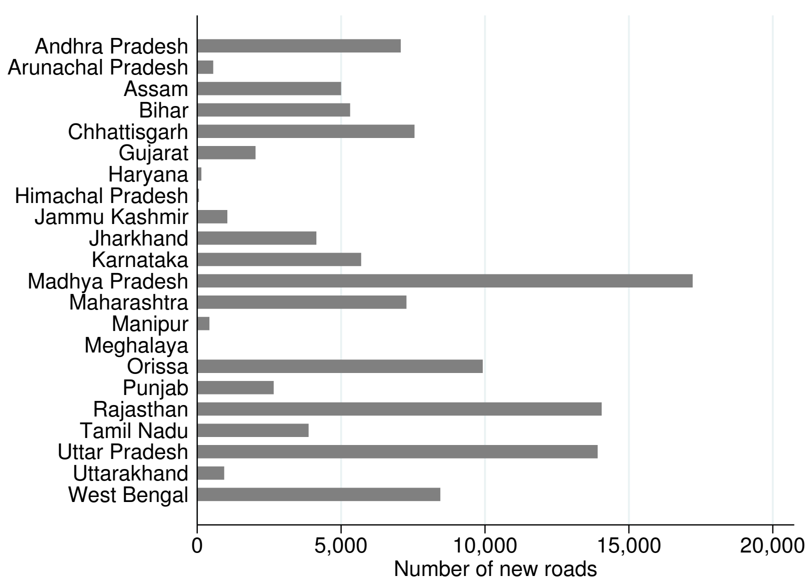
The table reports panel estimates of the effect of new road construction on village log middle-school enrollment, fully interacted with binary predictors of the size of the opportunity cost effect, the returns to education effect and the income/liquidity effect. The size of the opportunity cost effect is proxied by the district-level mean unskilled urban wage minus the mean unskilled rural wage. The size of the returns to education effect is proxied by the difference between the urban and rural Mincerian returns to one additional year of education. The size of income and liquidity effects are proxied by the share of households in a village reporting zero assets in 2002. All these interactions take the value of one if the underlying variable is above the value of the median village. The table shows linear combinations of interaction terms that describe the treatment effect in each of the eight partitions of the data according to the binary mechanism indicators. The specification is based on Equation 1, with added treatment interactions. Wage and education data comes from the 55th round of the NSS Employment and Unemployment Survey (1999-2000), and assets are from the Below Poverty Line survey (2002). All specifications have state-year fixed effects and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Figure 1
PMGSY New Road Summary Statistics

Panel A



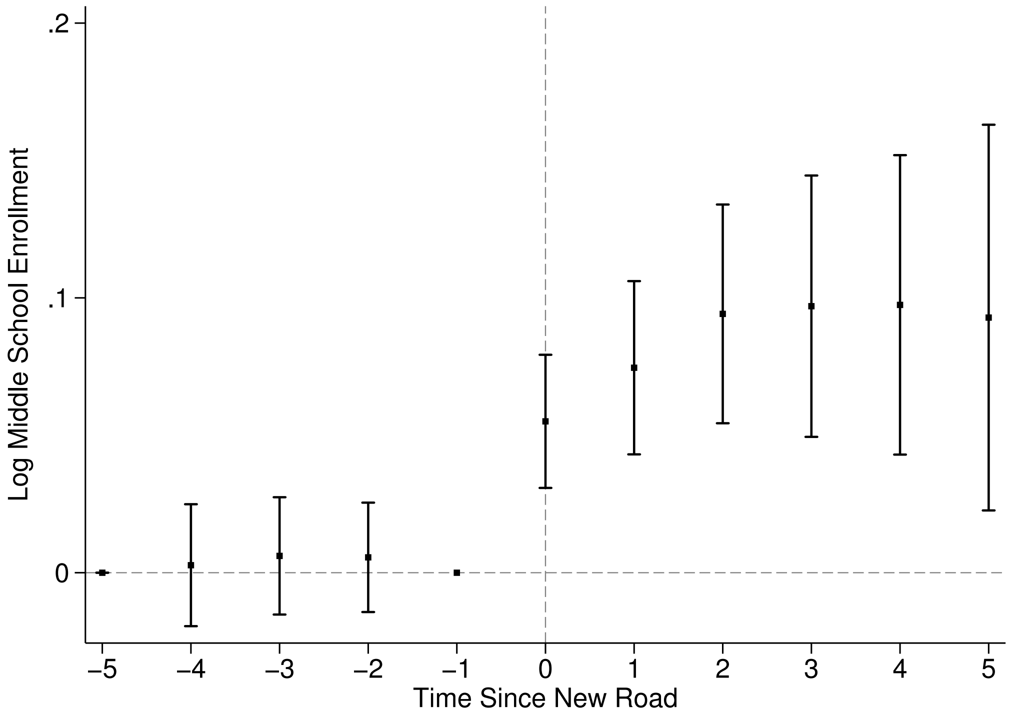
Panel B



The panels in this figure describe the distribution of new roads built under PMGSY between 2001 and 2015, across years and states. Graphs show new roads according to their registered completion dates. Data source: PMGSY Online Monitoring and Management System.

Figure 2

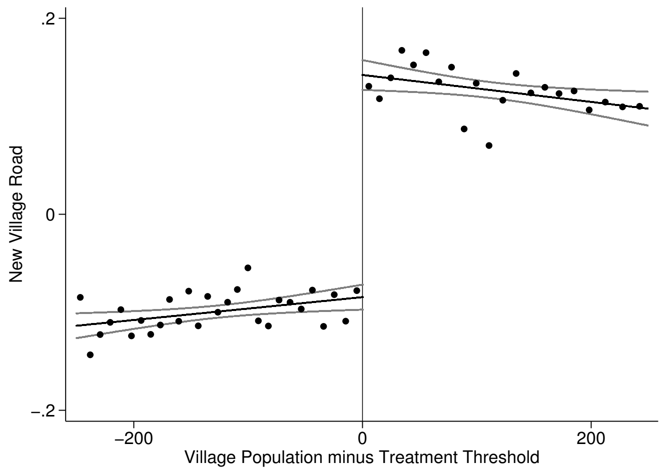
Impact of Roads on Middle-School Enrollment:
Treatment Effect Time Series



The figure shows coefficient estimates from a panel regression of log middle-school enrollment on a set of dummy variables indicating the number of years before or since a road was constructed, along with a set of state-by-year fixed effects and village fixed effects. The estimating equation is Equation 2. Year 0 is the first year in which a road was present when enrollment data were collected on September 30. Years $t = -1$ and $t = -5$ are omitted, following Borusyak and Jaravel (2017). 95% confidence intervals are displayed around each point estimate. Standard errors are clustered at the village level.

Figure 3

Regression Discontinuity First Stage:
Share of Villages Treated by Population

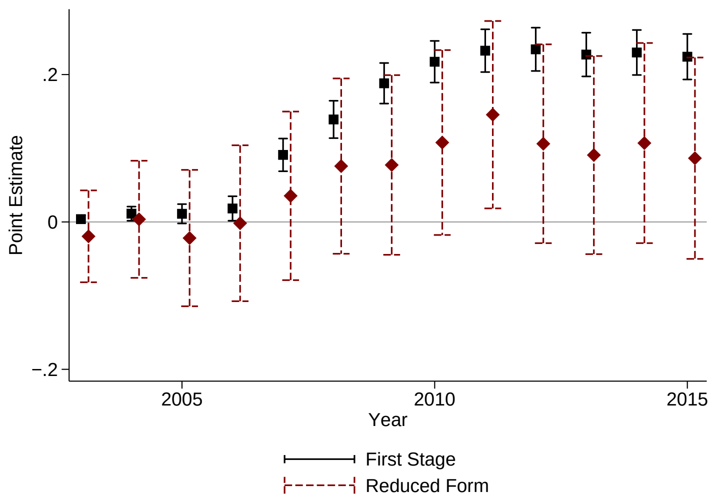


The graph plots the conditional expectation function of a dummy variable indicating that a village has received a road before 2011, conditioning on the village population as reported in the 2001 Population Census. Each point represents the mean of all villages in the given population bin (328 villages per bin). Population has been centered around the state-specific threshold used for road eligibility, which is either 500 or 1000, depending on the state. Points to the right of the center line represent villages with a higher prioritization under PMGSY, according to program rules.

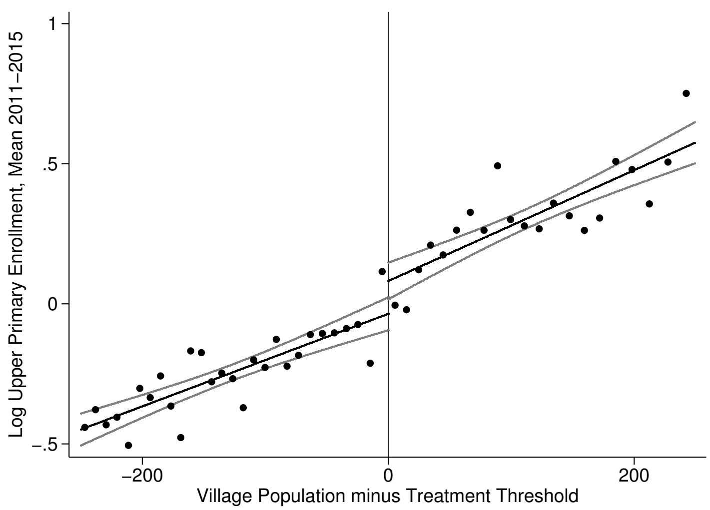
Figure 4

Regression Discontinuity Impacts of New Roads on Log Middle-School Enrollment Growth

Panel A: Reduced Form and First Stage Estimates By Year



Panel B: Log Middle School Enrollment by Population (2011-2015)



Panel A shows reduced form and first stage estimates from Equation 3, estimated on each sample year from 2003 to 2015. Each square and solid error bar describes a single estimate from Equation 3, where the dependent variable is an indicator taking the value one if a village received a new road before the year on the X axis. The diamonds and dashed error bars describe the reduced form RD estimate of the effect of being above the population threshold on village log middle school enrollment. Error bars show 95% confidence intervals. Panel B plots the conditional expectation function of average log middle school enrollment between 2011-2015. Population is centered around the state-specific threshold used for program eligibility, which is either 500 or 1000. Each point represents the mean of approximately 328 villages in the given population bin. Estimates in both panels control for baseline log middle-school enrollment, literacy rate, number of primary and middle schools, the log number of non-farm jobs in the village, and district fixed effects.

Appendix: Additional Figures and Tables for Online Publication Only

Table A1

Urban vs. Rural Wages and Mincerian Returns to Education

	Rural	Urban
Unskilled Wage	43.6	73.3
	(0.2)	(0.5)
Skilled Wage	114.3	166.0
	(0.9)	(0.8)
Return to Education	0.068	0.080
	(0.001)	(0.001)
Sample Size	46120	34024

The table shows mean district-level wages and returns to education from the 55th round of the NSS Employment and Unemployment Survey (1999-2000), separately for urban and rural areas. Wages are daily wages in Indian Rupees (in 1999, approximately 59 INR = 1 USD); the Mincerian returns to education is the coefficient on education from a regression of log wages on years of education, age, age squared, and log of household land. An individual is considered skilled if he or she has attained middle school or higher. Standard errors of means are shown in parentheses.

Table A2

Impact of New Roads on Middle-School Enrollment:
District-Year Fixed Effects

Dependent Variable	All, log	Girls, log	Boys, log	All, levels	Girls, levels	Boys, levels
	(1)	(2)	(3)	(4)	(5)	(6)
New Road	0.061***	0.053***	0.047***	1.957***	1.003***	0.954***
	(0.015)	(0.013)	(0.013)	(0.548)	(0.288)	(0.295)
N	146440	146440	146440	146440	146440	146440
r2	0.81	0.82	0.81	0.81	0.79	0.79

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

The table reports panel estimates of the effect of new road construction on village-level log middle-school enrollment, estimated with Equation 1. Specifications are identical to Table 2, but with district-by-year fixed effects instead of state-by-year fixed effects. Column 1 presents the primary balanced panel specification. The dependent variable in Columns 2 and 3 is log middle-school enrollment for boys and girls respectively. Column 4 estimates the same regression with the level of middle-school enrollment as the dependent variable, and Columns 5 and 6 do the same for boys and girls respectively. All specifications include district-year fixed effects and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table A3

Impact of New Roads on Primary-School Enrollment

	Panel			Reduced Form	IV
	(1)	(2)	(3)	(4)	(5)
New Road	-0.005	-0.004	-0.005		0.033
	(0.004)	(0.003)	(0.005)		(0.086)
Above Population Threshold				0.008	
				(0.020)	
N	146678	146678	237281	66663	66663
r2	0.87	0.92	0.88	0.30	0.30

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

The table reports estimates of the effect of new road construction on village log *primary*-school enrollment. Columns 1 through 3 present panel estimates, and Columns 4 and 5 present RD estimates. Column 1 presents the main balanced panel specification. Column 2 adds village-specific time trends, and Column 3 repeats the main specification in the unbalanced panel. Column 4 shows the reduced form estimate of the effect on log primary-school enrollment growth of being just above the eligiblity threshold, and Column 5 presents the RD IV estimates of the impact of the new road. Standard errors are clustered at the village level.

Table A4

Regression Discontinuity Baseline Tests

Variable	RD Estimate
Number of schools (DISE)	0.003
	(0.021)
Enrollment Divided by Population	-0.000
	(0.006)
Log Total Enrollment (grades 1-8)	-0.011
	(0.018)
Log Primary Enrollment (grades 1-5)	-0.018
	(0.019)
Log Middle Enrollment (grades 6-8)	0.012
	(0.053)
Log Students Passing Exam	-0.060
	(0.058)
Log Students with Distinction on Exam	-0.020
	(0.027)
Literacy Rate	0.000
	(0.005)
Scheduled Caste Population Share	0.007
	(0.006)
Distance to Nearest Town (km)	0.050
	(0.583)
Share of Asset-Poor Households	-0.001
	(0.006)
Number of Observations	17639

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

The table reports regression discontinuity estimates of the change in baseline variables across the PMGSY eligibility threshold, using Equation 3. Literacy, scheduled caste share and town distance are measured in 2001, enrollment, school variables and asset share are measured in 2002, and exam scores in 2005. All specifications include district fixed effects and control linearly for population (the running variable) on each side of the treatment threshold. Standard errors are in parentheses.

Table A5

Impact of New Roads on
Primary-School Completion Examinations

	Exam Taken	Exam Passed	High Exam Score
	(1)	(2)	(3)
New Road	0.028*	0.021	0.024
	(0.016)	(0.016)	(0.017)
State-Year F.E.	Yes	Yes	Yes
Village F.E.	Yes	Yes	Yes
Panel Sample	Balanced	Balanced	Balanced
N	31671	31671	31671
r2	0.73	0.71	0.61

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

The table reports panel estimates of the effect of new road construction on village-level *primary* school examination performance, estimated with Equation 1. All columns use a balanced panel specification, analogous to Column 1 in Table 2. The dependent variable in Columns 1 through 3 is, respectively: (1) the log of the number of students sitting for the primary-school completion examination; (2) the log number of students who pass this exam; (3) the log of the number of students who pass this exam with distinction. All estimations have state-year fixed effects and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table A6

Treatment Heterogeneity in Road Impacts:
Quartile Results

Panel A: Opportunity Cost Effect Quartiles

	(1)	(2)	(3)	(4)
New Road	0.096**	0.136***	0.023	0.027
	(0.047)	(0.030)	(0.033)	(0.031)
N	19544	33614	31584	28322
r2	0.78	0.82	0.81	0.83

Panel B: Returns to Education Effect Quartiles

	(1)	(2)	(3)	(4)
New Road	0.033	0.049	0.144***	0.068**
	(0.033)	(0.033)	(0.039)	(0.033)
N	29134	30016	23128	29204
r2	0.82	0.79	0.81	0.83

Panel C: Income/Liquidity Effect Quartiles

	(1)	(2)	(3)	(4)
New Road	0.086**	0.033	0.128***	0.060*
	(0.039)	(0.033)	(0.032)	(0.033)
N	22372	29946	30170	28924
r2	0.81	0.82	0.81	0.80

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

The table reports panel estimates of the effect of new road construction on village log middle-school enrollment. The estimates are calculated for separate samples defined by quartiles of the mechanism proxies for the opportunity cost effect (Panel A), the returns to education effect (Panel B), and the income/liquidity effects (Panel C). The size of the opportunity cost effect is proxied by the district-level mean unskilled urban wage minus the mean unskilled rural wage. The size of the returns to education effect is proxied by the difference between the urban and rural Mincerian returns to one additional year of education. The size of income and liquidity effects are proxied by the share of households in a village reporting zero assets in 2002. The estimating equation is Equation 1. All specifications include state-year fixed effects and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table A7

Panel and Regression Discontinuity Estimates of
Impact of Roads on School Infrastructure

Dependent Variable	Balanced Panel	Unbalanced Panel	RD
Piped Water	0.001	0.002	0.005
	(0.004)	(0.003)	(0.007)
Toilet	0.003	0.016***	0.000
	(0.005)	(0.004)	(0.008)
Electricity	0.003	0.004**	-0.002
	(0.002)	(0.002)	(0.006)
Library	0.000	0.006	0.004
	(0.005)	(0.004)	(0.009)
Computer	-0.004**	-0.002	0.001
	(0.002)	(0.002)	(0.004)
Perimeter Wall	0.001	0.002	0.005
	(0.004)	(0.003)	(0.009)
Playground	0.009**	0.007*	0.011
	(0.004)	(0.004)	(0.009)
Log Number of Schools	0.000	0.001	0.006
	(0.000)	(0.002)	(0.005)

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

The table reports panel estimates of the effect of new road construction on village-level school infrastructure, estimated with Equation 1 (Columns 1-2) and Equation 3 (Column 3). Each entry in the table shows a treatment effect analogous to the "New Road" row in Table 2, and thus each entry represents a distinct regression. The left column shows the dependent variable for each regression, and the column header describes the sample. Column 1 presents the main balanced panel specification. Column 2 presents results from the unbalanced panel. Columns 1 and 2 include state-year fixed effects and village fixed effects, and standard errors are clustered at the village level. Column 3 presents reduced form regression discontinuity estimates of the impact on the infrastructure variable of being in a village just above the treatment threshold.

Table A8

Regression Discontinuity Placebo Estimates:
Other Public Goods

Dep. Var.	Prim. School	Mid. School	Sec. School	Electricity	Health Center	Bank
	(1)	(2)	(3)	(4)	(5)	(6)
Above Population Threshold	-0.008	0.012	-0.001	0.016	0.002	0.002
	(0.005)	(0.013)	(0.006)	(0.013)	(0.002)	(0.002)
N	16973	16973	16973	16973	16973	16973
r2	0.37	0.32	0.15	0.36	0.09	0.08

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

The table shows reduced form regression discontinuity estimates of the change in public goods *other than roads* across the PMGSY population treatment threshold, using Equation 3. The dependent variable, column by column, is (i) presence of primary school; (ii) presence of middle school; (iii) presence of secondary school; (iv) village access to electric power; (v) presence of a primary health center; and (vi) presence of a commercial bank. All specifications include district fixed effects and control for baseline log middle-school enrollment, literacy rate, number of primary schools, number of middle schools, and the log number of non-farm jobs in the village.

Table A9

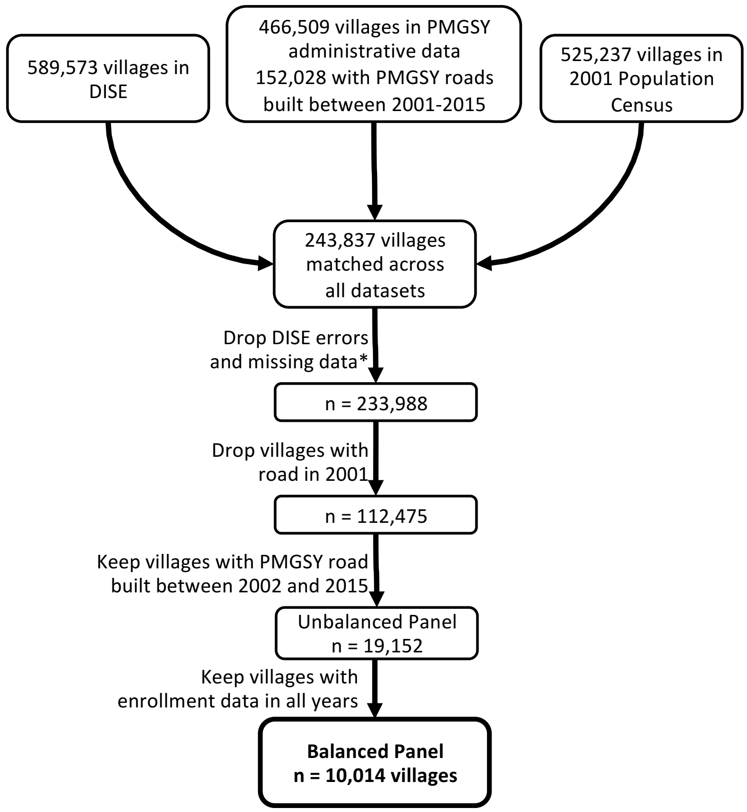
Impact of Roads on Middle-School Enrollment:
Spatial Effects

	Spillovers		Village Area		Nearby Eligible Kids	
	3 km	5 km	Low	High	Low	High
New Road	-0.011	0.002	0.083***	0.089***	0.075***	0.062**
	(0.016)	(0.010)	(0.018)	(0.018)	(0.026)	(0.026)
State-Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Village F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Panel Sample	Balanced	Balanced	Balanced	Balanced	Balanced	Balanced
N	93730	93730	126270	108624	46872	46858
r2	0.86	0.84	0.76	0.77	0.79	0.80

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

This table shows panel estimates of the impact of road construction on log middle-school enrollment. Columns 1 and 2 show the impact of a new road on middle-school enrollment in *nearby* villages, respectively those within a 3 km and 5 km radius. Columns 3 and 4 divide the sample into villages with above-median land area per person and below-median land area per person, and report effects separately. Columns 5 and 6 divide the sample into villages according to their proximity to children in villages *without* middle schools. Column 5 shows the effect of new roads on middle-school enrollment in villages with few nearby children in villages without middle schools; Column 6 shows estimates in villages where there are many nearby underserved schoolchildren. All specifications include state-year fixed effects and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

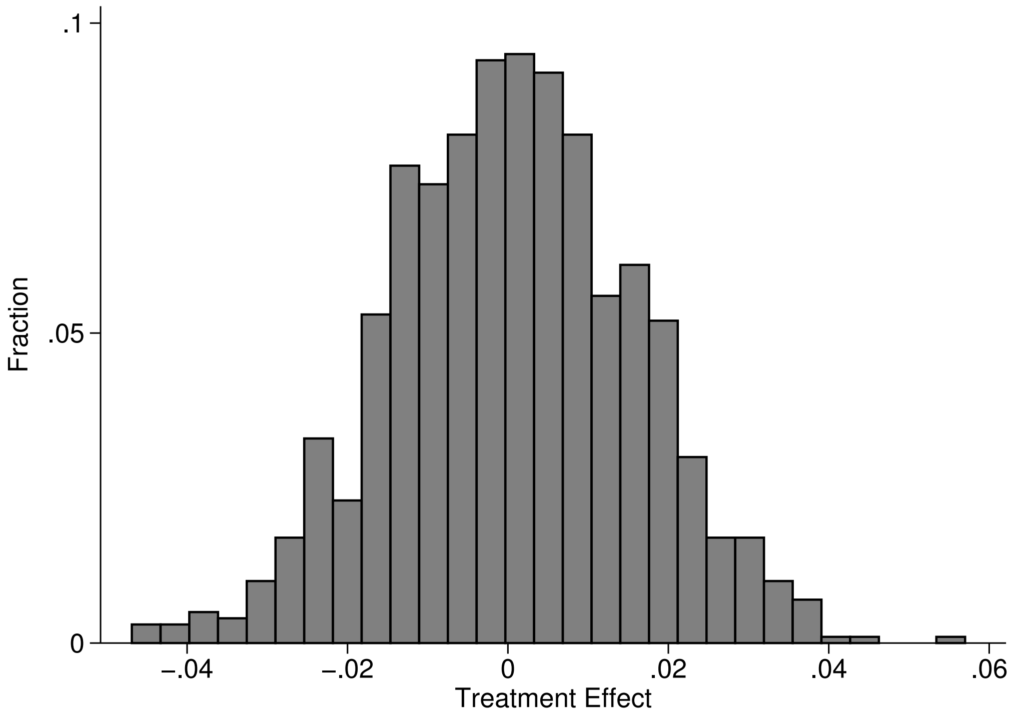
Figure A1
Sample Construction



The figure shows how we arrived at our final number of observations from the original datasets. DISE = District Information System for Education. PMGSY = Prime Minister's Road Building Program. All observation counts indicate number of villages at each stage.
*Observations were dropped if DISE reported grade one to eight enrollment greater than 60% of village population (99th percentile).

Figure A2

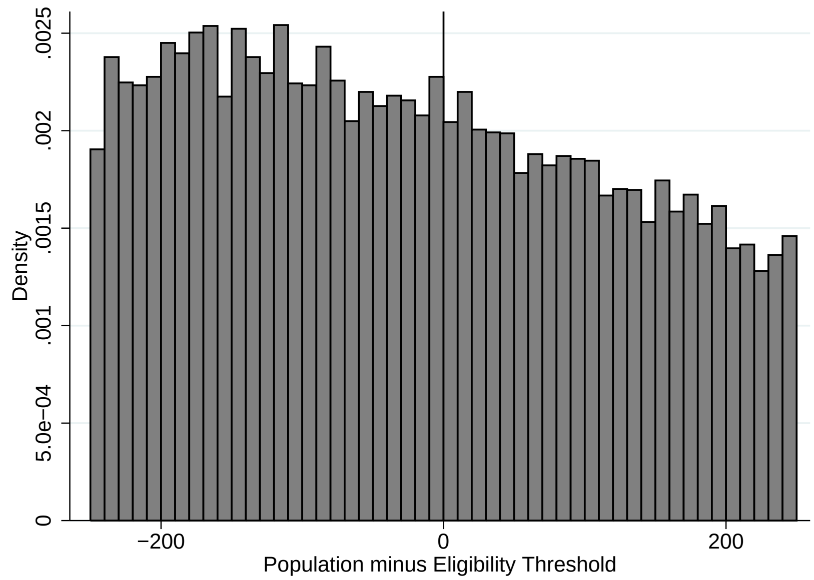
Panel Estimates of Effect of Roads on
Middle-School Enrollment: Permutation Test

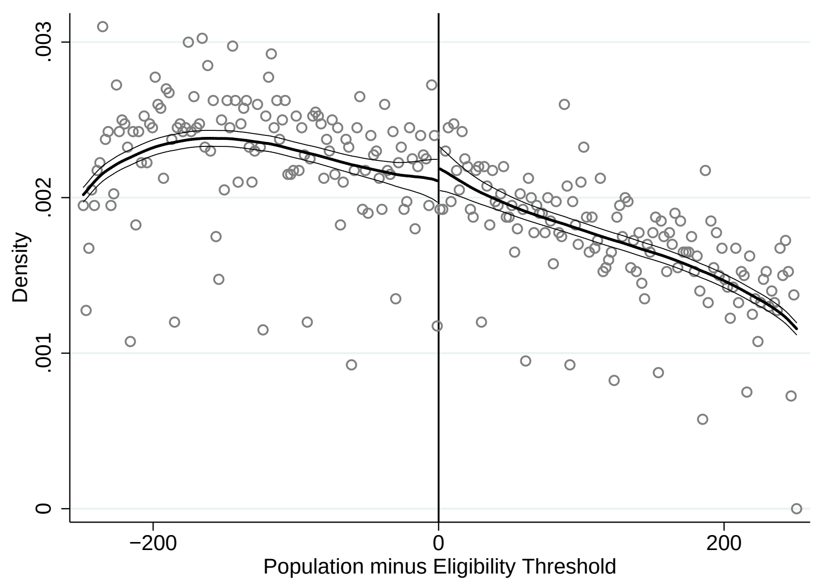


The figure shows the distribution of estimates from a placebo permutation test of the main panel specification presented in Column 1 of Table 2. For each village in the main sample, we randomly generated a placebo year of road completion, and then estimated Equation 1. We ran this estimation 1000 times; the graph shows the distribution of estimates of β, which would be the impact of a new road on log middle-school enrollment.

Figure A3

Regression Discontinuity: Continuity of Running Variable

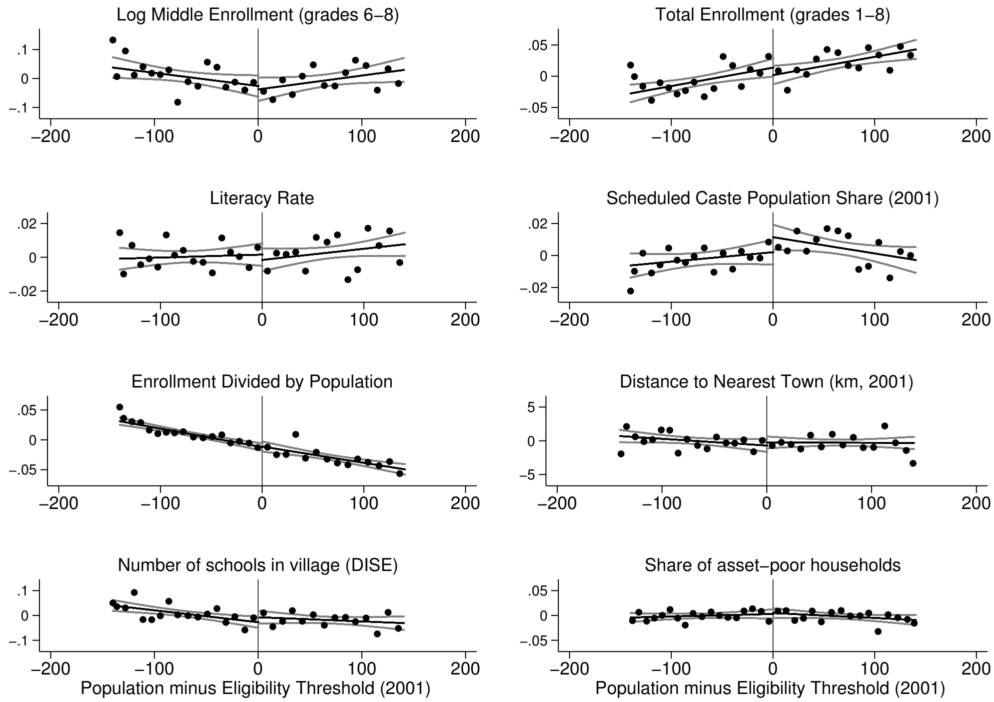




The figures show the distribution of village population in the set of villages in our sample. The top panel shows a histogram of village population, centered around the treatment threshold. In the bottom panel, we plot a non-parametric regression to each half of the distribution following McCrary (2008), testing for a discontinuity at the treatment threshold.

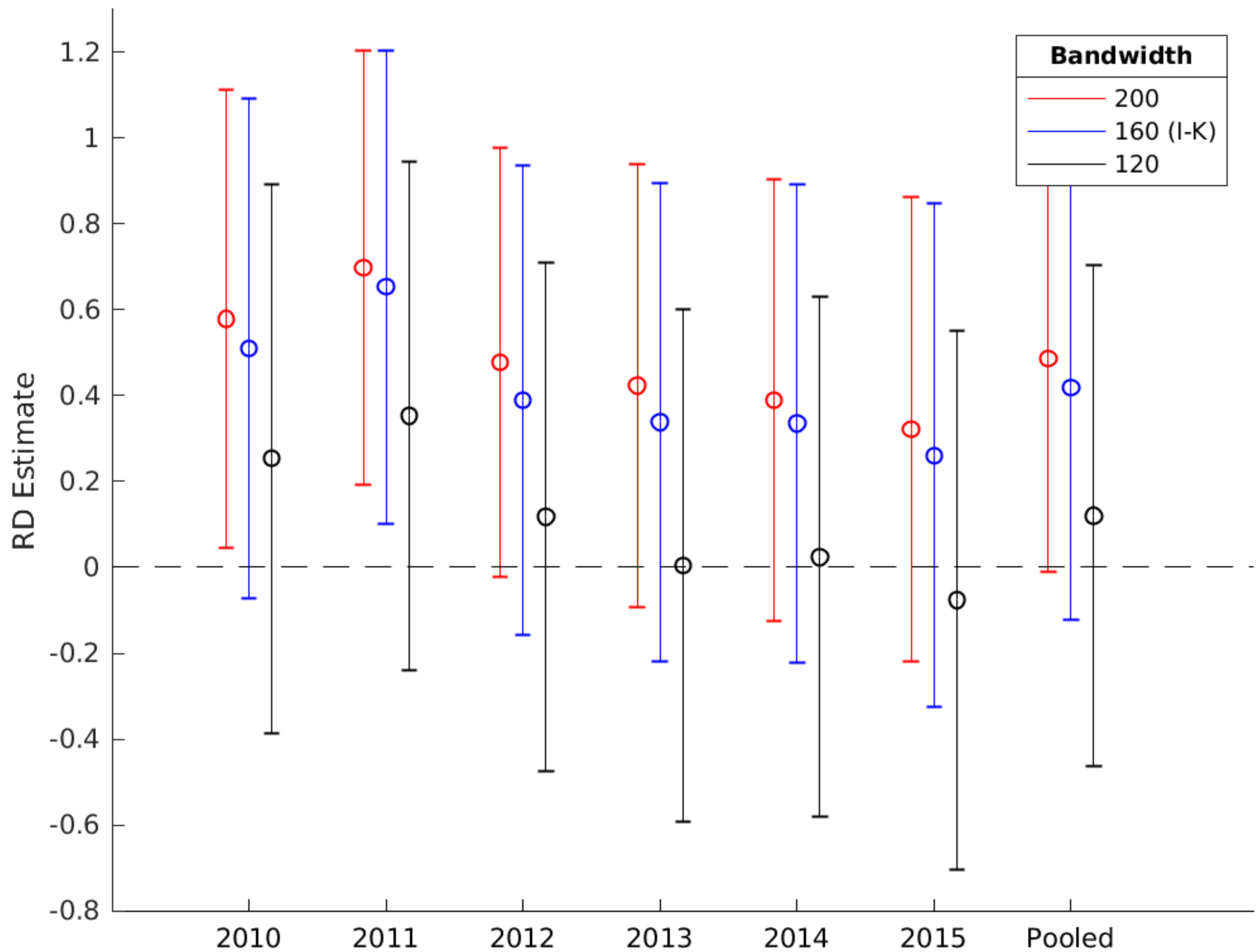
Figure A4

Regression Discontinuity: Continuity of Baseline Variables



The graphs show the distribution of baseline variables against the regression discontinuity running variable, population. We have subtracted the treatment eligibility threshold from the population variable so that eligibility for the road program rises discontinuously at zero. Each point in the graphs represents the mean baseline value of the variable in the set of villages within a given population bin. We fit a linear function to the data on each side of the treatment threshold, and show 95% confidence intervals.

Figure A5
Impacts of New Roads on Middle-School Enrollment:
RD Estimates by Year and Bandwidth

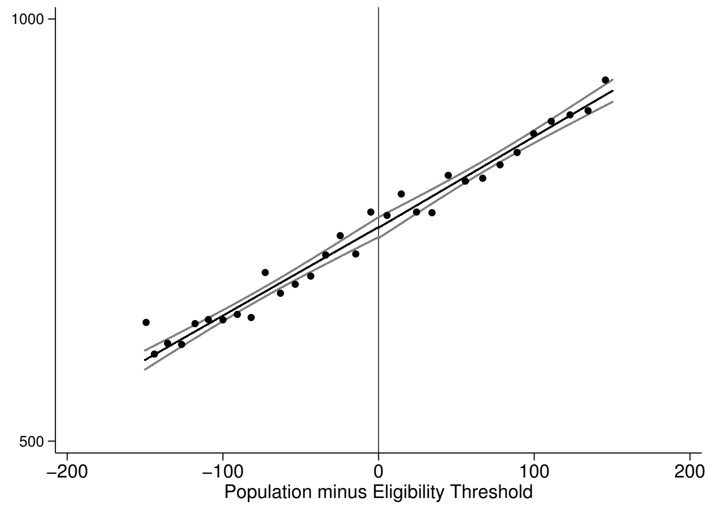


The figure shows IV estimates from Equation 3, estimated on different sample years, and at bandwidths 25% higher and lower than the optimal bandwidth of 120 selected with the algorithm of Imbens and Kalyanaraman (2012). Each point represents a single RD estimate of the impact of new roads on log middle school enrollment. Error bars show 95% confidence intervals. The pooled estimate corresponds to that from Table 4, and pools years 2011-2015, clustering standard errors at the village level. All specifications control for baseline log middle-school enrollment, literacy rate, number of primary schools, number of middle schools, and the log number of non-farm jobs in the village.

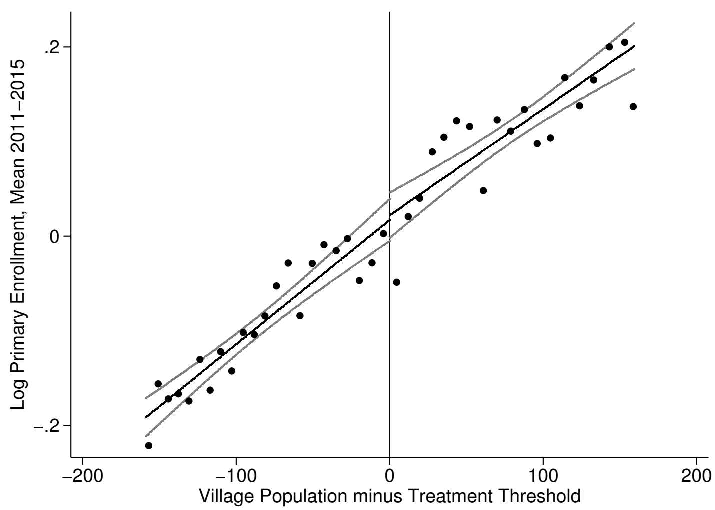
Figure A6

Regression Discontinuity Reduced Form:
Population and Primary School

Panel A: Log Population (2011)



Panel B: Log Primary School Enrollment (2011-2015)



The figure shows the conditional expectation function of the mean of annualized village-level population in 2011 (Panel A) and the mean of log primary school enrollment in 2011-2015 (Panel B), conditioning on the village population in 2001. 2001 Population (the X axis) is normalized to be centered around the state-specific threshold used for program eligibility.