WORKING PAPER · NO. 2017-08

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

Anjali Adukia, Sam Asher, Paul Novosad

MARCH 2019

Becker Friedman Institute
Development Economics Initiative

1126 E. 59th St, Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

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

Anjali Adukia†

Sam Asher‡

Paul Novosad§

March 2019

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 effects 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. Heterogeneity in treatment effects supports a standard human capital investment model: enrollment increases most when nearby labor markets offer high returns to education and least when they imply high opportunity costs of schooling.

JEL Codes: I25; O18; J24.

*We thank Martha Bailey, Chris Blattman, Liz Cascio, Dave Donaldson, Esther Duflo, 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, EBRD, EEA, the Federal Reserve Banks of Chicago (CHERP) and New York, Georgetown University, Michigan State University, NEUDC, the NBER Education Program Meetings, Northwestern University School of Law, the OECD, PAA, PacDev, Stanford, University of California-Berkeley, University of Chicago, University of Connecticut, University of Illinois at Chicago, University of Michigan, University of Missouri, the Urban Institute, and Yale, and the anonymous referees for their helpful comments and suggestions. We thank Srinivas Balasubramanian, Jack Landry, 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, 1307 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 connections to international markets. A key trade-off for individuals 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. As educational investment responds to market integration, it shapes the long-run economic impacts of policies that are increasingly integrating markets in developing countries.

We examine the human capital investment response when a paved road is built to a previously unconnected village, effectively connecting it to outside markets. India's national rural road construction program (PMGSY) built high quality roads to 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. The impacts of new road connections on schooling are theoretically ambiguous: they may raise the returns to education, raise the opportunity cost of schooling, and/or have important income or liquidity effects.

A major challenge in estimating causal effects of new roads is the endogeneity of road placement. If roads are targeted to wealthier or poorer regions, for example, then comparisons of villages with and without roads will be biased. To overcome this bias, we exploit the timing of road completion in each village, estimating a panel regression with village and

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

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

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. We thereby compare educational outcomes in villages before and after a road is built, flexibly controlling for time-variant regional shocks and static differences between villages that receive roads in different years.

We use village-level school enrollment data from India's national annual census of primary and middle schools (District Information System for Education, DISE, 2002-2015). Through a combination of human and machine fuzzy matching, we linked DISE data to administrative data from the national rural road construction program. The result is a panel of 300,000 villages across 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 heterogeneous impacts in subsamples of the data. Our sample spans a broad range of economic conditions in India today, similar to the variation across many places worldwide that remain unreached by paved roads.

We find that road construction significantly increases middle school enrollment. 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.

We do not find enrollment effects for primary school children, for whom there are fewer labor market opportunities. We do find small increases in primary school performance, however, suggesting that students may be increasing school effort on the intensive margin.

[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.

We then explore heterogeneity in the treatment effects on middle school children, guided by predictions from a standard model of human capital investment. The model predicts four primary mechanisms through which educational investment in rural areas may be affected by road connections to nearby labor markets. We model roads as leading to factor price equalization across areas, which is then predicted to: (i) raise the low-skill 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] The model suggests that the importance of each of these effects will be different across regions, depending on local market characteristics. Newly connected villages will experience larger opportunity cost effects when the urban-rural low skill wage gap is large. Returns to education effects will be larger when the urban-rural gap in Mincerian returns to education is larger. To predict the expected importance of income and liquidity effects, we use a measure of asset poverty.[5]

The estimated variation in treatment effects across these three measures supports the predictions of the model. Partitioning our data according to these measures, we estimate substantial treatment effect heterogeneity across villages, with effects that are positive and statistically significant in 39% of villages and positive but insignificant in 52% of villages. Market integration has (small and statistically insignificant) negative effects only in the 9% of villages where 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.

We explore several other treatment mechanisms, for which we do not find support in the data: (i) supply-side improvements in school infrastructure; (ii) migration; (iii) displacement to/from nearby villages; and (iv) improved access for children on the outskirts of villages.

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

[4]Because roads may change factor prices in many markets, many other effects are also possible. We focus on effects predicted from the literature on road construction.

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

to drive further long-run growth through increased educational attainment. Enrollment and exam performance respond positively to increased economic opportunities. Our results also provide a causal mechanism that underlies the strong correlation around the world between education, growth, and trade.

This study is related to the literature on the impact of labor demand shocks on schooling decisions, which finds both positive and negative schooling impacts from new economic opportunities.[6] The estimated heterogeneity in treatment effects in our study is consistent with the heterogeneity found in the literature, and well explained by the standard human capital model: new labor market opportunities affect the opportunity costs of schooling, but also affect the long-run benefits of schooling and demand for schooling through income and liquidity effects.

Our paper also contributes to the literature on the development impacts of transport infrastructure.[7] Relative to earlier work on roads and schooling, our large village-level sample and research design allow a more precise estimation of the causal effects of road construction. Our results suggest that road construction and domestic market integration may have greater long-run impacts on economic development by increasing educational investment. Finally, we contribute to a wide body of research on improving educational attainment in

[6]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 (Cascio and Narayan, 2015; Atkin, 2016; Shah and Steinberg, 2017). Our estimates are also related to impacts on human capital accumulation from India's national public works program (MGNREGA), which has found small decreases in enrollment for middle school students across India (Das and Singh, 2013; Islam and Sivasankaran, 2015; Li and Sekhri, n.d.; Shah and Steinberg, 2015; Adukia, 2018).

[7]Some examples include Jacoby (2000); Jacoby and Minten (2009); Gibson and Olivia (2010); Mu and van de Walle (2011); Casaburi et al. (2013); Donaldson and Hornbeck (2016); Donaldson (2018). For a detailed review, including studies on the impacts of highways and regional roads, see Hine et al. (2016). Asher and Novosad (2018) find that PMGSY road construction leads to changes in occupations but has little effect on village assets, incomes, or consumption using regression discontinuity exploiting village population thresholds. Mukherjee (2012) uses a similar approach to find that PMGSY road construction in India increases school enrollment. We present comparable regression discontinuity estimates, but we focus on panel estimates that have much greater statistical precision and allow for analysis of treatment heterogeneity. Using district-level data from India, Aggarwal (2018) 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.

developing countries (see Evans and Popova (2016) and Glewwe and Muralidharan (2016) for reviews of this literature). Our results highlight that investments outside the education sector can have important 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 results, Section VII explores the mechanisms suggested by the model of human capital investment, 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 both by characteristics of villages and by characteristics of local labor market conditions outside the village.

In this framework, the key decision point is an individual's trade-off between the long-run benefits of schooling and the short-run return to labor. A two-period model is sufficient to highlight the essential comparative statics. In the first period, a person chooses between working for a low-skill wage and obtaining schooling. In the second period, the person works and receives either a high or a low wage, depending upon the schooling choice in the first period. The person consumes in both periods, drawing from an initial endowment and wages earned in each period that the person works. The person can save at some interest rate, but may be restricted in borrowing. The person'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.[8]

[8]This framework underlies much of the theoretical literature on child labor and human capital invest-

When a village becomes connected to an external market via a new road, there is a change in the parameters underlying this trade-off between education and early participation in the labor market. Reduced transportation costs affect worker wages in both periods by inducing factor price equalization across areas.[9] In equilibrium, urban areas have higher wages than rural areas for both low- and high-skilled workers, and higher Mincerian returns to education (see Appendix Table A1).[10] Connecting a village to its external market is therefore likely to: (i) increase the low-skill wage; and (ii) increase the returns to education.[11]

An increase in the low-skill wage raises the opportunity cost of schooling and discourages human capital investment, which we call the opportunity cost effect. A relative increase in the high-skill wage raises the returns to education and encourages human capital investment, which we call the returns to education effect. Changes in wages could also affect human capital investment through income effects or liquidity effects. As wages rise, income effects will increase human capital investment if schooling is a normal good. Increases in household liquidity may also affect human capital investment if credit constrained households cannot afford to pay school fees or require children to work. In principle, these effects could go in either direction, but based on urban-rural wage gaps and skill gaps in India, we expect the opportunity cost effect to reduce schooling, and the returns to education, income, and liquidity effects to increase schooling.

Predictions for how human capital investment is affected by factor price equalization in goods and capital markets are less clear, as many prices can change simultaneously.[12] Rural

---

ment decisions. See, for example, Ranjan (1999) or Baland and Robinson (2000). We abstract away from intra-household bargaining.

[9]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. Asher and Novosad (2018) show that new PMGSY roads increase the number of people working for wages outside of villages.

[10]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, unobserved education quality differences are unlikely to 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).

[11]We can think of these effects as changes in real wages, such that changes in local goods prices due to new roads are subsumed in the above effects.

[12]For example, in general equilibrium, there may be various changes in the prices of different intermediate

road construction could also affect schooling decisions through many other channels, such as information flows, marriage markets, and healthcare access, but we focus on impacts through labor markets and wages.

To explore which of these mechanisms are more important, we identify places where these mechanisms are likely to generate heterogeneous impacts on schooling decisions. We generate measures of regional labor market conditions, which should influence the magnitude of changes in low-skill wages and the returns to education when a village becomes more integrated with nearby labor markets. The opportunity cost effect should be particularly large when the low-skill regional wage is much larger than the low-skill wage in the village becoming connected. The effect on returns to education should be larger when regional returns to education are greater than returns to education in the village becoming connected. We expect that income and liquidity effects on schooling would be greater in villages that are liquidity constrained or have low incomes, though the economic opportunities created by new roads may also differ in these villages. In the absence of shocks that separately affect liquidity and income, these liquidity and income effects are difficult to disentangle (Edmonds, 2006), and so we consider them together.

III Background and Details of the Road Construction Program

School enrollment increased substantially in India over our study period, from 2002 to 2015, paralleling a global increase in educational attainment. Increasing educational attainment has been a national priority in India, with several national initiatives aimed toward achieving universal primary education. Educational attainment and rates of economic development vary substantially across India. Indian policy-makers in the past have 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 and Somanathan, 2007; Dreze and Sen, 2013).

Many rural villages have limited connections to regional markets even while major cities in

goods and final goods. Capital market integration could also affect interest rates, changing the impact of liquidity constraints and changing the return on savings.

India have become increasingly connected to world markets. High costs of road construction and rapid degradation have historically constrained the ability of the Indian 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 Village Road Program, or PMGSY), a national program that aimed to eventually build a paved road to every village in India. While the federal government issued implementation guidelines, decisions on village-level allocations of roads were ultimately made at the district level. The unit of targeting for road construction was the habitation, which is the smallest rural administrative unit in India. A village is typically comprised of between one and three habitations; there are approximately 600,000 villages in India and 1.5 million habitations. 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) little economic data is available at the habitation level.

Road construction was targeted initially toward villages with larger populations. In some states, this took the form of a strict population threshold for road construction eligibility, while other criteria were used in other states. Given the program rules, early-treated villages tended to have larger populations, but were not substantially different from late-treated villages in other characteristics.[13] There were initially 80,000 villages eligible for the road construction program, a number that has grown as guidelines have been expanded to include smaller villages.

By 2015, over 115,000 villages had paved roads built or upgraded under the PMGSY program. These construction projects were most often managed through subcontracts with larger firms, and were built with capital-intensive methods and external labor; the building

[13]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.

of the road itself was therefore not a major local labor demand shock. These PMGSY roads are distinct from new roads being built under the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), which are lower quality roads built with labor-intensive methods.[14]

Figure 1 shows the distribution of road construction by state and over time. The median road length was 4.4 kilometers. Given the difficult terrain around many of these villages, a new paved road represents multiple hours saved on a round trip to or from the village.

FIGURE 1 ABOUT HERE.

IV Data

We constructed a village-level panel data set, combining data on road construction with education outcomes and other village characteristics. We matched an annual census of Indian schools, the District Information System for Education (DISE, 2002-2015), to administrative data from the implementation of the road program (2001-2015) and three successive Indian Population Censuses (1991, 2001, 2011). We matched locations based on village, block, and district names using a set of fuzzy matching algorithms.[15]

DISE is an annual census of primary and middle schools in India. It includes data on student enrollment, exam completion, and school infrastructure. This data set 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 cover every registered Indian government primary and middle school beginning in 2005.[16] We also have DISE data for a smaller sample of schools from 2002-2004, a period when the

[14]Major highway projects during this period, such as the Golden Quadrilateral, were planned and executed independently of PMGSY. There is not evidence of coordination of PMGSY road construction with the construction of the Golden Quadrilateral or other district road improvement projects.

[15]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 is posted at github.com/paulnov/masala-merge. 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. We matched 80 percent of census blocks; within census blocks, we matched 81 percent of villages.

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

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.[17] DISE data are based on interviews with school headmasters, and there are potential concerns of misreporting and inflated enrollment. Because new roads lower the cost of monitoring enrollment numbers, we expect that changes in misreporting would bias downward the estimated impacts of roads. It is also unlikely to affect only middle school students.

Our main 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. Our main focus is on outcomes for middle school children (grades 6-8), because the transition to middle school is a natural breakpoint in a child's schooling at which educational milestones are often measured, but we report some outcomes for primary school children as well. Younger children also have fewer labor market opportunities. DISE did not report enrollment information for high school over our sample period. DISE does not report the total number of school-age children in a village, so we are unable to calculate enrollment rates directly. 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 end-of-school examinations for primary schools and middle schools. These exams 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 high marks. We also use DISE data on school infrastructure, which describe the school-level presence of piped water, sanitation facilities, electricity, a library, a computer, a boundary wall, and a playground.

For data on road construction, we use administrative records of the PMGSY program that

[17] Our preferred sample drops villages that reported total enrollment (first through eighth grades) greater than 60 percent of total population (the 99th percentile of this statistic), which appear to be measured with error. By comparison, in 2001 only 22.4 percent of the population was of primary school age 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.

are used to track and implement the program.[18] Road data are reported at the village level, or at the smaller habitation level that we aggregate 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.

Appendix Figure A1 shows how we define our main sample of villages. We restrict our sample to villages that did not have a paved road in 2001. We further limit the sample to villages that received new PMGSY roads between September 2003 and September 2015, and we exclude villages where roads were categorized as upgrades rather than as new roads. Our main sample includes 10,014 villages that received roads between 2003 and 2015, for which we have enrollment data for at least one pre-treatment year and one post-treatment year. We find similar results when we extend our sample to an unbalanced sample (n=19,152) or include villages that never received PMGSY roads (n=112,475).

For data on district-level rural wages and urban wages, we use data from all individuals reporting wages from the 55th round of the NSS Employment and Unemployment Survey (1999-2000). For data on village population and other characteristics, we use data from the Population Censuses of India in 1991, 2001, and 2011, and the 1998 Economic Census.

Table 1 shows summary statistics of villages at baseline. The enrollment drop-off 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.

TABLE 1 ABOUT HERE.

V Empirical Strategy

Our goal is to estimate the causal impact of roads on educational choices. Cross-sectional estimates of the relationship between village roads and schooling decisions are likely to be

[18] We obtained these data from the government's public reporting portal for PMGSY, hosted at http://omms.nic.in.

biased estimates of the impact of roads on schooling, because villages that do not have access to paved roads are different from more accessible villages along many dimensions. For example, villages without paved roads are likely to be smaller, have more difficult terrain, and be more politically marginalized. Our main empirical specification is a panel fixed effects regression that exploits the timing of road construction, within the set of all villages that received new roads by 2015 under the PMGSY program.

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 variable for 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.

The coefficient of interest, β, measures the impact of a new road on village-level outcomes (such as log school 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.

The identification assumption is that in the absence of the PMGSY, village-level outcomes would have followed the same path over time in villages that receive a paved road in different years, after partialling out the location and time fixed effects. The state-year fixed effects control flexibly for differential enrollment growth across states. This alleviates 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.

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 reports that a new paved road leads to a seven percent increase in middle school enrollment in a village (95% confidence interval: 4.1 – 9.9 percent). This effect corresponds to approximately three additional students in middle school, given the sample mean of 39 students enrolled in middle school.[19] In columns 2 and 3 of Table 2, we split the analysis by gender, and find similar effects on the enrollment of girls and boys. Columns 4 through 6 show comparable estimates using the level of middle school enrollment as the dependent variable, rather than log enrollment.

TABLE 2 ABOUT HERE.

To explore further the changes in school enrollment before and after a new road is built, we regress log middle school enrollment on a set of relative time dummies that indicate the number of years before or after road construction in the village. The estimating equation is:

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

where τ indicates the year relative to when a new road was built (i.e., $\tau = -1$ is the year

[19]This effect reflects a treatment period of 3.7 years after a road is built, on average. The estimate is 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.

before road construction). As in Equation 1, we include state-by-year fixed effects and village fixed effects.

In Figure 2, we plot the identified τ coefficients. We omit the relative time coefficients from the year before treatment and the first year available, following the suggestion of Borusyak and Jaravel (2017). The regression above can only be estimated with two relative time coefficients omitted because all villages in our sample are eventually treated.[20] We can therefore identify trend breaks, but cannot test either average trends or pre-trends. The F-test of the pre-treatment coefficients in Figure 2, which tests for non-linear pre-trends, is insignificant (p=0.94).

FIGURE 2 ABOUT HERE.

Figure 2 shows that increases in school enrollment correspond to the timing of new road construction, and these effects appear to be persistent. The timing and persistence of this change in enrollment suggests that the treatment effects are not driven by labor demand shocks from road construction itself, which would occur as the road is being built and disappear thereafter.

VI.B Robustness: Specifications and Sample Definitions

Table 3 shows that the estimated average effect on middle school enrollment is robust to a range of empirical specifications and sample definitions. In column 1, we allow for village-specific linear time trends to control for potential differential trends across villages that receive a paved road in different years.[21] In column 2, we control for interactions between year fixed effects and baseline village characteristics: population, share of irrigated land, number of schools, log middle school and primary school enrollment, literacy rate, popu-

[20] 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. 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 pre-trend.

[21] 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 robust to the inclusion of village time trends.

lation share of Scheduled Castes, and distance to nearest town. In column 3, we expand the sample to an unbalanced panel by including villages with missing data in one or more years, and column 4 shows the unbalanced panel estimates with village-specific 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 similar in magnitude and statistical significance. The stability of the treatment effect suggests that these estimates are not driven by different types of villages being treated at different times. Appendix Table A2 reports specifications from Table 2, with district-by-year fixed effects, and shows similar estimates.

TABLE 3 ABOUT HERE.

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, which gives us a non-parametric distribution of test statistics under the sharp null hypothesis. The placebo estimates are centered around zero and, consistent with Table 2, none of the thousand estimates attains our main estimate of the effect of a new road on log enrollment (0.07 increase in log enrollment).

VI.C Robustness: Regression Discontinuity

In this section, we present regression discontinuity estimates of the impact of new roads on schooling. Under PMGSY road construction 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, however, and even then, states followed the guidelines to different degrees because there were often several conflicting guidelines.[22] 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 treated sooner. Population above a guideline threshold is therefore an imperfect predictor of treatment status.

Figure 3 shows the relationship between the probability of receiving a new road by 2011 and the population relative to the treatment threshold. There is a clear discontinuity in treatment status at the population threshold. By contrast, 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.[23]

FIGURE 3 ABOUT HERE.

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

(3)

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

$Y_{i,s,t}$ is log enrollment in village i, region s, and time t; P is the population threshold; $pop_{i,s,2001}$ is baseline village population (i.e., the running variable); $X_{i,s,2001}$ is a vector of village controls measured at baseline; and η_s is a region fixed effect.[24] The change in the

[22] 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.

[23] 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 A3 and Figure A4 present regression discontinuity estimates and graphs showing that baseline village covariates do not vary systematically at the treatment threshold.

[24] 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).

outcome variable across the population threshold P 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 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, which shows an increase in enrollment above the treatment threshold.

FIGURE 4 ABOUT HERE.

In Table 4, panel A presents regression discontinuity estimates for the pooled 2011-2015 sample.[25] 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 impact on log middle school enrollment from crossing the population threshold. Column 3 presents the IV estimate, which yields a large but imprecisely estimated impact of road construction on middle school enrollment (p=0.103).

TABLE 4 ABOUT HERE.

The regression discontinuity 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

[25]Figure A5 shows regression discontinuity 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.

of roads on enrollment for regression discontinuity complier villages is substantially higher than for villages in the diff-in-diff sample, we note that the 95% confidence interval of the regression discontinuity estimate includes the panel estimate, and we are hesitant to put a large weight on the specific point estimate.[26]

Villages may otherwise differ across the population thresholds, and so as a placebo exercise we estimate Equation 3 for states that did not follow the PMGSY population threshold guidelines.[27] Panel B of Table 4 shows that there is no substantive first stage in these states (column 1) and a reduced-form treatment effect close to zero (column 2). This provides reassurance that villages above the population threshold would not have otherwise experienced differential changes in school enrollment.

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 also make the regression discontinuity estimates less representative of impacts across India. We therefore focus on the panel specifications in the section on treatment heterogeneity below.

VI.D Average Impacts on School Achievement

Increasing middle school enrollment may not directly translate into greater learning, especially if school quality is low or if there is increased school crowding. To measure student

[26]Indeed, in Section VII we find subgroup estimates approaching this level in villages where we expect treatment effects to be particularly large. In Appendix Table A4, we estimate panel specifications on samples of villages that are similar to the regression discontinuity sample. While some of these have larger point estimates, note that it is impossible to set the sample to the regression discontinuity complier villages, as we do not know which villages above the population threshold would be untreated at lower populations, and which villages below the threshold would be treated at higher populations. The regression discontinuity sample also includes villages that never received roads, whereas our main panel estimates use only villages that received roads at some point.

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

learning, we estimate impacts on student examination outcomes. 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 end-of-year exams, which were required to certify completion of middle school. Column 1 shows the estimated effect of roads on the log number of students who appear for the exam. Column 2 shows the effect on the log number of students who pass the exam, and column 3 shows the effect on the log number who score high marks.[28] For exam appearance and exam passing, we find similar effects to the enrollment effects: six percent more students take and pass exams in villages after the construction of a new paved road.[29] We find a positive but smaller three percent increase in the number of students scoring high marks.

TABLE 5 ABOUT HERE.

The estimated impacts on exam outcomes reflect the net impact on student 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 fewer high marks), 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 they are of lower ability than students not on the margin of dropping out), but students who would have stayed in school regardless of road construction are now performing better on exams. Non-marginal students could perform better, for example, if they begin to perceive higher returns to human capital accumulation due to increased access to labor markets outside the village. It is difficult to disentangle these two scenarios. Under both interpretations, we can reject the possibility that school enrollment is increasing without corresponding increases in academic achievement.

[28] Sample size is smaller for the exam estimates than for enrollment estimates because we were only able to obtain examination results for all states in our sample for the years 2004-2009. In each case, we report effects on the log number of students plus one. The estimates are similar for an unbalanced panel.

[29] The number of students achieving these exam outcomes is smaller than the enrollment effects because for every ten students enrolled in the 8th grade, only six appear for the exam, five pass the exam, and two pass the exam with distinction.

VI.E Impacts on Primary School Outcomes

In this section, we explore impacts on primary school enrollment and exam scores. In panel A of Table 6, columns 1 through 3 show the estimated difference-in-differences impact of road construction on log primary school enrollment. We do not find effects on primary school enrollment, consistent with our expectation that children under the age of twelve have few labor market opportunities. Columns 4 and 5 show reduced-form and IV results from the regression discontinuity estimation, which also do not indicate an impact on primary school enrollment. The precision of the regression discontinuity estimate is much lower, however, so we cannot reject meaningful regression discontinuity impacts in either direction. The corresponding regression discontinuity figure is in panel A of Appendix Figure A6.

TABLE 6 ABOUT HERE.

Panel B of Table 6 shows estimated impacts on primary school completion exam outcomes. We find weakly positive impacts on student exam outcomes with point estimates between 2 and 3 percent, but they are of marginal statistical significance. The results are suggestive of increased effort among enrolled children, which could be due to anticipated increases in attending middle school or anticipated increases in returns to education in the labor market.

VII Mechanisms

VII.A Human Capital Investment Incentives

In this section, we examine the mechanisms underlying the estimated average impact of new rural roads on human capital accumulation. Our analysis is guided by the conceptual framework outlined in Section II. We begin by focusing on three primary channels and sources of heterogeneity: a negative impact on human capital investment through increased opportunity costs of schooling, a positive impact on human capital investment through increased returns to education, and an impact on human capital investment through income or liquidity effects. We analyze the combined impact of income and liquidity effects because, given the available data, it is difficult to distinguish between the effects of higher lifetime income and greater cash-in-hand. Our estimation focuses on identifying subsets of villages where each mechanism is likely to be especially prominent.

To examine these mechanisms, we begin by assuming 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 low-skill wage gap between the village and surrounding market is high, the village low-skill wage will rise more than if the low-skill wage gap is small. We therefore expect the largest increases in the opportunity cost of schooling to occur in places with the largest gaps in low-skill wages between the village and its surrounding labor market. We proxy for 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. We use data on urban and rural wages 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 expected size of the returns to education effect, we again aim to identify the difference in returns to education between each village and its regional labor market. The underlying assumption is that a new rural road will shift the returns to education in a village toward the returns to education in the broader regional labor market. We calculate district-level returns to education by running Mincerian regressions at the district level, separately for individuals in rural and urban areas, using data from the 55th round of the NSS. We call this difference the urban-rural returns gap, or the skill premium gap.[30] We assume the returns to education effect is stronger when this skill premium gap is higher.

Finally, to proxy for the importance of income and liquidity effects, we assume that households with few assets are more likely to be liquidity constrained, and that a given change in wages for these households has a larger income effect. We measure average baseline assets at the village level using data from the 2002 Below Poverty Line Census. We define a village

[30]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 land owned or including state fixed effects. We exclude districts with no urban data.

as having low assets (and thus high potential income and liquidity effects) if the share of households reporting zero durable assets is above the sample median.[31] Similarly, we define binary indicator measures for the opportunity cost proxy and the returns to education proxy, based on whether each proxy is above the sample median.

We then estimate our previous panel regression, including additional interaction terms between the treatment indicator for road construction and the indicator variable for each of the three mechanisms. If the estimated impact of road construction varies with our proxy measure, and the interaction term is important in magnitude, it provides suggestive evidence of that mechanism being an important channel through which new roads affect schooling decisions.

Table 7 shows the results from estimating these interactions. Column 1 repeats the main specification, without interaction terms, in the sample for which each interaction term is measured.[32] Columns 2 through 4 include each interaction term separately, and column 5 includes the three interaction terms together.

TABLE 7 ABOUT HERE.

The estimated interaction effects are consistent with the predictions from a standard model of human capital investment. Road construction has the smallest effects on middle school enrollment in districts where these roads are expected to most raise the opportunity cost of schooling. The largest effects of road construction on middle school enrollment are in districts where road connections are expected to raise the skill premium the most and to have the largest income and liquidity effects. The opportunity cost effect interaction is strongly statistically significant ($p < 0.01$), the returns to education effect interaction is marginally statistically significant ($p = 0.08$), and the income/liquidity effect interaction is in the expected direction but statistically insignificant ($p = 0.37$).[33] The greater magnitude of the opportunity cost effect may be in part because the urban-rural wage gap is much larger than

[31]The surveyed assets are a radio, a television, a telephone, and a motorcycle.
[32]This analysis excludes districts without NSS data for both urban and rural areas in 1999-2000.
[33]For completeness, Appendix Table A5 shows results by quartile of each mechanism proxy.

the urban-rural skill premium gap (see Appendix Table A1).

While the estimated interaction effects are consistent with a standard model of human capital investment, note that there could be other district-level characteristics that influence the size of treatment effects and are correlated with the proxies we use. For example, high rural-urban wage gaps are correlated with greater remoteness, worse infrastructure, lower returns to education, and tend to be in the North. The estimated interaction effects are robust to the inclusion of interactions with these other variables, but there are myriad other unobserved district-level characteristics. Therefore, we see these estimates not as definitive but rather as suggestive indications of the mechanisms underlying the main estimates.

By fully interacting the three binary mechanism variables, we can obtain treatment effects in eight partitions of the sample based on the model's predictions. Table 8 shows 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 effect, low returns to education effect, and low income/liquidity effect, which represents 9% of all villages. This is precisely the group where a standard model predicts that roads would have the most adverse effects on education. Treatment effects are positive and significant only in the 39% of villages where at least two of the mechanisms are favorable.

TABLE 8 ABOUT HERE.

This treatment heterogeneity is consistent with the heterogeneity in results from earlier research 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 dominant mechanism in these studies is likely an increase in the return to education, since spoken English is a requirement for these call center jobs. Conversely, Shah and Steinberg (2017) find that children are more likely to attend school in drought years, when there are fewer agricultural jobs available. The more important mechanism in that setting is likely to be an opportunity cost effect, as the low-skill wage is declining when there are fewer agricultural jobs available (or less need for children to substitute into home production while parents work agricultural jobs). The small negative effects on schooling from India's workfare MGNREGA program (Islam and Sivasankaran, 2015; Das and Singh, 2013; Li and Sekhri, n.d.; Shah and Steinberg, 2015; Adukia, 2018) are likely driven by similar mechanisms, as MGNREGA hires workers for labor-intensive construction and increases labor demand for lower-skill workers. The same model helps to explain the variation in estimated impacts on schooling 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 jobs in Bangladesh (Heath and Mobarak, 2015)). In each case, individual schooling choices appear to respond to the skill requirements of the 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 our finding 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 Potential Mechanisms

In this section, we explore several other mechanisms through which road construction might impact schooling outcomes.

School Quality. We have focused on how road construction affects the incentives for human capital investment (*i.e.*, changes in the demand for schooling), though road construction could also affect school quality or the number of schools available (*i.e.*, changes in the supply of schooling). We use village-level DISE data to examine impacts of road construction on the number of schools and on measures of school quality, as proxied by physical characteristics of a school.

Appendix Table A6 reports no impact of road construction on the number of schools, and no systematic impact on school infrastructure characteristics.[34] While a minority of specifications show statistically significant effects on school infrastructure, none approach the size of the enrollment effects presented above. The standard errors in column 1 rule out a 2 percentage point change in the presence of any of these kinds of infrastructure. Overall, we do not find systematic evidence that road construction substantially affects the number of schools and their physical characteristics.

Other Government Programs. To alleviate 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. We observe village public goods only in the decennial population census, and so we are unable to estimate panel regressions. Instead, we use the regression discontinuity approach to test for discontinuities in village public goods around the PMGSY eligibility thresholds. Appendix Table A7 shows no discontinuity in the presence of schools, as above using DISE data, and also shows no discontinuity in village access to electric power, a primary health center, or a commercial bank. We can rule out a one percentage point increase in the existence of primary or secondary schools, health centers and banks, and a four percentage point increase in middle schools and electrification status.

Migration. We next explore whether the estimated increases in middle school enrollment could be driven by increased migration into villages that receive roads, or reduced outmigration from those villages. Note that we did not find impacts of road construction on primary school enrollment (Table 6), which would presumably also be affected if the increase in middle school enrollment was driven by migration responses to road construction.

To test for migration effects, we use the regression discontinuity specification to examine

[34]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 have access to a particular kind of infrastructure.

impacts of roads on village population in 2011.[35] The result is shown in panel A of Appendix Figure A6; there is no effect of the treatment threshold on population. The point estimate is close to zero, and the 95% confidence interval rejects the net entry or exit of more than four people from a treated village. Changing migration patterns are thus unlikely to explain the effects of roads on school enrollment.

Cross-Village Displacement. Relatedly, we explore whether our estimated impacts on school enrollment could be driven by displacement effects, in which increased enrollment in treated villages is counterbalanced by decreased enrollment in nearby villages. We calculate total middle school enrollment for all other villages within a 3 km or 5 km radius of each village that received a new road.[36] Using the panel specification, columns 1 and 2 of Appendix Table A8 report the estimated impact of roads on log middle school enrollment in these surrounding villages. We do not find impacts on school enrollment in these surrounding villages; the 95% confidence interval rules out a 4% decrease in enrollment in a 3 km radius of the village, and a 2% decrease in a 5 km radius.

School Accessibility. Finally, we examine the possibility that road construction increases school enrollment by decreasing the students' costs of traveling to school.[37] Children generally walk to village schools, though paved roads could make schools more accessible, especially during the rainy season. We explore this possibility by estimating whether the impact of roads varies across villages that are more or less dispersed geographically. We expect that impacts through increased school accessibility would be more pronounced for villages in which children have further to walk to school. We measure village dispersion using village surface area, and divide the sample into villages with above-median and below-median surface area per capita.[38] Columns 3 and 4 of Appendix Table A8 show that the estimated impact of roads

[35]As with public goods, village population is measured only in the decennial censuses, so we cannot use the panel approach here.

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

[37]For example, Muralidharan and Prakash (2017) find that the provision of bicycles made girls more likely to attend middle school and high school.

[38]We find similar results using village surface area.

is similar across more dispersed and more dense villages. The point estimates are similar and we cannot reject equality between them. This suggests that a decreased cost of reaching school is not a main channel through which road construction is impacting school enrollment.

Similarly, road construction may increase middle school enrollment in a village by increasing its accessibility to children from nearby villages that do not have a middle school. This is related to the potential for cross-village displacement, discussed above, but this effect would not decrease middle school enrollment in those nearby villages. To explore this mechanism, we calculate the total number of school-age children within a 5 km radius of sample villages, who were living in villages without middle schools.[39] Columns 5 and 6 of Appendix Table A8 show that estimated impacts of roads on middle school enrollment are similar across villages with more or fewer under-served children in nearby villages; the point estimate is slightly higher in villages that do not have many underserved children living nearby. This provides suggestive evidence that the schooling increases we observe are not originating from nearby villages without middle schools.

VIII Conclusion

High local transportation costs are a central feature of the lives of the very poor around the world, leaving them isolated from external markets. Connecting remote villages to high-quality transportation networks is a major goal of developing country governments and international development agencies. These roads can bring access to new economic opportunities; however, a concern is that increased access to low-skill labor market opportunities could decrease investment in the human capital that is central to long-run increases in living standards and broader economic growth.

We examine this trade-off in the context of India's flagship rural road construction program, which built local paved roads to 115,000 villages in India between 2001 and 2015. These roads connected villages with nearby labor markets, potentially changing the incentives for

[39]We proxy for the number of middle school-aged children using the number of children aged 0-6 in 2001, as reported in the Population Census. We find similar results if we use total village population in villages without middle schools.

investment in human capital. We find that rural road construction increased adolescent schooling outcomes. Further, we find that a standard model of human capital investment has important predictive power for how schooling decisions respond differently across villages to the increases in economic opportunity from rural road construction. We highlight the competing influences of opportunity cost effects, returns to education effects, and income/liquidity effects.

Our analysis draws on the substantial heterogeneity in economic opportunities across India, allowing us to identify large positive effects on schooling in places where the relative return to high-skill work increases the most, as well as neutral or weakly negative effects on schooling in places where the return to low-skill work increases the most. Notably, treatment effects are negative (and small) in only a small subset of villages. Across most of rural India, local market integration substantially promoted increased investment in human capital.

Our paper also highlights an important but understudied impact of rural infrastructure investment. Investments in road improvements are usually premised on their potential to bring economic growth to rural areas, with a focus on contemporaneous economic gains in those areas. If road construction leads to increased investment in human capital in rural areas, then the long-run economic impacts will be greater than short-run estimates suggest and will reflect human capital dividends over subsequent generations.

References

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

——, "Spillover Impacts on Education from Employment Guarantees," 2018. Becker Friedman Institute Working Paper No. 2018-33.

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

Asher, Sam and Paul Novosad, "Rural Roads and Local Economic Development," 2018.

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

——, Azam Chaudhry, Shamyla Chaudry, Amit K. Khandelwal, and Eric Verhoogen, "Markup and Cost Dispersion across Firms: Direct Evidence from Producer Surveys in Pakistan," *The American Economic Review Papers and Proceedings*, 2015, *105* (5), 537–44.

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

Banerjee, Abhijit and Rohini Somanathan, "The Political Economy of Public Goods: Some Evidence from India," *Journal of Development Economics*, 2007, *82* (2).

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*, 2018, *108* (4-5), 899–934.

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

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," *The 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," 2015. 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.

Li, Tianshi and Sheetal Sekhri, "The Unintended Consequences of Employment-Based Safety Net Programs," *World Bank Economic Review (forthcoming).*

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 Effects in the Additive Model," *Oxford Bulletin of Economics and Statistics*, 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), 709–734.

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

Muralidharan, K and N Prakash, "Cycling to School: Increasing Secondary School Enrollment for Girls in India," *American Economic Journal: Applied Economics*, 2017, *9* (3). NBER Working Paper No.19305.

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 village-level variables at baseline, in the sample of villages that were matched across all analysis data sets. 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. 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 four years before each road is built and four years after. Different years are thus included for different villages, but each village has nine observations. Due to data availability, the sample in column 6 only includes villages with roads built between 2006 and 2012. All specifications have state-year fixed effects and village fixed effects. 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. 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 high marks. All specifications have state-year fixed effects and village fixed effects. Standard errors are clustered at the village level.

Table 6

Impact of New Roads on Primary School Outcomes

Panel A: 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 regression discontinuity 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 eligibility threshold, and column 5 presents the regression discontinuity IV estimates of the impact of the new road. Standard errors are clustered at the village level.

Panel B: 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 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 regression discontinuity 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 eligibility threshold, and column 5 presents the regression discontinuity IV estimates of the impact of the new road. Standard errors are clustered at the village level.

Table 7

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 measures of different potential mechanisms. The size of the opportunity cost effect is proxied by the district-level mean low-skill urban wage minus the mean low-skill 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. 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 come from the 55th round of the NSS Employment and Unemployment Survey (1999-2000), and asset data are from the Below Poverty Line Census (2002). All specifications have state-year fixed effects and village fixed effects. Standard errors are clustered at the village level.

Table 8

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)	

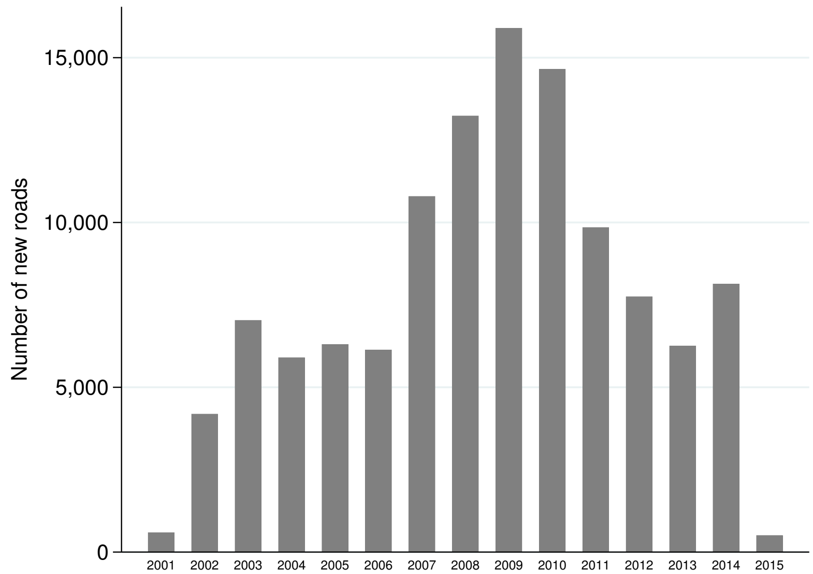
*p<0.10, **p<0.05, ***p<0.01

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 (as described in Table 7). 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. All specifications have state-year fixed effects and village fixed effects. Standard errors are clustered at the village level.

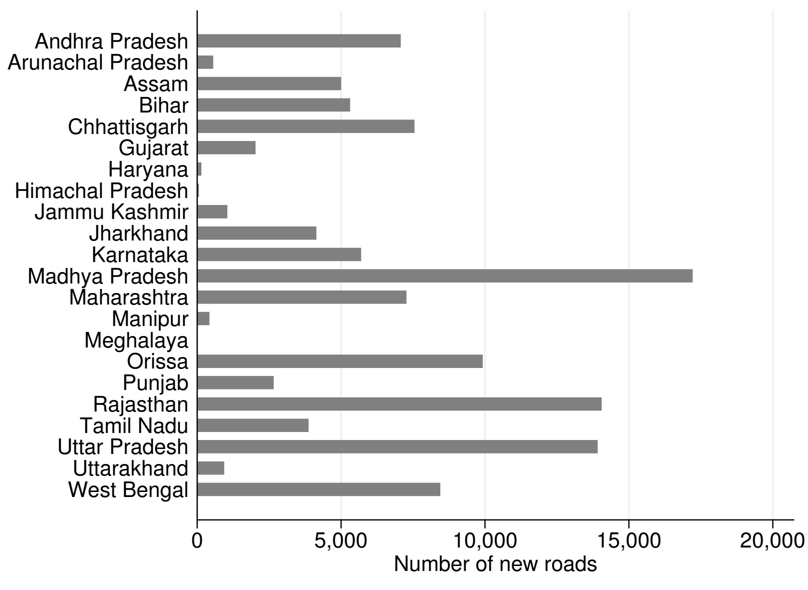
Figure 1

Sumamry of Road Construction under PMGSY

Panel A



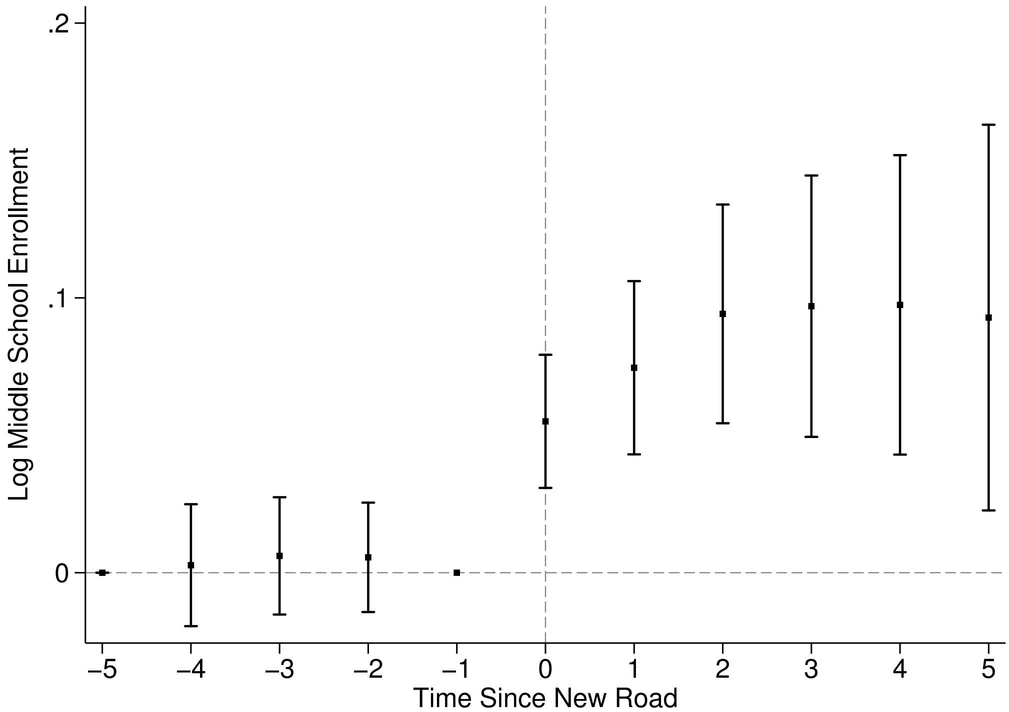
Panel B



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

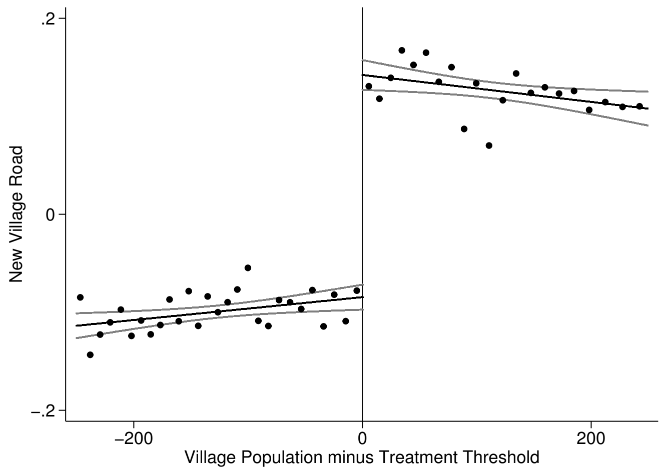
Impact of Roads on Middle School Enrollment:
Treatment Effect Time Series



The figure shows coefficient estimates from a panel regression of log middle school enrollment on a set of indicator 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. 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

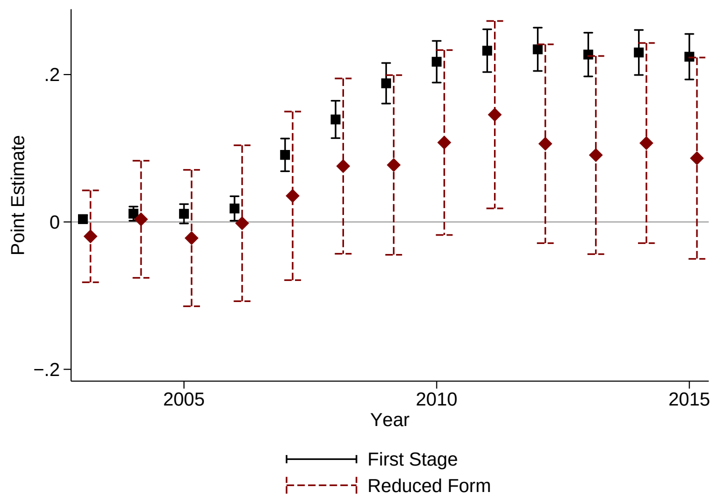


The graph plots the conditional expectation function of an indicator variable indicating that a village has received a road before 2011, conditioning on the village population as reported in the 2001 Population Census. Each point represents the mean of all villages in the given population bin (328 villages 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.

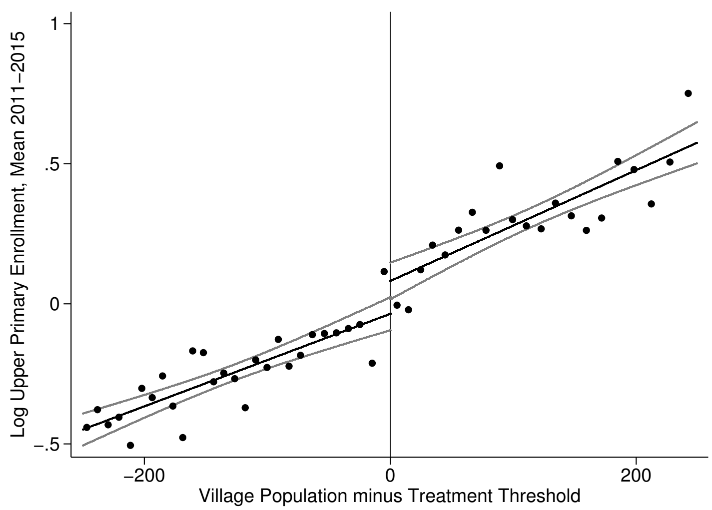
Figure 4

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 regression discontinuity 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.

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

Anjali Adukia*

Sam Asher†

Paul Novosad‡

March 2019

*University of Chicago, 1307 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

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); within each group, the Mincerian return 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. Standard errors are clustered at the village level.

Table A3

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 A4

Panel Estimates in Regression Discontinuity Sample

Dependent Variable	Full Sample	RD States	RD Villages	RD Villages with Untreated
	(1)	(2)	(3)	(4)
New Road	0.070***	0.082***	0.040**	0.165***
	(0.015)	(0.016)	(0.020)	(0.021)
N	146678	110740	71148	165606
r2	0.80	0.81	0.80	0.80

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

The table shows panel estimates of the effect of new road construction, focusing on samples that are more similar to the regression discontinuity analysis in Table 4. Column 1 repeats the main estimate from column 1 of Table 2. Column 2 limits the sample to the five states used in the regression discontinuity analysis. Column 3 limits the sample to the set of regression discontinuity villages with roads completed between 2003 and 2015. Note that this sample excludes the untreated regression discontinuity villages. The majority of villages in this sample were connected between 2007 and 2009, limiting the variation available for the difference-in-differences estimation. Column 4 limits the sample to the set of villages in the regression discontinuity sample, but (unlike Column 2 and unlike the other panel estimates) includes villages that never received roads. Thus, unlike the other panel estimates in the paper, this estimation compares treated villages to never-treated villages (as well as comparing pre- and post-treatment periods in treated villages).

Table A5
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 low-skill urban wage minus the mean low-skill 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. Standard errors are clustered at the village level.

Table A6

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 A7
Regression Discontinuity 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 difference in other public goods across the PMGSY population treatment threshold, using Equation 3. The dependent variable, column by column, is: (1) presence of primary school; (2) presence of middle school; (3) presence of secondary school; (4) village access to electric power; (5) presence of a primary health center; and (6) 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 A8

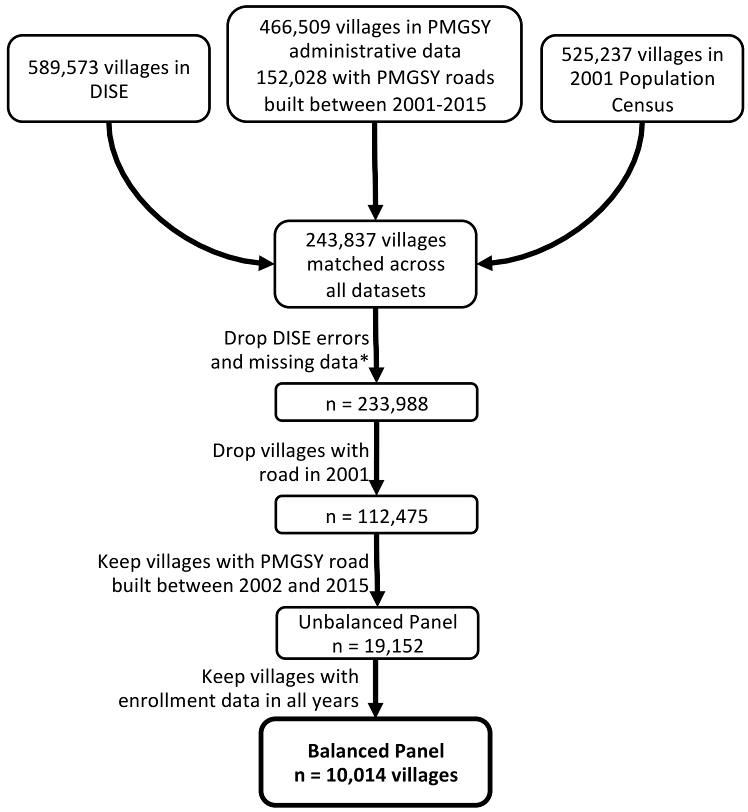
Impact of Roads on Middle School Enrollment:
Spatial Effects

	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

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

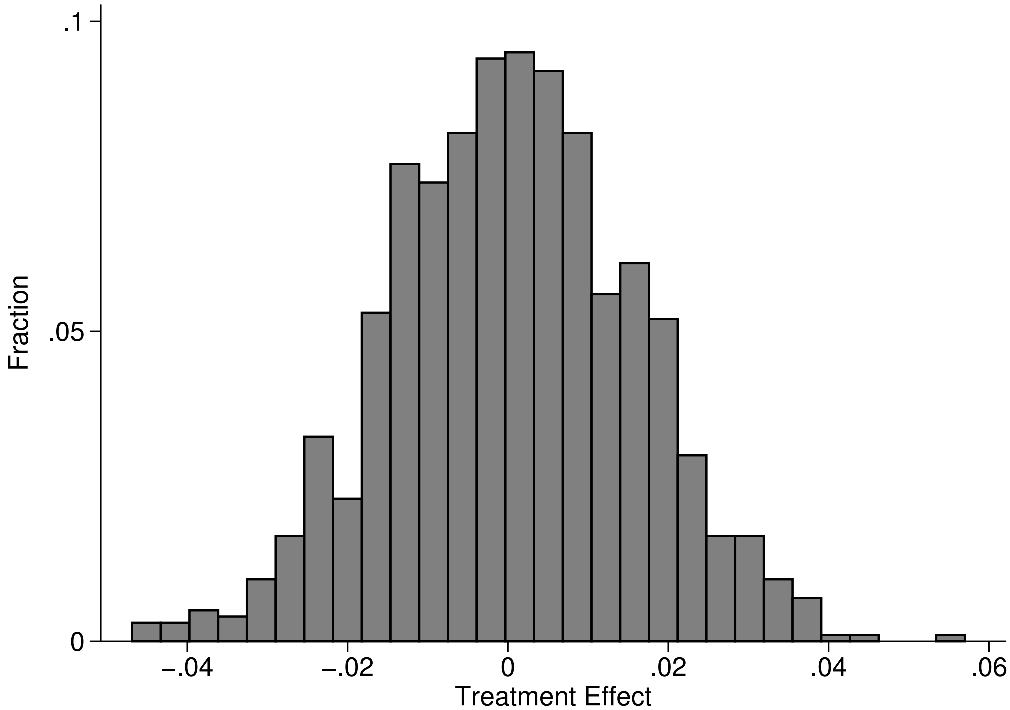
This table shows panel estimates of the impact of road construction on log middle school enrollment. Columns 1 and 2 show the impact of a new road on middle school enrollment in nearby villages, measured as those villages within a 3 km or 5 km radius, respectively. Columns 3 and 4 divide the sample into villages with above-median land area per capita and below-median land area per capita, and report effects separately. Columns 5 and 6 divide the sample into villages according to the number of children in nearby 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 under-served children. All specifications include state-year fixed effects and village fixed effects. Standard errors are clustered at the village level.

Figure A1
Sample Construction



The figure shows how the sample was constructed from the original datasets. DISE refers to the District Information System for Education. PMGSY refers to the Prime Minister's Village Road Program. Observation counts indicate the number of villages at each stage. * Observations were dropped if DISE reported enrollment for grades one through eight to be greater than 60% of the village population (or greater than the 99th percentile.

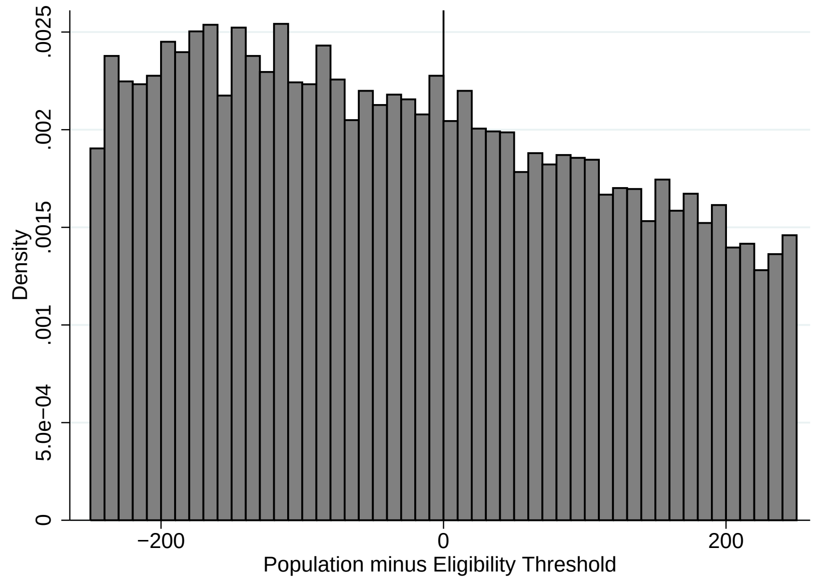
Figure A2
Panel Estimates of Effect of Roads on
Middle School Enrollment: Permutation Test

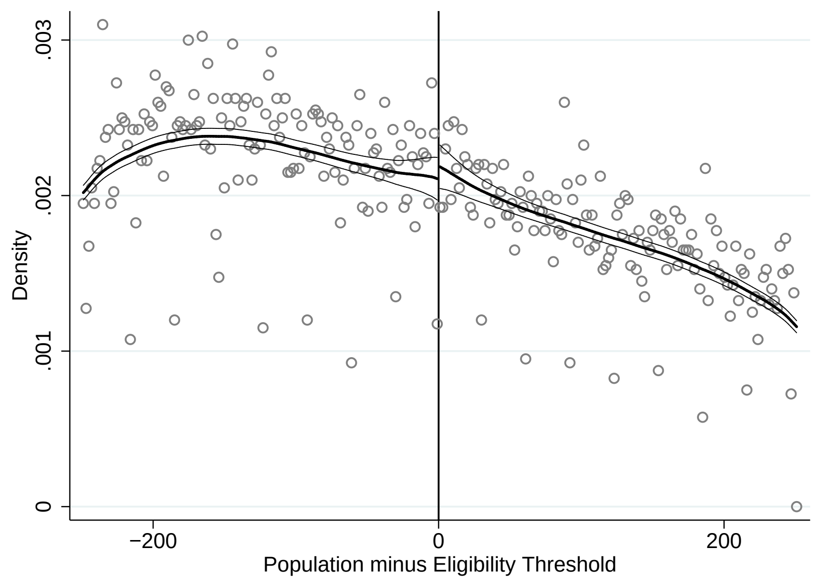


The figure shows the distribution of estimates from a placebo permutation test of the main panel specification presented in column 1 of Table 2. For each village in the main sample, we randomly generated a placebo year of road completion, and then estimated Equation 1. We ran this estimation 1000 times; 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

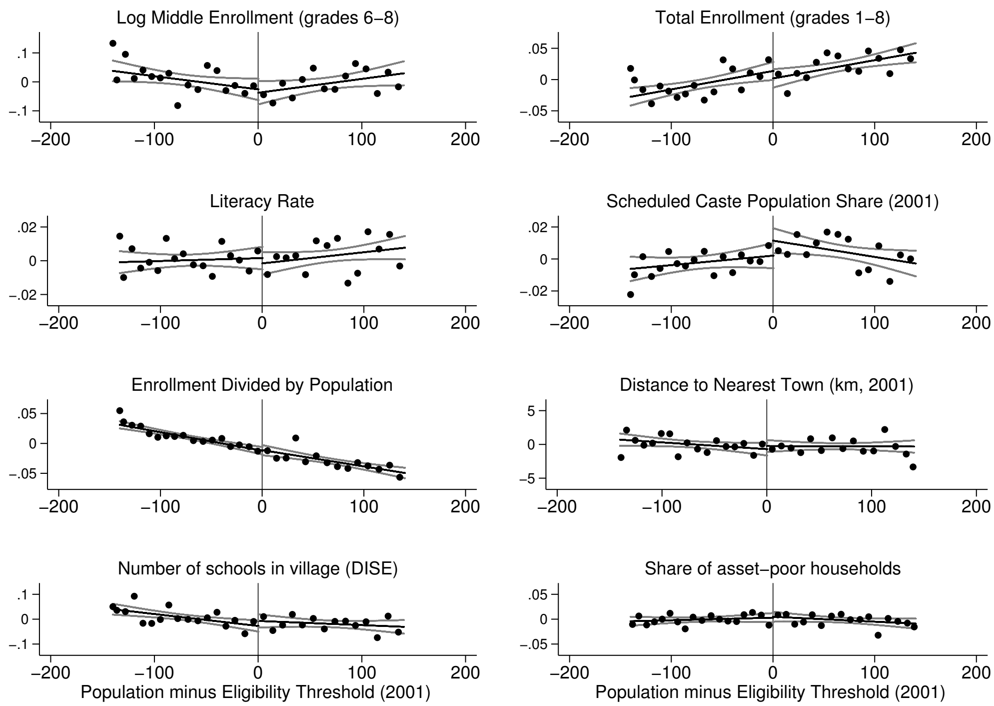




The figures show the distribution of village population in the set of villages in our sample. The top panel shows a histogram of village population, centered around the treatment threshold. In the bottom panel, we plot a non-parametric regression to each half of the distribution following McCrary (2008),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

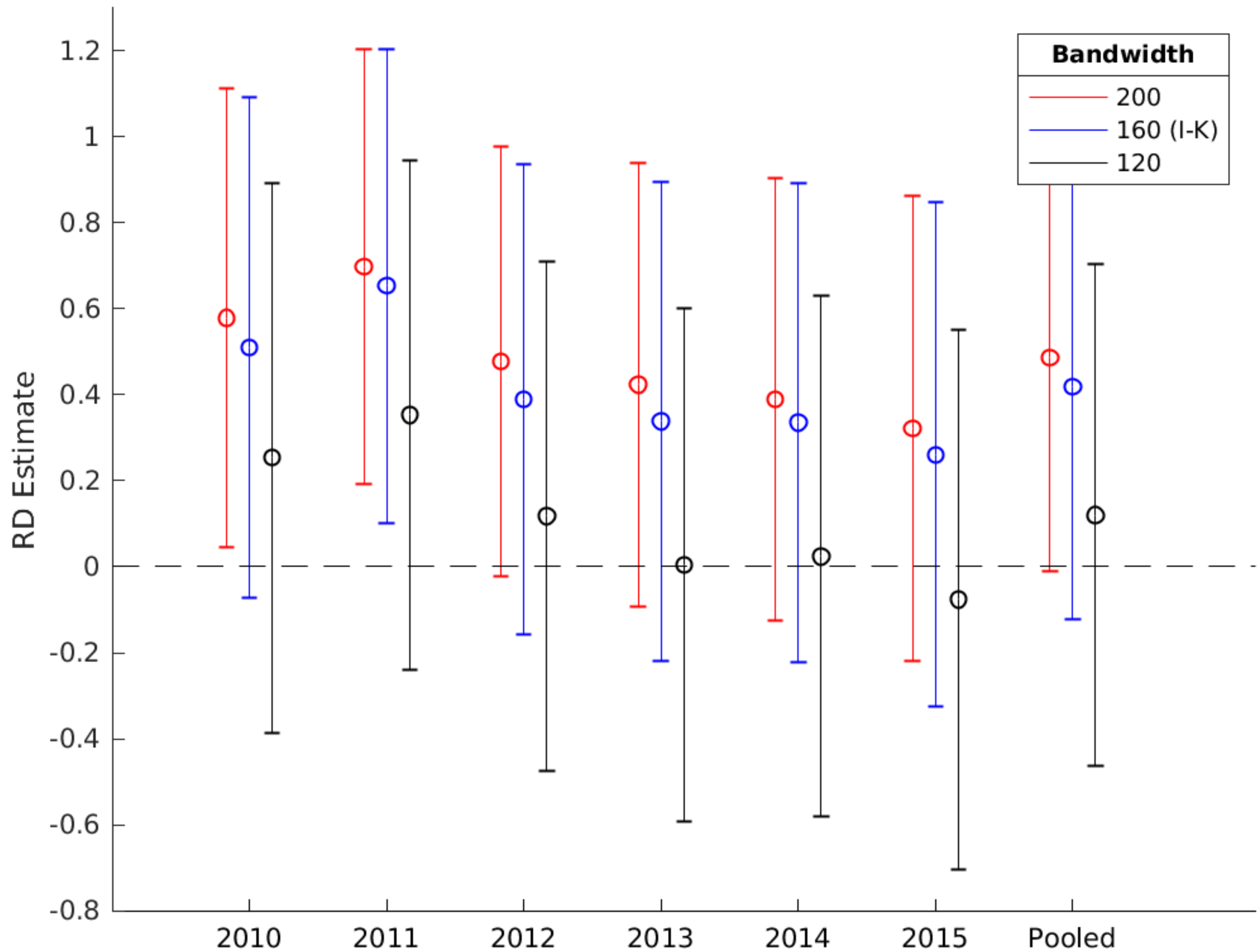
Figure A4

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:
Regression Discontinuity Estimates by Year and Bandwidth

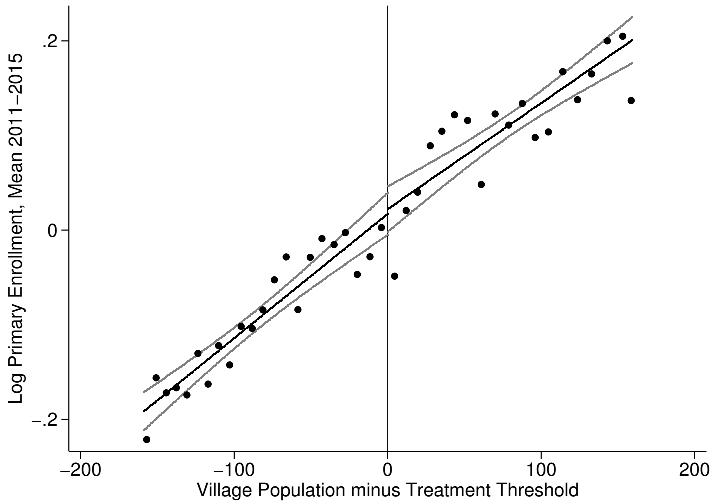


The figure shows IV estimates from Equation 3, estimated on different sample years, and at bandwidths 25% higher and lower than the optimal bandwidth of 160 selected with the algorithm of Imbens and Kalyanaraman (2012). Each point represents a single regression discontinuity estimate of the impact of 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.

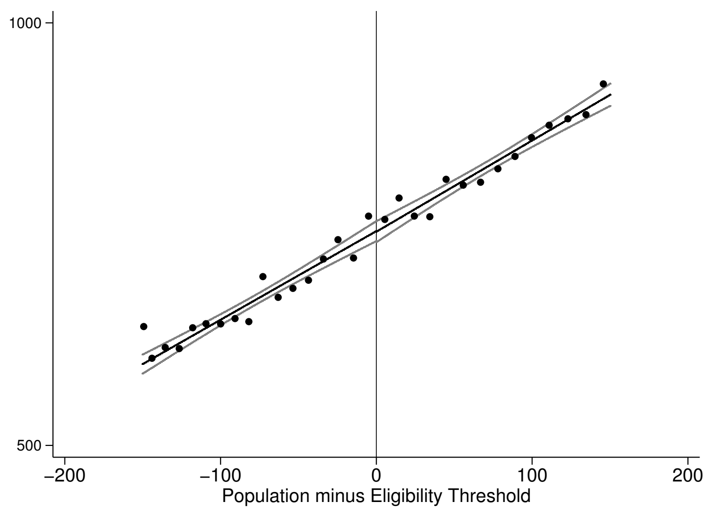
Figure A6

Regression Discontinuity Reduced Form:
Population and Primary School

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



Panel B: Log Population (2011)



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.

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

Anjali Adukia†

Sam Asher‡

Paul Novosad§

March 2019

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 effects 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. Heterogeneity in treatment effects supports a standard human capital investment model: enrollment increases most when nearby labor markets offer high returns to education and least when they imply high opportunity costs of schooling.

JEL Codes: I25; O18; J24.

*We thank Martha Bailey, Chris Blattman, Liz Cascio, Dave Donaldson, Esther Duflo, 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, EBRD, EEA, the Federal Reserve Banks of Chicago (CHERP) and New York, Georgetown University, Michigan State University, NEUDC, the NBER Education Program Meetings, Northwestern University School of Law, the OECD, PAA, PacDev, Stanford, University of California-Berkeley, University of Chicago, University of Connecticut, University of Illinois at Chicago, University of Michigan, University of Missouri, the Urban Institute, and Yale, and the anonymous referees for their helpful comments and suggestions. We thank Srinivas Balasubramanian, Jack Landry, 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, 1307 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 connections to international markets. A key trade-off for individuals 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. As educational investment responds to market integration, it shapes the long-run economic impacts of policies that are increasingly integrating markets in developing countries.

We examine the human capital investment response when a paved road is built to a previously unconnected village, effectively connecting it to outside markets. India's national rural road construction program (PMGSY) built high quality roads to 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. The impacts of new road connections on schooling are theoretically ambiguous: they may raise the returns to education, raise the opportunity cost of schooling, and/or have important income or liquidity effects.

A major challenge in estimating causal effects of new roads is the endogeneity of road placement. If roads are targeted to wealthier or poorer regions, for example, then comparisons of villages with and without roads will be biased. To overcome this bias, we exploit the timing of road completion in each village, estimating a panel regression with village and

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

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

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. We thereby compare educational outcomes in villages before and after a road is built, flexibly controlling for time-variant regional shocks and static differences between villages that receive roads in different years.

We use village-level school enrollment data from India's national annual census of primary and middle schools (District Information System for Education, DISE, 2002-2015). Through a combination of human and machine fuzzy matching, we linked DISE data to administrative data from the national rural road construction program. The result is a panel of 300,000 villages across 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 heterogeneous impacts in subsamples of the data. Our sample spans a broad range of economic conditions in India today, similar to the variation across many places worldwide that remain unreached by paved roads.

We find that road construction significantly increases middle school enrollment. 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.

We do not find enrollment effects for primary school children, for whom there are fewer labor market opportunities. We do find small increases in primary school performance, however, suggesting that students may be increasing school effort on the intensive margin.

[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.

We then explore heterogeneity in the treatment effects on middle school children, guided by predictions from a standard model of human capital investment. The model predicts four primary mechanisms through which educational investment in rural areas may be affected by road connections to nearby labor markets. We model roads as leading to factor price equalization across areas, which is then predicted to: (i) raise the low-skill 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] The model suggests that the importance of each of these effects will be different across regions, depending on local market characteristics. Newly connected villages will experience larger opportunity cost effects when the urban-rural low skill wage gap is large. Returns to education effects will be larger when the urban-rural gap in Mincerian returns to education is larger. To predict the expected importance of income and liquidity effects, we use a measure of asset poverty.[5]

The estimated variation in treatment effects across these three measures supports the predictions of the model. Partitioning our data according to these measures, we estimate substantial treatment effect heterogeneity across villages, with effects that are positive and statistically significant in 39% of villages and positive but insignificant in 52% of villages. Market integration has (small and statistically insignificant) negative effects only in the 9% of villages where 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.

We explore several other treatment mechanisms, for which we do not find support in the data: (i) supply-side improvements in school infrastructure; (ii) migration; (iii) displacement to/from nearby villages; and (iv) improved access for children on the outskirts of villages.

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

[4]Because roads may change factor prices in many markets, many other effects are also possible. We focus on effects predicted from the literature on road construction.

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

to drive further long-run growth through increased educational attainment. Enrollment and exam performance respond positively to increased economic opportunities. Our results also provide a causal mechanism that underlies the strong correlation around the world between education, growth, and trade.

This study is related to the literature on the impact of labor demand shocks on schooling decisions, which finds both positive and negative schooling impacts from new economic opportunities.[6] The estimated heterogeneity in treatment effects in our study is consistent with the heterogeneity found in the literature, and well explained by the standard human capital model: new labor market opportunities affect the opportunity costs of schooling, but also affect the long-run benefits of schooling and demand for schooling through income and liquidity effects.

Our paper also contributes to the literature on the development impacts of transport infrastructure.[7] Relative to earlier work on roads and schooling, our large village-level sample and research design allow a more precise estimation of the causal effects of road construction. Our results suggest that road construction and domestic market integration may have greater long-run impacts on economic development by increasing educational investment. Finally, we contribute to a wide body of research on improving educational attainment in

[6]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 (Cascio and Narayan, 2015; Atkin, 2016; Shah and Steinberg, 2017). Our estimates are also related to impacts on human capital accumulation from India's national public works program (MGNREGA), which has found small decreases in enrollment for middle school students across India (Das and Singh, 2013; Islam and Sivasankaran, 2015; Li and Sekhri, n.d.; Shah and Steinberg, 2015; Adukia, 2018).

[7]Some examples include Jacoby (2000); Jacoby and Minten (2009); Gibson and Olivia (2010); Mu and van de Walle (2011); Casaburi et al. (2013); Donaldson and Hornbeck (2016); Donaldson (2018). For a detailed review, including studies on the impacts of highways and regional roads, see Hine et al. (2016). Asher and Novosad (2018) find that PMGSY road construction leads to changes in occupations but has little effect on village assets, incomes, or consumption using regression discontinuity exploiting village population thresholds. Mukherjee (2012) uses a similar approach to find that PMGSY road construction in India increases school enrollment. We present comparable regression discontinuity estimates, but we focus on panel estimates that have much greater statistical precision and allow for analysis of treatment heterogeneity. Using district-level data from India, Aggarwal (2018) 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.

developing countries (see Evans and Popova (2016) and Glewwe and Muralidharan (2016) for reviews of this literature). Our results highlight that investments outside the education sector can have important 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 results, Section VII explores the mechanisms suggested by the model of human capital investment, 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 both by characteristics of villages and by characteristics of local labor market conditions outside the village.

In this framework, the key decision point is an individual's trade-off between the long-run benefits of schooling and the short-run return to labor. A two-period model is sufficient to highlight the essential comparative statics. In the first period, a person chooses between working for a low-skill wage and obtaining schooling. In the second period, the person works and receives either a high or a low wage, depending upon the schooling choice in the first period. The person consumes in both periods, drawing from an initial endowment and wages earned in each period that the person works. The person can save at some interest rate, but may be restricted in borrowing. The person'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.[8]

[8]This framework underlies much of the theoretical literature on child labor and human capital invest-

When a village becomes connected to an external market via a new road, there is a change in the parameters underlying this trade-off between education and early participation in the labor market. Reduced transportation costs affect worker wages in both periods by inducing factor price equalization across areas.[9] In equilibrium, urban areas have higher wages than rural areas for both low- and high-skilled workers, and higher Mincerian returns to education (see Appendix Table A1).[10] Connecting a village to its external market is therefore likely to: (i) increase the low-skill wage; and (ii) increase the returns to education.[11]

An increase in the low-skill wage raises the opportunity cost of schooling and discourages human capital investment, which we call the opportunity cost effect. A relative increase in the high-skill wage raises the returns to education and encourages human capital investment, which we call the returns to education effect. Changes in wages could also affect human capital investment through income effects or liquidity effects. As wages rise, income effects will increase human capital investment if schooling is a normal good. Increases in household liquidity may also affect human capital investment if credit constrained households cannot afford to pay school fees or require children to work. In principle, these effects could go in either direction, but based on urban-rural wage gaps and skill gaps in India, we expect the opportunity cost effect to reduce schooling, and the returns to education, income, and liquidity effects to increase schooling.

Predictions for how human capital investment is affected by factor price equalization in goods and capital markets are less clear, as many prices can change simultaneously.[12] Rural

---

ment decisions. See, for example, Ranjan (1999) or Baland and Robinson (2000). We abstract away from intra-household bargaining.

[9] 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. Asher and Novosad (2018) show that new PMGSY roads increase the number of people working for wages outside of villages.

[10] 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, unobserved education quality differences are unlikely to 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).

[11] We can think of these effects as changes in real wages, such that changes in local goods prices due to new roads are subsumed in the above effects.

[12] For example, in general equilibrium, there may be various changes in the prices of different intermediate

road construction could also affect schooling decisions through many other channels, such as information flows, marriage markets, and healthcare access, but we focus on impacts through labor markets and wages.

To explore which of these mechanisms are more important, we identify places where these mechanisms are likely to generate heterogeneous impacts on schooling decisions. We generate measures of regional labor market conditions, which should influence the magnitude of changes in low-skill wages and the returns to education when a village becomes more integrated with nearby labor markets. The opportunity cost effect should be particularly large when the low-skill regional wage is much larger than the low-skill wage in the village becoming connected. The effect on returns to education should be larger when regional returns to education are greater than returns to education in the village becoming connected. We expect that income and liquidity effects on schooling would be greater in villages that are liquidity constrained or have low incomes, though the economic opportunities created by new roads may also differ in these villages. In the absence of shocks that separately affect liquidity and income, these liquidity and income effects are difficult to disentangle (Edmonds, 2006), and so we consider them together.

III Background and Details of the Road Construction Program

School enrollment increased substantially in India over our study period, from 2002 to 2015, paralleling a global increase in educational attainment. Increasing educational attainment has been a national priority in India, with several national initiatives aimed toward achieving universal primary education. Educational attainment and rates of economic development vary substantially across India. Indian policy-makers in the past have 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 and Somanathan, 2007; Dreze and Sen, 2013).

Many rural villages have limited connections to regional markets even while major cities in

goods and final goods. Capital market integration could also affect interest rates, changing the impact of liquidity constraints and changing the return on savings.

India have become increasingly connected to world markets. High costs of road construction and rapid degradation have historically constrained the ability of the Indian 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 Village Road Program, or PMGSY), a national program that aimed to eventually build a paved road to every village in India. While the federal government issued implementation guidelines, decisions on village-level allocations of roads were ultimately made at the district level. The unit of targeting for road construction was the habitation, which is the smallest rural administrative unit in India. A village is typically comprised of between one and three habitations; there are approximately 600,000 villages in India and 1.5 million habitations. 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) little economic data is available at the habitation level.

Road construction was targeted initially toward villages with larger populations. In some states, this took the form of a strict population threshold for road construction eligibility, while other criteria were used in other states. Given the program rules, early-treated villages tended to have larger populations, but were not substantially different from late-treated villages in other characteristics.[13] There were initially 80,000 villages eligible for the road construction program, a number that has grown as guidelines have been expanded to include smaller villages.

By 2015, over 115,000 villages had paved roads built or upgraded under the PMGSY program. These construction projects were most often managed through subcontracts with larger firms, and were built with capital-intensive methods and external labor; the building

[13] 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.

of the road itself was therefore not a major local labor demand shock. These PMGSY roads are distinct from new roads being built under the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), which are lower quality roads built with labor-intensive methods.[14]

Figure 1 shows the distribution of road construction by state and over time. The median road length was 4.4 kilometers. Given the difficult terrain around many of these villages, a new paved road represents multiple hours saved on a round trip to or from the village.

FIGURE 1 ABOUT HERE.

IV Data

We constructed a village-level panel data set, combining data on road construction with education outcomes and other village characteristics. We matched an annual census of Indian schools, the District Information System for Education (DISE, 2002-2015), to administrative data from the implementation of the road program (2001-2015) and three successive Indian Population Censuses (1991, 2001, 2011). We matched locations based on village, block, and district names using a set of fuzzy matching algorithms.[15]

DISE is an annual census of primary and middle schools in India. It includes data on student enrollment, exam completion, and school infrastructure. This data set 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 cover every registered Indian government primary and middle school beginning in 2005.[16] We also have DISE data for a smaller sample of schools from 2002-2004, a period when the

[14] Major highway projects during this period, such as the Golden Quadrilateral, were planned and executed independently of PMGSY. There is not evidence of coordination of PMGSY road construction with the construction of the Golden Quadrilateral or other district road improvement projects.

[15] 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 is posted at github.com/paulnov/masala-merge. 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. We matched 80 percent of census blocks; within census blocks, we matched 81 percent of villages.

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

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.[17] DISE data are based on interviews with school headmasters, and there are potential concerns of misreporting and inflated enrollment. Because new roads lower the cost of monitoring enrollment numbers, we expect that changes in misreporting would bias downward the estimated impacts of roads. It is also unlikely to affect only middle school students.

Our main 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. Our main focus is on outcomes for middle school children (grades 6-8), because the transition to middle school is a natural breakpoint in a child's schooling at which educational milestones are often measured, but we report some outcomes for primary school children as well. Younger children also have fewer labor market opportunities. DISE did not report enrollment information for high school over our sample period. DISE does not report the total number of school-age children in a village, so we are unable to calculate enrollment rates directly. 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 end-of-school examinations for primary schools and middle schools. These exams 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 high marks. We also use DISE data on school infrastructure, which describe the school-level presence of piped water, sanitation facilities, electricity, a library, a computer, a boundary wall, and a playground.

For data on road construction, we use administrative records of the PMGSY program that

[17] Our preferred sample drops villages that reported total enrollment (first through eighth grades) greater than 60 percent of total population (the 99th percentile of this statistic), which appear to be measured with error. By comparison, in 2001 only 22.4 percent of the population was of primary school age 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.

are used to track and implement the program.[18] Road data are reported at the village level, or at the smaller habitation level that we aggregate 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.

Appendix Figure A1 shows how we define our main sample of villages. We restrict our sample to villages that did not have a paved road in 2001. We further limit the sample to villages that received new PMGSY roads between September 2003 and September 2015, and we exclude villages where roads were categorized as upgrades rather than as new roads. Our main sample includes 10,014 villages that received roads between 2003 and 2015, for which we have enrollment data for at least one pre-treatment year and one post-treatment year. We find similar results when we extend our sample to an unbalanced sample (n=19,152) or include villages that never received PMGSY roads (n=112,475).

For data on district-level rural wages and urban wages, we use data from all individuals reporting wages from the 55th round of the NSS Employment and Unemployment Survey (1999-2000). For data on village population and other characteristics, we use data from the Population Censuses of India in 1991, 2001, and 2011, and the 1998 Economic Census.

Table 1 shows summary statistics of villages at baseline. The enrollment drop-off 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.

TABLE 1 ABOUT HERE.

V Empirical Strategy

Our goal is to estimate the causal impact of roads on educational choices. Cross-sectional estimates of the relationship between village roads and schooling decisions are likely to be

[18] We obtained these data from the government's public reporting portal for PMGSY, hosted at http://omms.nic.in.

biased estimates of the impact of roads on schooling, because villages that do not have access to paved roads are different from more accessible villages along many dimensions. For example, villages without paved roads are likely to be smaller, have more difficult terrain, and be more politically marginalized. Our main empirical specification is a panel fixed effects regression that exploits the timing of road construction, within the set of all villages that received new roads by 2015 under the PMGSY program.

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 variable for 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.

The coefficient of interest, β, measures the impact of a new road on village-level outcomes (such as log school 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.

The identification assumption is that in the absence of the PMGSY, village-level outcomes would have followed the same path over time in villages that receive a paved road in different years, after partialling out the location and time fixed effects. The state-year fixed effects control flexibly for differential enrollment growth across states. This alleviates 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.

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 reports that a new paved road leads to a seven percent increase in middle school enrollment in a village (95% confidence interval: 4.1 – 9.9 percent). This effect corresponds to approximately three additional students in middle school, given the sample mean of 39 students enrolled in middle school.[19] In columns 2 and 3 of Table 2, we split the analysis by gender, and find similar effects on the enrollment of girls and boys. Columns 4 through 6 show comparable estimates using the level of middle school enrollment as the dependent variable, rather than log enrollment.

TABLE 2 ABOUT HERE.

To explore further the changes in school enrollment before and after a new road is built, we regress log middle school enrollment on a set of relative time dummies that indicate the number of years before or after road construction in the village. The estimating equation is:

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

where τ indicates the year relative to when a new road was built (i.e., $\tau = -1$ is the year

[19]This effect reflects a treatment period of 3.7 years after a road is built, on average. The estimate is 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.

before road construction). As in Equation 1, we include state-by-year fixed effects and village fixed effects.

In Figure 2, we plot the identified τ coefficients. We omit the relative time coefficients from the year before treatment and the first year available, following the suggestion of Borusyak and Jaravel (2017). The regression above can only be estimated with two relative time coefficients omitted because all villages in our sample are eventually treated.[20] We can therefore identify trend breaks, but cannot test either average trends or pre-trends. The F-test of the pre-treatment coefficients in Figure 2, which tests for non-linear pre-trends, is insignificant (p=0.94).

FIGURE 2 ABOUT HERE.

Figure 2 shows that increases in school enrollment correspond to the timing of new road construction, and these effects appear to be persistent. The timing and persistence of this change in enrollment suggests that the treatment effects are not driven by labor demand shocks from road construction itself, which would occur as the road is being built and disappear thereafter.

VI.B Robustness: Specifications and Sample Definitions

Table 3 shows that the estimated average effect on middle school enrollment is robust to a range of empirical specifications and sample definitions. In column 1, we allow for village-specific linear time trends to control for potential differential trends across villages that receive a paved road in different years.[21] In column 2, we control for interactions between year fixed effects and baseline village characteristics: population, share of irrigated land, number of schools, log middle school and primary school enrollment, literacy rate, popu-

[20] 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. 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 pre-trend.

[21] 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 robust to the inclusion of village time trends.

lation share of Scheduled Castes, and distance to nearest town. In column 3, we expand the sample to an unbalanced panel by including villages with missing data in one or more years, and column 4 shows the unbalanced panel estimates with village-specific 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 similar in magnitude and statistical significance. The stability of the treatment effect suggests that these estimates are not driven by different types of villages being treated at different times. Appendix Table A2 reports specifications from Table 2, with district-by-year fixed effects, and shows similar estimates.

TABLE 3 ABOUT HERE.

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, which gives us a non-parametric distribution of test statistics under the sharp null hypothesis. The placebo estimates are centered around zero and, consistent with Table 2, none of the thousand estimates attains our main estimate of the effect of a new road on log enrollment (0.07 increase in log enrollment).

VI.C Robustness: Regression Discontinuity

In this section, we present regression discontinuity estimates of the impact of new roads on schooling. Under PMGSY road construction 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, however, and even then, states followed the guidelines to different degrees because there were often several conflicting guidelines.[22] 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 treated sooner. Population above a guideline threshold is therefore an imperfect predictor of treatment status.

Figure 3 shows the relationship between the probability of receiving a new road by 2011 and the population relative to the treatment threshold. There is a clear discontinuity in treatment status at the population threshold. By contrast, 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.[23]

FIGURE 3 ABOUT HERE.

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

(3)

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

$Y_{i,s,t}$ is log enrollment in village i, region s, and time t; P is the population threshold; $pop_{i,s,2001}$ is baseline village population (i.e., the running variable); $X_{i,s,2001}$ is a vector of village controls measured at baseline; and η_s is a region fixed effect.[24] The change in the

[22] 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.

[23] 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 A3 and Figure A4 present regression discontinuity estimates and graphs showing that baseline village covariates do not vary systematically at the treatment threshold.

[24] 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).

outcome variable across the population threshold P 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 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, which shows an increase in enrollment above the treatment threshold.

FIGURE 4 ABOUT HERE.

In Table 4, panel A presents regression discontinuity estimates for the pooled 2011-2015 sample.[25] 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 impact on log middle school enrollment from crossing the population threshold. Column 3 presents the IV estimate, which yields a large but imprecisely estimated impact of road construction on middle school enrollment (p=0.103).

TABLE 4 ABOUT HERE.

The regression discontinuity 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

[25]Figure A5 shows regression discontinuity 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.

of roads on enrollment for regression discontinuity complier villages is substantially higher than for villages in the diff-in-diff sample, we note that the 95% confidence interval of the regression discontinuity estimate includes the panel estimate, and we are hesitant to put a large weight on the specific point estimate.[26]

Villages may otherwise differ across the population thresholds, and so as a placebo exercise we estimate Equation 3 for states that did not follow the PMGSY population threshold guidelines.[27] Panel B of Table 4 shows that there is no substantive first stage in these states (column 1) and a reduced-form treatment effect close to zero (column 2). This provides reassurance that villages above the population threshold would not have otherwise experienced differential changes in school enrollment.

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 also make the regression discontinuity estimates less representative of impacts across India. We therefore focus on the panel specifications in the section on treatment heterogeneity below.

VI.D Average Impacts on School Achievement

Increasing middle school enrollment may not directly translate into greater learning, especially if school quality is low or if there is increased school crowding. To measure student

[26]Indeed, in Section VII we find subgroup estimates approaching this level in villages where we expect treatment effects to be particularly large. In Appendix Table A4, we estimate panel specifications on samples of villages that are similar to the regression discontinuity sample. While some of these have larger point estimates, note that it is impossible to set the sample to the regression discontinuity complier villages, as we do not know which villages above the population threshold would be untreated at lower populations, and which villages below the threshold would be treated at higher populations. The regression discontinuity sample also includes villages that never received roads, whereas our main panel estimates use only villages that received roads at some point.

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

learning, we estimate impacts on student examination outcomes. 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 end-of-year exams, which were required to certify completion of middle school. Column 1 shows the estimated effect of roads on the log number of students who appear for the exam. Column 2 shows the effect on the log number of students who pass the exam, and column 3 shows the effect on the log number who score high marks.[28] For exam appearance and exam passing, we find similar effects to the enrollment effects: six percent more students take and pass exams in villages after the construction of a new paved road.[29] We find a positive but smaller three percent increase in the number of students scoring high marks.

TABLE 5 ABOUT HERE.

The estimated impacts on exam outcomes reflect the net impact on student 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 fewer high marks), 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 they are of lower ability than students not on the margin of dropping out), but students who would have stayed in school regardless of road construction are now performing better on exams. Non-marginal students could perform better, for example, if they begin to perceive higher returns to human capital accumulation due to increased access to labor markets outside the village. It is difficult to disentangle these two scenarios. Under both interpretations, we can reject the possibility that school enrollment is increasing without corresponding increases in academic achievement.

[28]Sample size is smaller for the exam estimates than for enrollment estimates because we were only able to obtain examination results for all states in our sample for the years 2004-2009. In each case, we report effects on the log number of students plus one. The estimates are similar for an unbalanced panel.

[29]The number of students achieving these exam outcomes is smaller than the enrollment effects because for every ten students enrolled in the 8th grade, only six appear for the exam, five pass the exam, and two pass the exam with distinction.

VI.E Impacts on Primary School Outcomes

In this section, we explore impacts on primary school enrollment and exam scores. In panel A of Table 6, columns 1 through 3 show the estimated difference-in-differences impact of road construction on log primary school enrollment. We do not find effects on primary school enrollment, consistent with our expectation that children under the age of twelve have few labor market opportunities. Columns 4 and 5 show reduced-form and IV results from the regression discontinuity estimation, which also do not indicate an impact on primary school enrollment. The precision of the regression discontinuity estimate is much lower, however, so we cannot reject meaningful regression discontinuity impacts in either direction. The corresponding regression discontinuity figure is in panel A of Appendix Figure A6.

TABLE 6 ABOUT HERE.

Panel B of Table 6 shows estimated impacts on primary school completion exam outcomes. We find weakly positive impacts on student exam outcomes with point estimates between 2 and 3 percent, but they are of marginal statistical significance. The results are suggestive of increased effort among enrolled children, which could be due to anticipated increases in attending middle school or anticipated increases in returns to education in the labor market.

VII Mechanisms

VII.A Human Capital Investment Incentives

In this section, we examine the mechanisms underlying the estimated average impact of new rural roads on human capital accumulation. Our analysis is guided by the conceptual framework outlined in Section II. We begin by focusing on three primary channels and sources of heterogeneity: a negative impact on human capital investment through increased opportunity costs of schooling, a positive impact on human capital investment through increased returns to education, and an impact on human capital investment through income or liquidity effects. We analyze the combined impact of income and liquidity effects because, given the available data, it is difficult to distinguish between the effects of higher lifetime income and greater cash-in-hand. Our estimation focuses on identifying subsets of villages where each mechanism is likely to be especially prominent.

To examine these mechanisms, we begin by assuming 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 low-skill wage gap between the village and surrounding market is high, the village low-skill wage will rise more than if the low-skill wage gap is small. We therefore expect the largest increases in the opportunity cost of schooling to occur in places with the largest gaps in low-skill wages between the village and its surrounding labor market. We proxy for 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. We use data on urban and rural wages 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 expected size of the returns to education effect, we again aim to identify the difference in returns to education between each village and its regional labor market. The underlying assumption is that a new rural road will shift the returns to education in a village toward the returns to education in the broader regional labor market. We calculate district-level returns to education by running Mincerian regressions at the district level, separately for individuals in rural and urban areas, using data from the 55th round of the NSS. We call this difference the urban-rural returns gap, or the skill premium gap.[30] We assume the returns to education effect is stronger when this skill premium gap is higher.

Finally, to proxy for the importance of income and liquidity effects, we assume that households with few assets are more likely to be liquidity constrained, and that a given change in wages for these households has a larger income effect. We measure average baseline assets at the village level using data from the 2002 Below Poverty Line Census. We define a village

[30]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 land owned or including state fixed effects. We exclude districts with no urban data.

as having low assets (and thus high potential income and liquidity effects) if the share of households reporting zero durable assets is above the sample median.[31] Similarly, we define binary indicator measures for the opportunity cost proxy and the returns to education proxy, based on whether each proxy is above the sample median.

We then estimate our previous panel regression, including additional interaction terms between the treatment indicator for road construction and the indicator variable for each of the three mechanisms. If the estimated impact of road construction varies with our proxy measure, and the interaction term is important in magnitude, it provides suggestive evidence of that mechanism being an important channel through which new roads affect schooling decisions.

Table 7 shows the results from estimating these interactions. Column 1 repeats the main specification, without interaction terms, in the sample for which each interaction term is measured.[32] Columns 2 through 4 include each interaction term separately, and column 5 includes the three interaction terms together.

TABLE 7 ABOUT HERE.

The estimated interaction effects are consistent with the predictions from a standard model of human capital investment. Road construction has the smallest effects on middle school enrollment in districts where these roads are expected to most raise the opportunity cost of schooling. The largest effects of road construction on middle school enrollment are in districts where road connections are expected to raise the skill premium the most and to have the largest income and liquidity effects. The opportunity cost effect interaction is strongly statistically significant ($p < 0.01$), the returns to education effect interaction is marginally statistically significant ($p = 0.08$), and the income/liquidity effect interaction is in the expected direction but statistically insignificant ($p = 0.37$).[33] The greater magnitude of the opportunity cost effect may be in part because the urban-rural wage gap is much larger than

[31]The surveyed assets are a radio, a television, a telephone, and a motorcycle.
[32]This analysis excludes districts without NSS data for both urban and rural areas in 1999-2000.
[33]For completeness, Appendix Table A5 shows results by quartile of each mechanism proxy.

the urban-rural skill premium gap (see Appendix Table A1).

While the estimated interaction effects are consistent with a standard model of human capital investment, note that there could be other district-level characteristics that influence the size of treatment effects and are correlated with the proxies we use. For example, high rural-urban wage gaps are correlated with greater remoteness, worse infrastructure, lower returns to education, and tend to be in the North. The estimated interaction effects are robust to the inclusion of interactions with these other variables, but there are myriad other unobserved district-level characteristics. Therefore, we see these estimates not as definitive but rather as suggestive indications of the mechanisms underlying the main estimates.

By fully interacting the three binary mechanism variables, we can obtain treatment effects in eight partitions of the sample based on the model's predictions. Table 8 shows 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 effect, low returns to education effect, and low income/liquidity effect, which represents 9% of all villages. This is precisely the group where a standard model predicts that roads would have the most adverse effects on education. Treatment effects are positive and significant only in the 39% of villages where at least two of the mechanisms are favorable.

TABLE 8 ABOUT HERE.

This treatment heterogeneity is consistent with the heterogeneity in results from earlier research 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 dominant mechanism in these studies is likely an increase in the return to education, since spoken English is a requirement for these call center jobs. Conversely, Shah and Steinberg (2017) find that children are more likely to attend school in drought years, when there are fewer agricultural jobs available. The more important mechanism in that setting is likely to be an opportunity cost effect, as the low-skill wage is declining when there are fewer agricultural jobs available (or less need for children to substitute into home production while parents work agricultural jobs). The small negative effects on schooling from India's workfare MGNREGA program (Islam and Sivasankaran, 2015; Das and Singh, 2013; Li and Sekhri, n.d.; Shah and Steinberg, 2015; Adukia, 2018) are likely driven by similar mechanisms, as MGNREGA hires workers for labor-intensive construction and increases labor demand for lower-skill workers. The same model helps to explain the variation in estimated impacts on schooling 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 jobs in Bangladesh (Heath and Mobarak, 2015)). In each case, individual schooling choices appear to respond to the skill requirements of the 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 our finding 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 Potential Mechanisms

In this section, we explore several other mechanisms through which road construction might impact schooling outcomes.

School Quality. We have focused on how road construction affects the incentives for human capital investment (*i.e.*, changes in the demand for schooling), though road construction could also affect school quality or the number of schools available (*i.e.*, changes in the supply of schooling). We use village-level DISE data to examine impacts of road construction on the number of schools and on measures of school quality, as proxied by physical characteristics of a school.

Appendix Table A6 reports no impact of road construction on the number of schools, and no systematic impact on school infrastructure characteristics.[34] While a minority of specifications show statistically significant effects on school infrastructure, none approach the size of the enrollment effects presented above. The standard errors in column 1 rule out a 2 percentage point change in the presence of any of these kinds of infrastructure. Overall, we do not find systematic evidence that road construction substantially affects the number of schools and their physical characteristics.

Other Government Programs. To alleviate 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. We observe village public goods only in the decennial population census, and so we are unable to estimate panel regressions. Instead, we use the regression discontinuity approach to test for discontinuities in village public goods around the PMGSY eligibility thresholds. Appendix Table A7 shows no discontinuity in the presence of schools, as above using DISE data, and also shows no discontinuity in village access to electric power, a primary health center, or a commercial bank. We can rule out a one percentage point increase in the existence of primary or secondary schools, health centers and banks, and a four percentage point increase in middle schools and electrification status.

Migration. We next explore whether the estimated increases in middle school enrollment could be driven by increased migration into villages that receive roads, or reduced outmigration from those villages. Note that we did not find impacts of road construction on primary school enrollment (Table 6), which would presumably also be affected if the increase in middle school enrollment was driven by migration responses to road construction.

To test for migration effects, we use the regression discontinuity specification to examine

[34]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 have access to a particular kind of infrastructure.

impacts of roads on village population in 2011.[35] The result is shown in panel A of Appendix Figure A6; there is no effect of the treatment threshold on population. The point estimate is close to zero, and the 95% confidence interval rejects the net entry or exit of more than four people from a treated village. Changing migration patterns are thus unlikely to explain the effects of roads on school enrollment.

Cross-Village Displacement. Relatedly, we explore whether our estimated impacts on school enrollment could be driven by displacement effects, in which increased enrollment in treated villages is counterbalanced by decreased enrollment in nearby villages. We calculate total middle school enrollment for all other villages within a 3 km or 5 km radius of each village that received a new road.[36] Using the panel specification, columns 1 and 2 of Appendix Table A8 report the estimated impact of roads on log middle school enrollment in these surrounding villages. We do not find impacts on school enrollment in these surrounding villages; the 95% confidence interval rules out a 4% decrease in enrollment in a 3 km radius of the village, and a 2% decrease in a 5 km radius.

School Accessibility. Finally, we examine the possibility that road construction increases school enrollment by decreasing the students' costs of traveling to school.[37] Children generally walk to village schools, though paved roads could make schools more accessible, especially during the rainy season. We explore this possibility by estimating whether the impact of roads varies across villages that are more or less dispersed geographically. We expect that impacts through increased school accessibility would be more pronounced for villages in which children have further to walk to school. We measure village dispersion using village surface area, and divide the sample into villages with above-median and below-median surface area per capita.[38] Columns 3 and 4 of Appendix Table A8 show that the estimated impact of roads

[35]As with public goods, village population is measured only in the decennial censuses, so we cannot use the panel approach here.

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

[37]For example, Muralidharan and Prakash (2017) find that the provision of bicycles made girls more likely to attend middle school and high school.

[38]We find similar results using village surface area.

is similar across more dispersed and more dense villages. The point estimates are similar and we cannot reject equality between them. This suggests that a decreased cost of reaching school is not a main channel through which road construction is impacting school enrollment.

Similarly, road construction may increase middle school enrollment in a village by increasing its accessibility to children from nearby villages that do not have a middle school. This is related to the potential for cross-village displacement, discussed above, but this effect would not decrease middle school enrollment in those nearby villages. To explore this mechanism, we calculate the total number of school-age children within a 5 km radius of sample villages, who were living in villages without middle schools.[39] Columns 5 and 6 of Appendix Table A8 show that estimated impacts of roads on middle school enrollment are similar across villages with more or fewer under-served children in nearby villages; the point estimate is slightly higher in villages that do not have many underserved children living nearby. This provides suggestive evidence that the schooling increases we observe are not originating from nearby villages without middle schools.

VIII Conclusion

High local transportation costs are a central feature of the lives of the very poor around the world, leaving them isolated from external markets. Connecting remote villages to high-quality transportation networks is a major goal of developing country governments and international development agencies. These roads can bring access to new economic opportunities; however, a concern is that increased access to low-skill labor market opportunities could decrease investment in the human capital that is central to long-run increases in living standards and broader economic growth.

We examine this trade-off in the context of India's flagship rural road construction program, which built local paved roads to 115,000 villages in India between 2001 and 2015. These roads connected villages with nearby labor markets, potentially changing the incentives for

[39] We proxy for the number of middle school-aged children using the number of children aged 0-6 in 2001, as reported in the Population Census. We find similar results if we use total village population in villages without middle schools.

investment in human capital. We find that rural road construction increased adolescent schooling outcomes. Further, we find that a standard model of human capital investment has important predictive power for how schooling decisions respond differently across villages to the increases in economic opportunity from rural road construction. We highlight the competing influences of opportunity cost effects, returns to education effects, and income/liquidity effects.

Our analysis draws on the substantial heterogeneity in economic opportunities across India, allowing us to identify large positive effects on schooling in places where the relative return to high-skill work increases the most, as well as neutral or weakly negative effects on schooling in places where the return to low-skill work increases the most. Notably, treatment effects are negative (and small) in only a small subset of villages. Across most of rural India, local market integration substantially promoted increased investment in human capital.

Our paper also highlights an important but understudied impact of rural infrastructure investment. Investments in road improvements are usually premised on their potential to bring economic growth to rural areas, with a focus on contemporaneous economic gains in those areas. If road construction leads to increased investment in human capital in rural areas, then the long-run economic impacts will be greater than short-run estimates suggest and will reflect human capital dividends over subsequent generations.

References

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

——, "Spillover Impacts on Education from Employment Guarantees," 2018. Becker Friedman Institute Working Paper No. 2018-33.

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

Asher, Sam and Paul Novosad, "Rural Roads and Local Economic Development," 2018.

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

——**, Azam Chaudhry, Shamyla Chaudry, Amit K. Khandelwal, and Eric Verhoogen**, "Markup and Cost Dispersion across Firms: Direct Evidence from Producer Surveys in Pakistan," *The American Economic Review Papers and Proceedings*, 2015, *105* (5), 537–44.

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

Banerjee, Abhijit and Rohini Somanathan, "The Political Economy of Public Goods: Some Evidence from India," *Journal of Development Economics*, 2007, *82* (2).

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*, 2018, *108* (4-5), 899–934.

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

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," *The 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," 2015. 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.

Li, Tianshi and Sheetal Sekhri, "The Unintended Consequences of Employment-Based Safety Net Programs," *World Bank Economic Review (forthcoming).*

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 Effects in the Additive Model," *Oxford Bulletin of Economics and Statistics*, 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), 709–734.

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

Muralidharan, K and N Prakash, "Cycling to School: Increasing Secondary School Enrollment for Girls in India," *American Economic Journal: Applied Economics*, 2017, *9* (3). NBER Working Paper No.19305.

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 village-level variables at baseline, in the sample of villages that were matched across all analysis data sets. 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. 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 four years before each road is built and four years after. Different years are thus included for different villages, but each village has nine observations. Due to data availability, the sample in column 6 only includes villages with roads built between 2006 and 2012. All specifications have state-year fixed effects and village fixed effects. 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. 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 high marks. All specifications have state-year fixed effects and village fixed effects. Standard errors are clustered at the village level.

Table 6
Impact of New Roads on Primary School Outcomes

Panel A: 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 regression discontinuity 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 eligibility threshold, and column 5 presents the regression discontinuity IV estimates of the impact of the new road. Standard errors are clustered at the village level.

Panel B: 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 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 regression discontinuity 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 eligibility threshold, and column 5 presents the regression discontinuity IV estimates of the impact of the new road. Standard errors are clustered at the village level.

Table 7

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 measures of different potential mechanisms. The size of the opportunity cost effect is proxied by the district-level mean low-skill urban wage minus the mean low-skill 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. 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 come from the 55th round of the NSS Employment and Unemployment Survey (1999-2000), and asset data are from the Below Poverty Line Census (2002). All specifications have state-year fixed effects and village fixed effects. Standard errors are clustered at the village level.

Table 8

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)	

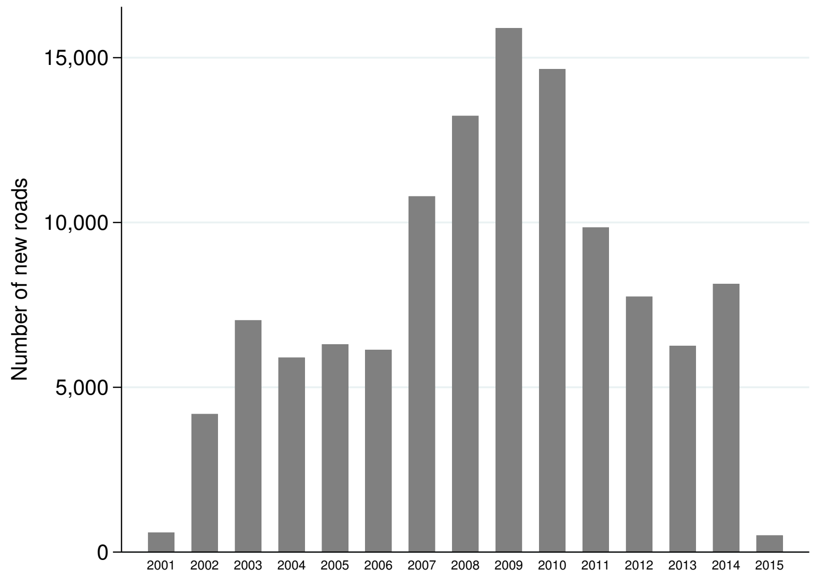
*p<0.10, **p<0.05, ***p<0.01

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 (as described in Table 7). 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. All specifications have state-year fixed effects and village fixed effects. Standard errors are clustered at the village level.

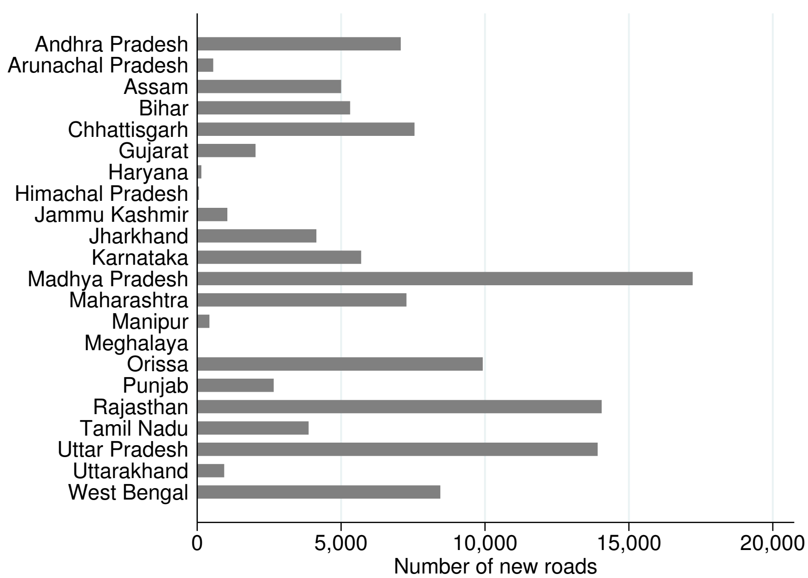
Figure 1

Sumamry of Road Construction under PMGSY

Panel A



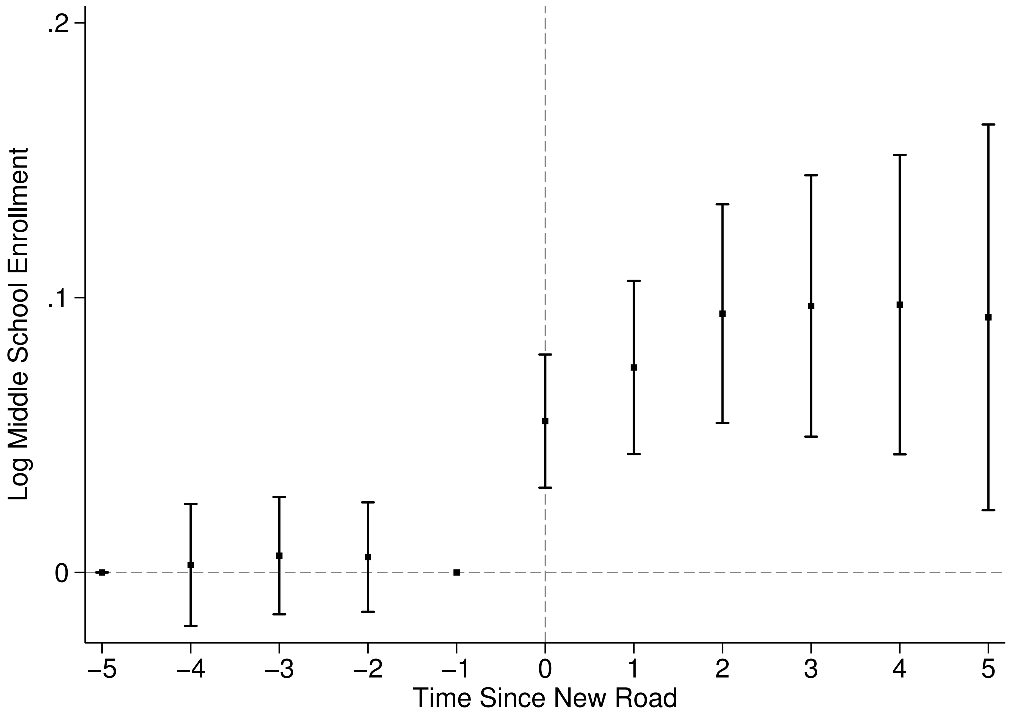
Panel B



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

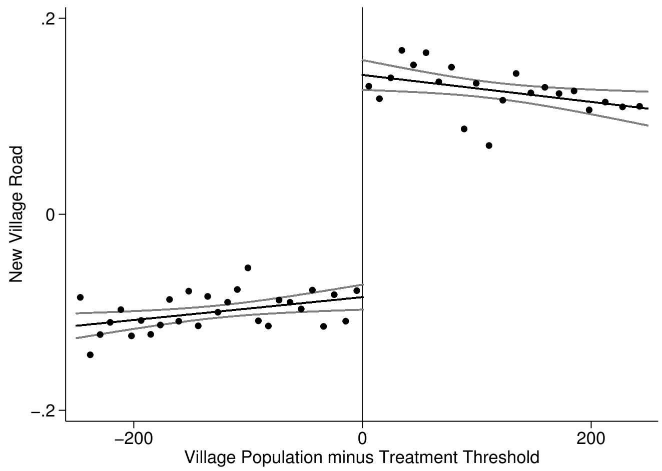
Impact of Roads on Middle School Enrollment:
Treatment Effect Time Series



The figure shows coefficient estimates from a panel regression of log middle school enrollment on a set of indicator 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. 95% confidence intervals are displayed around each point estimate. Standard errors are clustered at the village level.

Figure 3

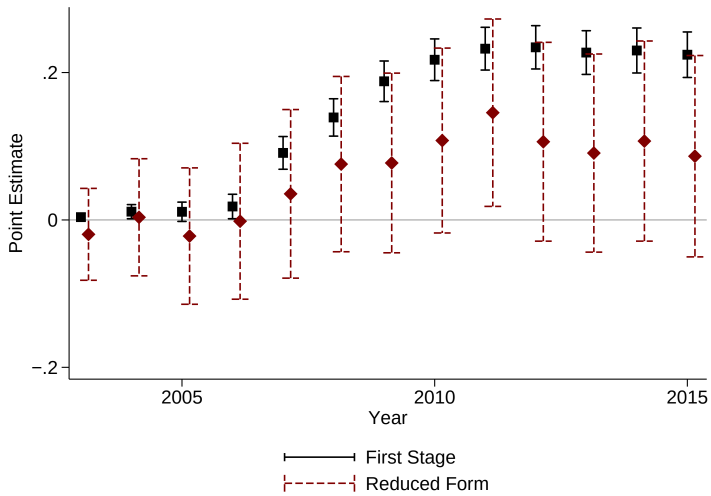
Regression Discontinuity First Stage:
Share of Villages Treated by Population



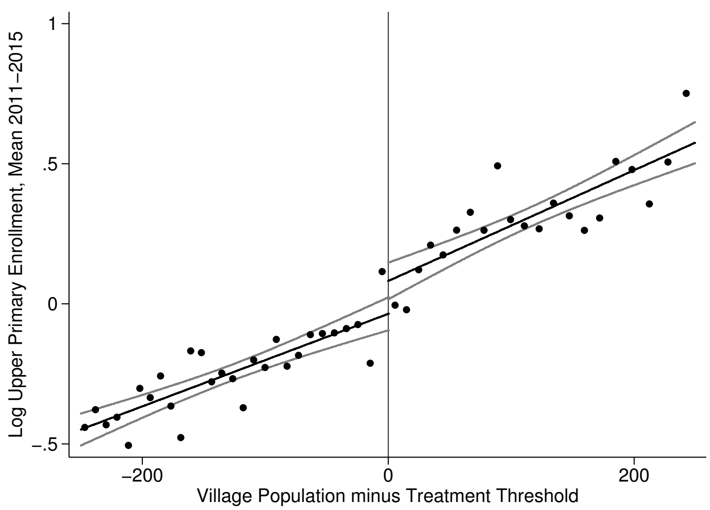
The graph plots the conditional expectation function of an indicator variable indicating that a village has received a road before 2011, conditioning on the village population as reported in the 2001 Population Census. Each point represents the mean of all villages in the given population bin (328 villages 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 regression discontinuity 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.

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

Anjali Adukia*

Sam Asher†

Paul Novosad‡

March 2019

*University of Chicago, 1307 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

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); within each group, the Mincerian return 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. Standard errors are clustered at the village level.

Table A3

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 A4

Panel Estimates in Regression Discontinuity Sample

Dependent Variable	Full Sample (1)	RD States (2)	RD Villages (3)	RD Villages with Untreated (4)
New Road	0.070***	0.082***	0.040**	0.165***
	(0.015)	(0.016)	(0.020)	(0.021)
N	146678	110740	71148	165606
r2	0.80	0.81	0.80	0.80

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

The table shows panel estimates of the effect of new road construction, focusing on samples that are more similar to the regression discontinuity analysis in Table 4. Column 1 repeats the main estimate from column 1 of Table 2. Column 2 limits the sample to the five states used in the regression discontinuity analysis. Column 3 limits the sample to the set of regression discontinuity villages with roads completed between 2003 and 2015. Note that this sample excludes the untreated regression discontinuity villages. The majority of villages in this sample were connected between 2007 and 2009, limiting the variation available for the difference-in-differences estimation. Column 4 limits the sample to the set of villages in the regression discontinuity sample, but (unlike Column 2 and unlike the other panel estimates) includes villages that never received roads. Thus, unlike the other panel estimates in the paper, this estimation compares treated villages to never-treated villages (as well as comparing pre- and post-treatment periods in treated villages).

Table A5

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 low-skill urban wage minus the mean low-skill 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. Standard errors are clustered at the village level.

Table A6

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 A7

Regression Discontinuity 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 difference in other public goods across the PMGSY population treatment threshold, using Equation 3. The dependent variable, column by column, is: (1) presence of primary school; (2) presence of middle school; (3) presence of secondary school; (4) village access to electric power; (5) presence of a primary health center; and (6) 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 A8

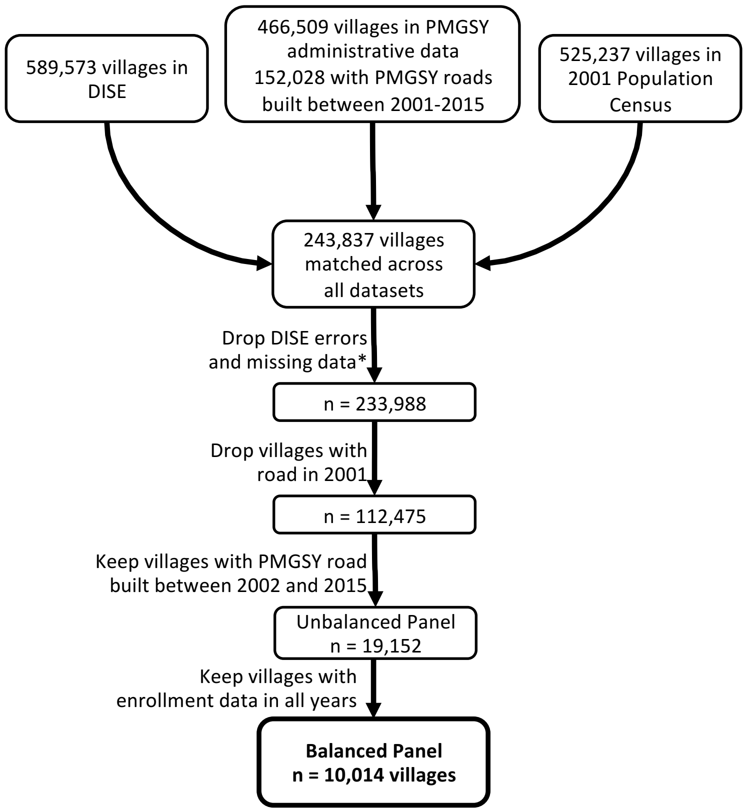
Impact of Roads on Middle School Enrollment:
Spatial Effects

	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, measured as those villages within a 3 km or 5 km radius, respectively. Columns 3 and 4 divide the sample into villages with above-median land area per capita and below-median land area per capita, and report effects separately. Columns 5 and 6 divide the sample into villages according to the number of children in nearby 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 under-served children. All specifications include state-year fixed effects and village fixed effects. Standard errors are clustered at the village level.

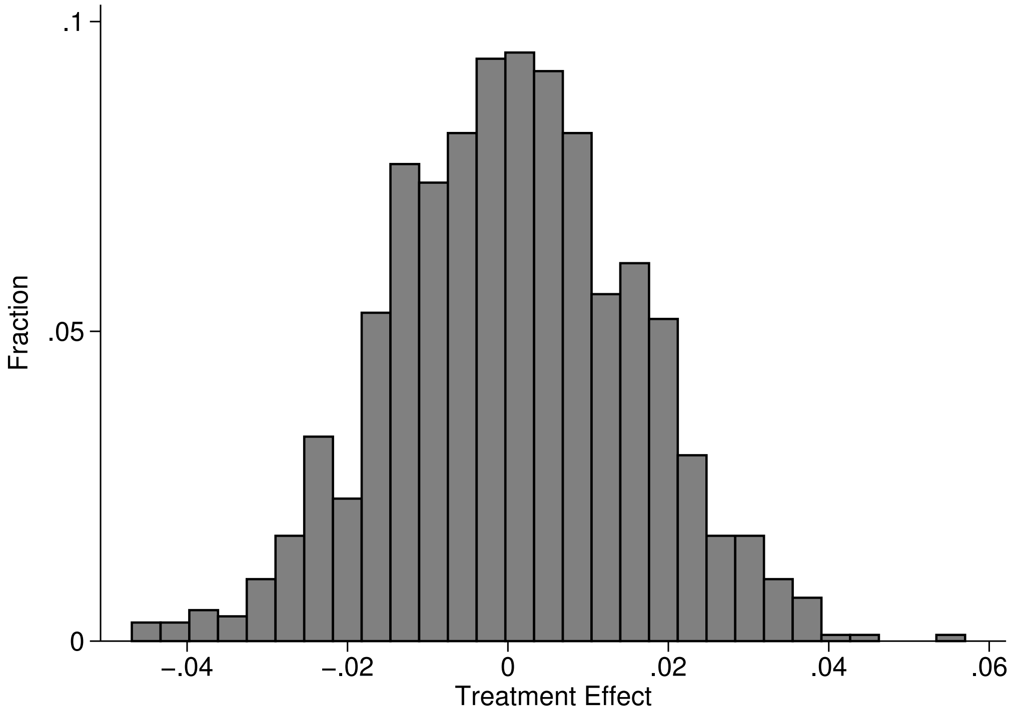
Figure A1
Sample Construction



The figure shows how the sample was constructed from the original datasets. DISE refers to the District Information System for Education. PMGSY refers to the Prime Minister's Village Road Program. Observation counts indicate the number of villages at each stage. * Observations were dropped if DISE reported enrollment for grades one through eight to be greater than 60% of the village population (or greater than the 99th percentile.

Figure A2

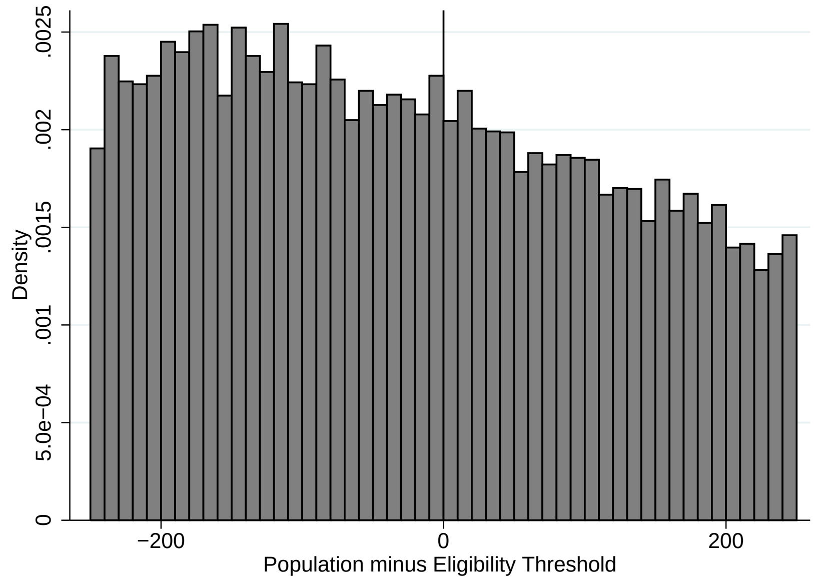
Panel Estimates of Effect of Roads on
Middle School Enrollment: Permutation Test

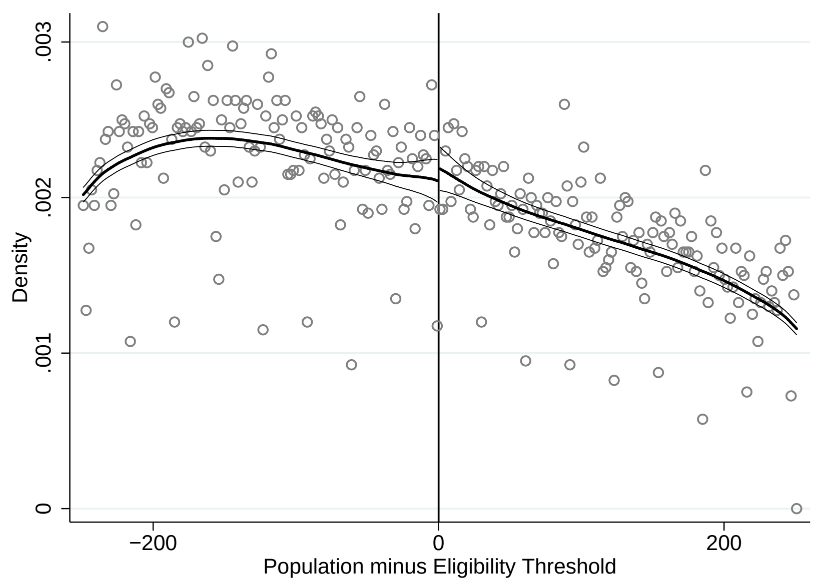


The figure shows the distribution of estimates from a placebo permutation test of the main panel specification presented in column 1 of Table 2. For each village in the main sample, we randomly generated a placebo year of road completion, and then estimated Equation 1. We ran this estimation 1000 times; 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

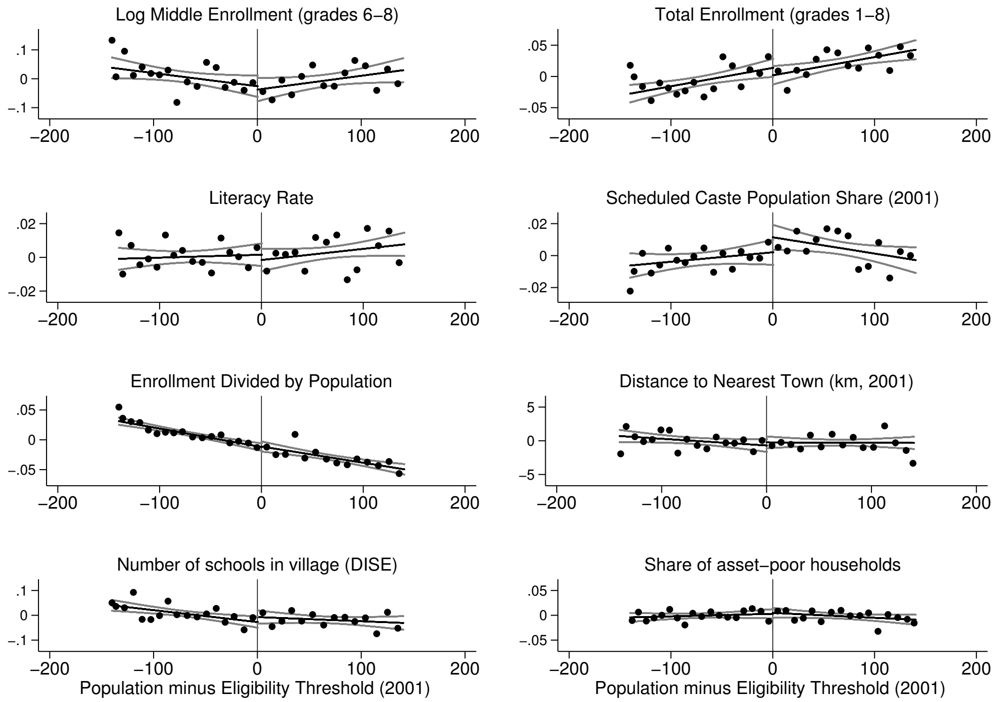
Regression Discontinuity: Continuity of Running Variable





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.

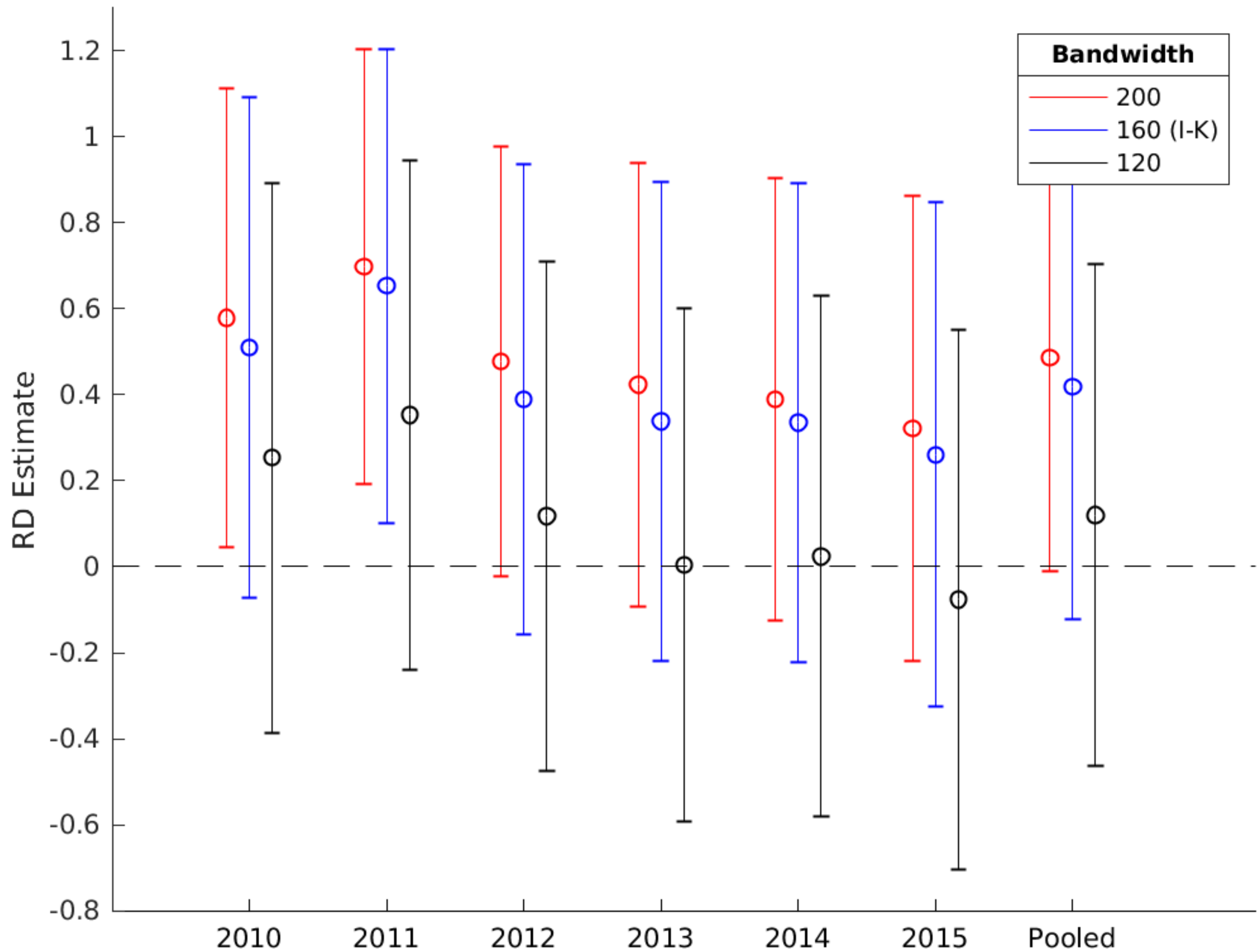
Figure A4
Regression Discontinuity: Continuity of Baseline Variables



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:
Regression Discontinuity Estimates by Year and Bandwidth

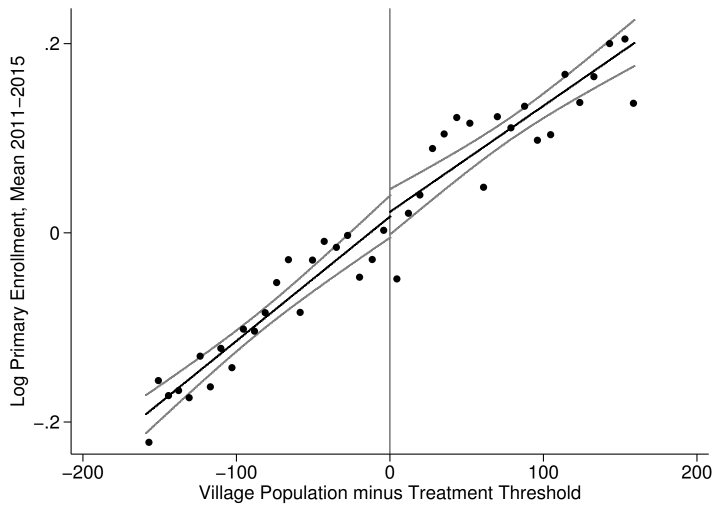


The figure shows IV estimates from Equation 3, estimated on different sample years, and at bandwidths 25% higher and lower than the optimal bandwidth of 160 selected with the algorithm of Imbens and Kalyanaraman (2012). Each point represents a single regression discontinuity estimate of the impact of 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.

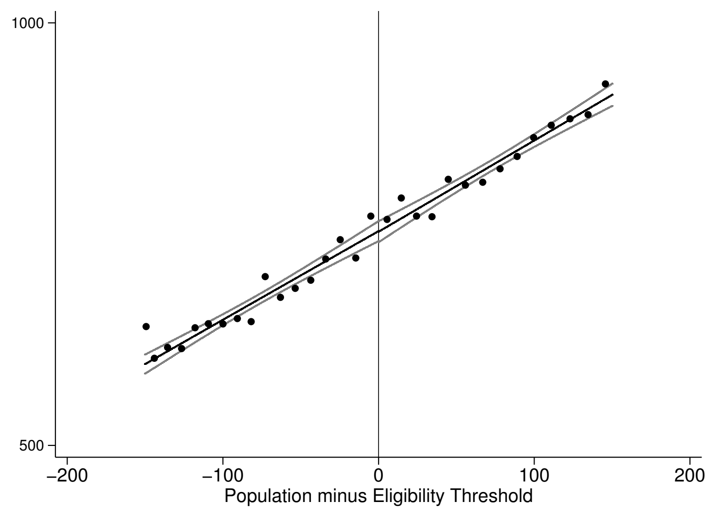
Figure A6

Regression Discontinuity Reduced Form:
Population and Primary School

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



Panel B: Log Population (2011)



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.