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

Anjali Adukia[†]

Sam Asher[‡]

Paul Novosad[§]

September 2016

Abstract

The rural poor in developing countries are increasingly gaining better access to domestic markets. The effects of this process on the schooling decisions of the poor is a central component of the long run impact of domestic market integration. But new economic opportunities have theoretically ambiguous impacts on human capital investment. We examine the impact of increased access to markets on educational choices, focusing on India's flagship road construction program, under which 115,000 new roads were built between 2001 and 2011. Using rural census data on school enrollment and administrative road program data, we find precise positive impacts of new roads on school enrollment. Children not only stay in school longer, but score higher and increasingly pass school completion exams. Treatment heterogeneity is broadly in line with a standard human capital investment model. Effect size is negatively correlated with predicted changes in opportunity cost of schooling, and positively correlated with predicted changes in returns to education. Effects are also largest in villages with few assets at baseline, suggesting important income or liquidity effects.

JEL Codes: I25; O18; J24.

*For helpful comments and guidance, we thank Liz Cascio, Eric Edmonds, Rick Hornbeck, Ofer Malamud, Doug Staiger, and participants of seminars at University of Chicago and CSWEP. We thank Srinivas Balasubramanian, Anwita Mahajan, Olga Namen, and Taewan Roh for excellent research assistance. We thank Arun Mehta and Aparna Mookerjee for help in data acquisition.

[†]University of Chicago, adukia@uchicago.edu

[‡]World Bank, sasher@worldbank.org

[§]Dartmouth College, paul.novosad@dartmouth.edu, corresponding author.

1 Introduction

Increased access to international markets can have important influences on schooling decisions, which are central to long run economic growth.[1] The integration of domestic markets is a parallel process, whose human capital impacts have not been as closely studied. But a large share of the world's poor are not well-connected to international markets, and it is new domestic market links that are of first order importance to them.[2] Critically, individuals face a tradeoff between long-run investment in human capital and immediate economic opportunities which might discourage increased schooling. Connections to new markets should lead to more education if increases in the return to education outweigh increases in the opportunity cost of schooling. However, lack of information or low liquidity may prevent poor households from making efficient investments in their children.

Our study examines the human capital investment response when a village gets connected to its surrounding towns and villages by a new paved road. The source of variation is the rollout of India's national rural road construction program, under which the government of India built high quality roads to over 115,000 villages across the country since 2001, connecting over 30 million households to new markets and opportunities. We focus on new rural feeder roads, which connect previously unconnected villages to the broader transportation network; they provide "last-mile" connections, and do not connect major population centers to each other. Ex ante, the impact of new road connections on schooling is theoretically ambiguous. If roads increase household earnings and education is a normal good, if roads ease a liquidity constraint, or if returns to education rise when a village's accessibility improves, then schooling should increase. Conversely, if integration with new markets increases the immediate return to household or wage work for school-aged individuals, then schooling

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

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

could decline.

The major barrier to estimating causal effects of new roads is the endogeneity of road placement. If roads are targeted either to growing or lagging regions, then comparisons of villages with and without roads will be biased. To overcome this bias, we estimate a panel regression of annual village-level enrollment, with village and state-time fixed effects, exploiting the year of road completion in each affected village. The state-time fixed effects control for any other time-variant state-level policies, as well as regional economic shocks and trends. The village fixed effects control for unobserved village-specific factors that may have influenced the time of treatment—for example, if certain types of villages were treated early and others were treated late. Essentially, we compare outcomes in villages before and after a road is built, flexibly controlling for time-variant regional shocks, as well as differences between early- and late-treated villages. Changes in fixed effect inclusion, including the addition of village-specific time trends, do not substantively affect the estimates, indicating that the exact year in which a program village received a road is not correlated with unobserved variables (Altonji et al., 2005).

Our village-level enrollment data is from India's national annual school enrollment census for the first through the eighth grades, the District Information System for Education (DISE, 2002-2011). Through a combination of human and machine-based fuzzy matching, we linked DISE data to administrative data from the national rural road construction program. The result is a panel of over 300,000 villages across all of India. The use of census data is essential to our analysis, since variation in the road program is at the village-level. It also allows us to precisely estimate program impacts even in small subsamples of the population. Our estimates are externally valid for all of India and particularly relevant for the places in the world that remain unreached by paved roads today.

We find that road construction significantly increases enrollment among middle-school children (defined in India as children in sixth through eighth grade). We estimate that

connecting a village with a new paved road causes a seven percent increase in middle school enrollment over the following two years.[3] The estimates are precise and highly statistically significant. We find similar effects on the number of students taking and scoring highly on middle-school completion exams, suggesting that human capital is improving in addition to school enrollment.[4] The results are robust to a range of variable definitions and specifications, as well as to a regression discontinuity specification (used in Asher and Novosad (2016), discussed below), which exploits a program rule that caused villages above specific village population thresholds to be targeted for treatment.

A standard human capital capital investment model suggests four possible channels through which the new economic opportunities brought by a road could influence schooling decisions: they could (i) raise the unskilled wage, and thus the opportunity cost of schooling; (ii) raise the return to education, and thus increase schooling; (iii) increase schooling through a household income effect; and (iv) increase schooling through a household liquidity effect. To test the importance of these channels, we identify villages where each of these effects is likely to be larger or smaller, based on equilibrium differences between urban and rural wage gaps, skill premia, and asset levels, and then examine how treatment varies across village type. The results are consistent with the standard model: schooling decisions are decreasing in opportunity costs, increasing in returns to education, and decreasing in asset levels, the last of which is a proxy for either an income or a liquidity effect. Market integration leads to statistically detectable *decreases* in schooling in only a small subset of villages. We present evidence that rules out several other channels for these effects, including (i) migration effects; (ii) increasing school quality or school infrastructure; (iii) displacement effects, where

[3]We are not able to estimate the impact on enrollment rates, because our data contains gross enrollment statistics, and we do not have village-level population disaggregated by age.

[4]In many cases, interventions that improve attendance and enrollment do not improve student test scores (*e.g.* Miguel et al. (2004), Behrman et al. (2008), Adukia (2016)), perhaps due to congestion. Congestion effects in our study may be counterbalanced by the fact that already-enrolled children may be working harder.

enrollment changes are driven by enrollment declines in nearby villages; and (iv) easier access to schools for students living on the outskirts of villages. Consistent with earlier literature, we find no effects on primary school children; there is less margin for improvement in this group, and younger children are less likely to be working as well.

Our findings suggest that human capital investment among the rural poor is constrained by limited integration with nearby markets. Despite the low quality of schools in rural India (see ASER Centre (2014) for a summary), the economic opportunities that come with new roads drive increased enrollment, suggesting even existing schools have positive value. Our result provides context for the strong correlation around the world between education, growth and trade; in India at least, improvements in access to markets at the margin motivate increases in educational attainment.

This study is related to a growing literature on the impact of exogenous economic shocks on schooling decisions; these studies find both positive and negative effects of new economic opportunities.[5] This ambiguity makes sense in the context of the human capital investment model: for any sectoral shock, the opportunity cost effect could dominate, or the combination of income, liquidity and returns to education effects could dominate. Our work is also related to studies of the impact of India's national public works program (the National Rural Employment Guarantee Scheme, or NREGS) on human capital accumulation.[6] A key difference in context is that NREGS strictly increases demand for unskilled labor, whereas new roads increase access to both high and low skilled opportunities; thus NREGS is unlikely to increase the return to education, though it could have liquidity or income effects.

[5]The opening of new outsourcing facilities in India, cut flower plants in Colombia, and garment factories and Bangladesh have driven increases in schooling (Jensen, 2012; Oster and Millett, 2013; Hernández, 2015; Heath et al., 2015). Positive agricultural shocks in India, and new opportunities in natural gas fracking in the United States and auto manufacturing in Mexico, have increased dropout rates, especially of middle school and older children (Shah and Steinberg, 2013; Atkin, 2016; Cascio and Narayan, 2015).

[6]Studies from Andhra Pradesh find that access to the workfare program increases children's enrollment (Afridi et al., 2013) and test scores (Mani et al., 2014). All-India studies consistently find increased enrollment for primary school aged children, but decreased enrollment for middle and high school children (Islam and Sivasankaran, 2014; Das and Singh, 2013; Li, Tianshi and Sekhri, 2015; Shah and Steinberg, 2015).

All of these papers, like ours, are concerned with exogenous shocks to local labor markets; but none of them are directly informative about the effect of improving village access to already-existing nearby markets. The latter is particularly relevant, as the degree of market integration between villages and their neighborhood is a direct consequence of infrastructure investment policy, whereas economic shocks are not entirely under the control of decision-makers, even if they can be influenced by policy.

Our paper also contributes to the literature on the development impacts of rural road construction.[7] Asher and Novosad (2016) estimate the economic impacts of the same road building program as this paper using the regression discontinuity approach described above, and find that roads lead to reallocation of village labor from agricultural to wage work, as well as increases in household income. In this paper, we focus on panel estimates rather than regression discontinuity estimates, because they are more precise and cover a wider range of villages.[8] The panel approach is not available in Asher and Novosad (2016), because unlike school enrollment data, village-level economic outcomes are not available on an annual basis. This literature supports the key assumption underlying the model, that rural roads lead to new wage-earning opportunities and increased income for residents of newly-connected villages. Surprisingly few of these studies examine impacts on schooling, perhaps because roads are predominantly understood by policy-makers as income-generating investments. Khandker et al. (2009), Khandker and Koolwal (2011) and Aggarwal (2015) find positive associations between new roads and human capital investments, comparing treated to untreated villages and districts. To our knowledge, our paper is the first to generate precise causal estimates of the impact of roads on schooling, using identification strategies not

[7]Some examples include Jacoby (2000); Jacoby and Minten (2009); Donaldson (2012); Gibson and Olivia (2010); Mu and van de Walle (2011); Donaldson and Hornbeck (2015); Casaburi et al. (2013); Shrestha (2015). We focus here on previous research on rural feeder roads, which connect remote villages to the transportation network. For a more detailed review, including studies on the impacts of highways and regional roads, see Asher and Novosad (2016) and Hine et al. (2016).

[8]The regression discontinuity requires discarding villages with population far from treatment thresholds, as well as states that did not implement according to program rules.

available in these other settings. Our paper thus highlights an important but understudied benefit of rural road construction.

Section 2 provides background on road construction and education in India. Section 3 presents a model describing human capital investment decisions and the role of accessibility. We describe the data in Section 4, and the empirical strategies in Section 5. Section 6 presents results and Section 7 concludes.

2 Background and Details of the Road Construction Program

The study period 2002-2011 represented a period of marked growth and substantial education reform in India. Several programs were put into place with the explicit goal of increasing school participation, including a national drive supporting the goal of universal primary education under the flagship program *Sarve Shiksha Abhiyan*. During this period, school enrollment significantly increased and the percentage of school-aged children who were out-of-school declined substantially, parallel to a similar global trend. From 2002 to 2011, the middle-school enrollment rate increased from 63.4 percent to 85.2 percent, while the primary enrollment rate increased from 95 percent to 110.5 percent (UNESCO, 2016).

Both educational attainment and economic growth vary substantially across India. Rural areas are meaningfully poorer and less educated than urban areas, and virtually all economic outcomes are monotonically deteriorating with distance from urban centers (Asher et al., 2016). Since the 1990s, Indian policy-makers have sought to allocate public goods with an aim to mitigate spatial inequality, but vast disparities remain and are at the center of public debate in India (Banerjee et al., 2007; Dreze and Sen, 2013). The high cost and poor durability of roads have constrained the ability of the government to connect every village: in 2001, 33 percent of Indian villages remained inaccessible by all-season roads. These villages

were characterized by greater poverty and lower educational attainment, and paid a greater economic penalty with distance from cities.

In 2000, the Government of India launched the Pradhan Mantri Gram Sadak Yojana (Prime Minister's Road Construction Program, or PMGSY), a national program with the goal of eventually building a paved road to every village in India. The program rules stipulated that unconnected habitations with high populations should be prioritized: those with populations above 1000 were to receive highest priority, followed by those with populations above 500.[9] Eligibility lists were prepared by district-level officials and reviewed at the state level. At the outset, about 170,000 habitations in approximately 80,000 villages were eligible for the program, a number that has grown as the guidelines have been expanded to include smaller habitations. By 2013, over 100,000 villages had access roads built or upgraded under the program. Construction projects were most often managed through subcontracts with larger firms, and were capital-intensive due to the high quality of road. These roads are distinct from new roads being built under the National Rural Employment Guarantee Scheme, which are less durable roads that are built with very labor intensive methods. Figure 1 shows the number of roads completed by year between 2001 and 2011. The median road length was 4.4km.

[9] A habitation is an administrative unit smaller than a village; under certain circumstances, proximate habitations could pool their populations to exceed this cutoff. We focus on villages as the unit of analysis, because (i) many villages have only one habitation; (ii) many habitations were pooled to the village level for the purposes of the program; and (iii) very little economic data is available at the habitation level.

3 Conceptual Framework: Schooling Decisions and Economic Opportunity

Our goal is to understand how individuals' choices over human capital investments change when their outside opportunities change.[10] An individual (or household) trades off the long-run benefits of human capital accumulation against the short-run returns to labor activity; the parameters of this tradeoff change when a village becomes connected to external labor and goods markets. The most useful framework for our analysis is a standard model of human capital investment with credit constraints; see Ranjan (1999) or Baland and Robinson (2000) for examples applied to child labor.

A two period model is sufficient to highlight the essential comparative statics. In the first period, an agent chooses between wage work and schooling; schooling leads to a higher wage in the second period. Consumption across both periods comes out of an initial endowment plus the wages earned in each period; agents can save, but may not be able to borrow. The endowment can reflect household wealth, or the wages of household adults, who have completed their schooling. If credit constrained, agents with small endowments may therefore forgo schooling even if the return to education is high. Finally, education may be a normal good, which households value independently of its impact on future wages.

A new road to a village can change the following parameters in the model: (i) the low skill wage; (ii) the high skill wage relative to the low skill wage; (iii) the endowment (via adult wages). Based on the existing literature on roads, as well as the cross-sectional differences between connected and unconnected villages, we expect all these parameters to increase in value when a village is serviced by a new road.[11] A new road is also likely to affect local

[10] Asher and Novosad (2016) present evidence that access to nearby labor markets causes individuals to shift out of agricultural cultivation into wage work, often in nearby towns, and that household income increases.

[11] In the cross-section of the Socioeconomic and Caste Census (2012), rural connected villages have higher wages, higher returns to education, and higher average household income than villages without roads. The

prices; price changes will affect schooling decisions through their impacts on the returns to different kinds of work, hence are captured in the parameters above.[12] An increase in the low skill wage raises the opportunity cost of schooling, and may motivate agents to reduce human capital investments. An increase in the high skill wage raises the return to education, and motivates increase human capital accumulation. An increase in adult wages or in the endowment can raise schooling levels through two mechanisms. First, it can increase lifetime income, and thus increase the quantity of education consumed. Second, in the presence of credit constraints, an increased endowment eases the liquidity constraint in the first period, which could lead to increases in schooling. There are thus four possible mechanisms through which schooling is likely to respond to a new road: (i) an opportunity cost effect; (ii) a returns to education effect; (iii) an income effect; and (iv) a liquidity effect. The first of these is negative; the remaining three are positive.

Which of these factors dominates is ultimately an empirical question, and an important one. While access to markets is generally considered to be beneficial to individuals, it may cause short-sighted individuals to make suboptimal educational investment decisions. For example, Atkin (2016) finds that individuals who were induced to drop out of school in response to new export market job opportunities ultimately end up with lower wages than their counterparts who stayed in school.

Our empirical approach is to first estimate the aggregate treatment effect of roads on all previously unconnected villages; we show that across all of India, the combined positive return, income and liquidity effects dominate the negative opportunity cost effect on schooling. Then, we identify characteristics of villages and districts where each of these individual effects are likely to be larger or smaller, to better understand the full distribution of treatment effects.

same is true when comparing towns to villages in the 68th National Sample Survey (2011-2012).

[12]Specifically, a road is likely to lower the prices of durable goods, and to increase the prices received for agricultural commodities. Both of these increase effective rural wages.

4 Data

We constructed a village panel, combining data on road construction with village characteristics and educational outcomes. We matched three successive Indian Population Censuses (1991, 2001, 2011) to an annual census of Indian schools, the District Information System for Education (DISE, 2002-2011), as well as the administrative data from the implementation of road program (2001-2014). All data were merged primarily through fuzzy matching of location names, though in some cases unique identifiers were available for subsets of the match.[13,14].

The DISE dataset is an annual census of primary and middle schools in India. It contains, among other variables, data on student enrollment, disaggregated by student grade, sex and age; teacher characteristics; and school infrastructure. This dataset was created by the Ministry of Human Resource Development of the Government of India and is administered by The National University of Educational Planning and Administration. DISE data are considered to comprehensively cover every registered Indian government primary and middle school beginning in 2005. We also have DISE data for a more limited sample of schools from 2002-2004, a period when they had established systematic data-collection mechanisms but were still rolling out the data-collection system to all districts. We refer to academic years (which begin in June or July) according to the beginning of the school year; i.e. we refer to academic year 2007-08 as 2007. We use the data that were available at the time of writing, which cover 2002-2011. Using DISE, we were able to replicate national survey-based statistics on enrollment, suggesting that the data are reliable.

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

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

Our primary outcome variable is log total school enrollment, which we define as the sum of enrollment of all schools in a village. As with the previous literature, we focus on middle school (grades 6-8), because there is little variation in dropout rates for younger children, and they have fewer labor market opportunities. Further, the transition to middle school is a natural breakpoint in a child's schooling at which educational milestones are often measured. While DISE does not report enrollment information for higher grades, middle school exam completion and scores are likely to be proxies for continuation to high school. DISE does not provide the total number of school-age children in a village, so it is not possible to calculate an enrollment rate. Some of the enrollment numbers reported by DISE are erroneous; we dropped all villages which reported total enrollment (first through eighth grades) greater than 50 percent of total population, which was the 95th percentile of this statistic.[15]

In addition to enrollment data, DISE collects information on examination outcomes in the set of states where public examinations are held in the terminal year of middle school, which is either the seventh or the eighth grade. These are used for promotion decisions and completion verification. The information collected includes the number of students that appeared for the exam, that passed the exam, and that scored with distinction. Examination data are available for years 2004-2009. Finally, we use DISE data on school infrastructure, which describes the presence of blackboards, electricity, sanitation facilities, water (by source), a playground, a library, a boundary wall, access to regular medical checkups, and access ramps.

For data on road construction, we use the administrative records which are used to track and implement the PMGSY program, which we scraped from the Online Management and Monitoring System, the government's public reporting portal for this program.[16] Road data are reported at either the village or habitation level; as discussed in Section 2, we aggregate

[15] In 2001, 11.3 percent of the population was of primary- or middle-school age (ages 6-15); population data from the BPL Census (described below) suggests that fewer than 37.5 percent of village residents are between 6 and 15 years of age in 99 percent of villages. Results are not materially affected by this threshold.

[16] At the time of writing, this was hosted at http://omms.nic.in.

PMGSY data to the village level. We defined a village as having a paved road at baseline if any habitation in that village had a paved road, and we define a village as receiving a new road if any habitation in the village received a new road. We define a village as treated by a road if the road completion date is before September 30 of a given school year, which is the date on which DISE records enrollment numbers. We restrict our sample to villages that did not have a paved road in 2001, and we discard village where PMGSY roads were categorized as upgrades rather than as new roads, because we do not know the baseline condition of the road.[17] We further limit the sample to villages that received new program roads between 2003 and 2010, so that we have at least one pre- and post-treatment year for each village. Figure 2 shows how we arrive at our final sample of villages.

For additional baseline village characteristics, we used data from the 2002 Below Poverty Line census, which is a census of rural assets, and includes information on household assets, and individual education and occupation. These internal administrative data were shared with us by the Ministry of Rural Development. Some of these data were corrupted, but we were ultimately able to match household BPL data to 80 percent of our sample villages. To calculate district-level rural and urban wages, we used the 68th round of the NSS Employment and Unemployment Survey, undertaken in 2011-12. Finally, we use data from the Population Census of India in 1991, 2001, and 2011, which include village population and other demographic data.

Table 1 shows summary statistics of villages at baseline. In the cross-section, villages without paved roads are smaller; relative to their population, they have lower non-agricultural employment, lower middle school enrollment, considerably lower middle school exam completion and performance, but higher primary school enrollment. The gross enroll-

[17]The "paved" status of a road is in fact a subjective measure, which is why we use both of these restrictions. When comparing the 2001 population census measure of a village's paved road status and the PMGSY's internal status report (based on an engineer's perception), the correlation coefficient is only 0.42. Our results are not materially affected by these choices.

ment numbers may be confounded by different demographic structure in villages without paved roads, because we do not have village-level data on population by age.

5 Empirical strategy

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

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:

$$Y_{i,s,t} = \beta_0 + \beta_1 ROAD_{i,s,t} + \boldsymbol{\gamma}_{s,t} + \boldsymbol{\eta}_i + \epsilon_{i,s,t}. \tag{1}$$

$Y_{i,s,t}$ is the outcome variable (log enrollment in the main specification), in village i and state s at time t. $ROAD_{i,s,t}$ is an indicator of whether the village has been connected by a paved road under PMGSY by time t. $\boldsymbol{\gamma}_{s,t}$ is a state-time fixed effect, and $\boldsymbol{\eta}_i$ is a village fixed effect. The error term, $\epsilon_{i,t}$, is clustered at the village level, to account for spatial and serial correlation in the dependent variable. β_1 is the coefficient of interest, and indicates the effect of the new road on the annual log change in village-level enrollment.

Our primary sample includes only villages for which $ROAD_{i,2002} = 0$ and $ROAD_{i,2011} = 1$, i.e. villages that received a road at some point under the program. We thus avoid any comparison between villages with and without roads, or villages with roads built through

programs other than PMGSY.[18]

The state-time fixed effects control for differential enrollment growth across states at any point in time. This controls for the possibility that states with effective governments build 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 effect takes into account the fact that villages that receive roads early in the program may be systematically different from those that receive roads late in the program. We also present specifications that control for village time trends.

As an additional robustness test, we conduct additional tests using a regression discontinuity specification (Lee and Lemieux, 2010). Under the program rules, states were instructed to first target villages with populations greater than 1000 in the population census, and then villages with population greater than 500. By comparing outcomes in villages just above the population threshold to villages just below the threshold, we can obtain causal program estimates under minimal assumptions, which we test in Section 6.3.1. The top panel of Figure 4 graphs the share of unconnected villages that received new roads under PMGSY before 2011 against their population. The treatment effects at the two population thresholds are highly visible.

The idea underlying the regression discontinuity approach is that the population cutoffs create variation in access to roads that is as good as random. For villages with populations very close to one of the program thresholds, there is unlikely to be a large underlying difference between villages just above the population cutoff and those just below it. Even if unobservable village-level factors influence the placement of roads in aggregate, they are unlikely to be correlated with whether a village is just above or just below this threshold. A comparison of outcomes in villages just-above and just-below can thus yield an unbiased

[18] We include a specification with the inclusion of all DISE-matched villages in our sample; the village fixed effects prevent these never-treated villages from influencing the coefficient β_1 directly, but they may help in estimating control variables more precisely.

program treatment effect. Asher and Novosad (2016) show that the program rules meet the assumptions for validity of the RD strategy. We replicate these tests in our sample data below.

We use the following implementation of a local linear estimator:

$$Y_{i,s} = \gamma_0 + \gamma_1 1\{pop_{i,s} \geq T\} + \gamma_2 pop_{i,s} + \gamma_3 pop_{i,s} * 1\{pop_{i,s} \geq T\} + \boldsymbol{\zeta} X_{i,s} + \eta_s + \upsilon_{i,s}. \quad (2)$$

$Y_{i,s}$ is the outcome in village i and geographic region s, T is the population threshold, $pop_{i,s}$ is baseline village population (the running variable), $X_{i,s}$ is a vector of village controls measured at baseline, and η_s is a region fixed effect. Village controls and region fixed effects are not necessary for identification but improve the efficiency of the estimation. The change in the outcome variable across the population threshold T is captured by $\gamma_1 + \gamma_3 * T$. For ease of exposition, we subtract the threshold value T from the population variable, such that $T = 0$, and γ_1 fully describes the change in outcome $Y_{i,s}$ at the treatment threshold. The population controls allow for different slopes on either side of the treatment threshold, and we test for robustness to a range of bandwidths.

The strength of the regression discontinuity approach is that it exploits variation in road treatment that is as good as random, generating unbiased treatment estimates under a minimal set of assumptions. The weakness, with respect to the panel estimates, is that the local average treatment effects are relevant only to the subset of villages within the treatment bandwidth, in the states that followed program rules. This weakens external validity, and also necessitates discarding more than half of the villages in the sample; treatment estimates therefore have larger standard errors than the panel estimates, leaving less scope for examination of treatment heterogeneity. For these reasons, our results focus on the panel estimates, and we rely on the regression discontinuity estimates primarily for robustness.

6 Results

6.1 Panel Estimates

Table 2 shows panel estimates of the impact of building a new road to a previously unconnected village, using Equation 3. The dependent variable is log enrollment in middle school, or grades six through eight. All estimates include village and state-year fixed effects. The coefficient on *New Road* thus represents the difference in log average enrollment between years when a village has a road and the years when it does not, after controlling for trends at the state-year level. Column 1 shows the balanced panel estimate from the 11,905 villages in our sample that were unconnected at baseline and received a road under PMGSY between 2003 and 2011. The estimate implies that a new road leads to a 7 percent increase in middle school in enrollment. The estimate is highly statistically significant, with a p-value less than 0.001. Given the sample mean of 41 students enrolled in middle school, this corresponds to three additional students in middle school, an average of two years after a road is built.[19] The overall middle school annual enrollment growth rate in the sample is 5.2 percent.

This finding is robust under a range of sample specifications. In Column 2, we add village-specific linear time trends, which leave the estimates unchanged.[20] In Column 3, we restrict the data to years after 2004, the years when the DISE data has the highest coverage of villages and schools. In Column 4, we expand to an unbalanced sample which includes villages with missing data in one or more years. Column 5 restricts the sample to a set of villages for which we have three observations before and three observations after the construction of the road; the sample is limited to those observations, hence seven observations per village. Columns 2 and 5 confirm that treatment effects are not driven by different enrollment trends

[19] Most villages are observed several times after being treated. The estimate is thus a weighted difference of enrollment in all treated years and enrollment in untreated years. The average number of treated years is two.

[20] 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).

in villages that received roads early or late in the sample period.[21]

We run a randomization test to verify that our p-values are estimated correctly. For each village, we randomly generate a placebo year of road completion, and then estimate Equation 3. We run this estimation 1000 times; Appendix Figure A1 shows the distribution of β_1, the placebo impacts of a new road on log middle school enrollment growth. As expected, the placebo estimates are centered around zero, and not a single one of the thousand estimates attains our primary estimate of the effect of a new road on log enrollment, of 0.07. This result is consistent with our finding of a p-value less than 0.001 on our main estimate.

Figure 3 shows individual coefficients from a regression of log enrollment on a set of relative time dummies, which indicate the number of years before or after treatment of a given observation. The estimating equation is:

$$Y_{i,s,t} = \beta_0 + \sum_{\tau \in (-4,+4), \tau \neq -1} \boldsymbol{\zeta}_\tau + \boldsymbol{\gamma}_{s,t} + \boldsymbol{\eta}_i + \epsilon_{i,s,t}, \tag{3}$$

where τ indicates the year *relative to when a new road was built*, i.e. $\tau = -1$ is the year before road construction. State-time and village fixed effects are included as above, and the year before the road is built ($\tau = -1$) is omitted. We plot the τ coefficients. The graph confirms that the enrollment increase corresponds to the timing of the construction of the new road.[22]

Increasing school enrollment may not imply increasing human capital, especially if school quality is low, or there are congestion effects. We turn to exam scores as a measure of what students are actually learning. Table 3 presents panel estimates of the impact of new roads on a set of dependent variables describing students' exam-taking decisions and exam

[21] Appendix Table A1 shows results with the inclusion of a lagged dependent variable, as well as results with a level dependent variable rather than a log. The estimates are all substantively similar; the point estimate in the specification with a lag term is smaller because it should be interpreted as an annual log growth effect rather than the average of approximately two years of growth.

[22] Given the standard errors on the estimates in the individual years after road construction, we do not make inferences about whether the impact is gradual or immediate.

performance. We focus on middle school completion exams, which are required if students are to go onto high school. Column 1 estimates the effect of roads on the log number of students who appear for completion exams. Column 2 estimates effects on the number of students who pass the exam, and Column 3 shows effects on the number who pass with distinction.[23] For exam appearance and passing, we find similar effects to the enrollment effects: 6 percent more students take and pass exams in villages after new feeder roads have been built. We find a positive but smaller 3 percent effect on those passing with distinction. While the percentage effects are similar, the number of students achieving these outcomes is smaller than the enrollment effects, because for every ten students enrolled in the 7th class, only six appear for the exam, five pass, and two pass with distinction. These results indicate that students are not only staying in school longer, they appear to be learning more and performing better. This is not surprising: students are staying in school because they perceive the human capital accumulation as valuable.

6.2 Mechanisms

In this section, we examine the mechanisms for an effect of new economic opportunities on human capital accumulation, as described in Section 3. The potential channels are: (i) a negative opportunity cost effect; (ii) a positive returns to education effect; (iii) a positive income effect; and (iv) a positive liquidity effect. Our approach is to study how the treatment effect varies in villages that are likely to be influenced by one of these channels more or less than others. We therefore seek proxies of the underlying factors that drive each channel.

The opportunity cost is changing in the extent to which a new road increases the return to unskilled labor. We proxy this with the urban-rural unskilled wage gap, which we calculate at the district level using the 68th round of the National Sample Survey. To proxy the

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

expected difference in returns to education between an unconnected village and the broader market, we use Mincerian regressions to estimate local returns to education.[24] We calculate rural returns to education using village-level Mincerian regressions, using comprehensive household income and education data from the Socioeconomic and Caste Census of 2012. We then calculate (i) average village-level returns to education; (ii) average district-level returns in villages without roads at baseline; and (iii) average district-level returns in all villages. For urban returns, we do not have town-specific microdata, so we generate district averages from the 68th round of the NSS. Finally, to proxy for the potential importance of the income and liquidity effects, we use the share of households in a village who report owning none of the surveyed assets in the 2002-2005 Below Poverty Line Census.[25] These are the households that are most likely to be liquidity constrained, as well as those for whom income effects are likely the largest.[26]

We run our standard panel regression, interacting each of these variables (the wage gap, the returns gap, and the baseline asset level) with the treatment indicator. If the interaction term is important, it provides suggestive evidence that the relevant mechanism is an important channel through which new roads affect schooling decisions. Table 4 shows the results. Column 1 repeats the main specification without interaction terms, in the set of data for which the interaction terms are non-missing. Columns 2 through 4 present results under three different definitions of the proxy for changing returns to education. In Column 2, we proxy the expected increase in returns to education with the district-level difference between urban and rural returns to education. In Column 3, we use the district-level difference in returns to education between villages that had roads in 2001 and those that did not have roads in 2001. In Column 4, we use the difference between village-level returns to education

[24] Specifically, we regress log income on years of education, age and age squared.

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

[26] We would prefer to calculate returns to education using the Below Poverty Line Census, since it was taken before many PMGSY roads were constructed. However, we are missing the education data in the BPL for over half of our sample, so we generated returns using the more recent SECC.

and district-level urban returns to education. In each case, we use a binary interaction term, that indicates whether a village or district is above the sample median of the given variable.[27]

All point estimates go in the direction predicted by the model. When the opportunity cost of education is likely to increase the most, new roads have smaller or negative impacts on schooling. When the returns to education increase the most, or when households have the fewest assets, new roads have more positive effects on school enrollment. The interactions with returns to education and low assets are high statistically significant; the interactions with the opportunity cost variable are not statistically significant. The results suggest that both returns to education and income/liquidity effects are strong. As highlighted by Edmonds (2006), it is very difficult to distinguish between liquidity and income effects; both are plausibly large in our context. The lack of significance on the wage gap coefficient should not be taken as evidence that the opportunity cost effect is not important; the variation across district wage gaps in our sample is less than the average urban-rural wage gap. It is therefore possible that the opportunity cost effect is substantial on average, but nevertheless does not predict significant treatment heterogeneity. The wage gap interaction variable is also negative and statistically significant in the unbalanced panel (shown in Appendix Table A3).

Finally, Appendix Table A4 shows results from a fully interacted regression, which allows estimation of treatment effects in each of the eight subgroups defined by the binary district categorization. For clarity, we show estimated treatment effects in each subgroup, rather than the coefficients on the interaction terms. Treatment effects are negative and statistically significant only in one subgroup: the subgroup where we would expect a high opportunity cost effect, and low income, liquidity and returns to education effects. This is exactly as the model would predict. This decomposition also helps identify the distribution of treatment

[27] Appendix Table A2 shows results with continuous rather than binary interaction terms, and Appendix Table A2 shows results in the unbalanced panel. Results are broadly similar.

effects: new roads lead to statistically significant (at the 10% level) reductions in schooling in 9% of our sample (745/7741 villages), and to significant increases in schooling in 56% of the sample (4368/7741).

We acknowledge that these interactions cannot be interpreted as causal, as there could be other important district-level characteristics that are correlated with our proxies for opportunity cost, returns to education and income/liquidity effects and also influence treatment heterogeneity. Nevertheless, the results are highly consistent a standard human capital investment model.

6.3 Robustness

6.3.1 Regression Discontinuity Estimates

The panel estimates indicate that the positive effects of new market opportunities (a combination of liquidity, income and returns effects) dominate the opportunity cost effect. In this section, we provide additional evidence by showing this finding is robust to a regression discontinuity specification. This specification generates causal evidence under few assumptions, using a partially overlapping sample of the data, and should alleviate any concern that the exact timing of road construction, as exploited in the panel estimates, is endogenous to unobserved village-level factors.

A regression discontinuity estimation provides unbiased causal estimates under the following two assumptions: (i) the density of the running variable is continuous around the treatment threshold; and (ii) baseline variables are discontinuous across the treatment threshold. Appendix Figure A2 presents graphical evidence from the McCrary test; we fit a nonparametric function to the village population density, with allowance for a discontinuity at the treatment threshold (McCrary, 2008). No such discontinuity is apparent; the p-value for the test is 0.31.

To test the second assumption, we run the regression discontinuity specification on a range of variables measured at baseline. The running variable in these specifications is a linearly transformed population variable, such that the treatment threshold is at zero. We subtract 500 from the population for villages in the vicinity of the low treatment threshold, and 1000 from the population of states that used the high treatment threshold. Appendix Table A5 shows coefficient estimates for all control variables and outcome variables measured in the period before any roads were built. None of the point estimates are significantly different from zero at the 10 percent threshold. Appendix Figures A3 and A4 present graphical evidence that these variables do not vary systematically at the treatment threshold. In these graphs, each point represents the mean outcome within a bin defined by a narrow range of the transformed population, the running variable. We fit linear functions to data points above and below the treatment threshold; none show a significant difference across the threshold.[28]

Table 5 presents regression discontinuity estimates of the impact of road treatment on school enrollment. Column 1 reports the first stage estimate; the dependent variable is a village-level indicator that takes the value one if a village received a road under PMGSY. A village just above the population treatment threshold is 24 percentage points more likely to receive a new road under PMGSY; the sample mean of this variable is 0.20. Figure 4 presents a graphical analog to this estimate. Columns 2 and 3 report reduced form estimates of the impact of treatment on village-level log middle school enrollment in 2011, respectively with district and subdistrict fixed effects.[29] The reduced form estimates suggest that enrollment is 9 to 10 percentage points higher in villages just above the treatment threshold; location fixed effects have little effect on the estimates. Columns 4 and 5 present estimates from

[28]For additional robustness tests indicating balance of this specification on other economic variables, as well as details on how the sample of states was chosen, see Asher and Novosad (2016). They additional show further that there is no evidence of any other village public goods systematically changing at treatment thresholds.

[29]The regression discontinuity sample consists of 4 states and 481 subdistricts. India has 35 states and approximately 4000 subdistricts.

the IV specification, respectively with district and subdistrict fixed effects; the treatment estimate indicates that a new PMGSY road increases middle school enrollment by about 0.41 log points. Figure 5 shows the graphical analog of these estimates; the discontinuity at zero in the graph reflects the increased enrollment in villages just above the eligibility cutoff for the roads program.

The RD estimates are larger and less precise than the panel estimates; the panel estimates fall within the RD confidence interval. There are two factors that increase the RD estimates relative to the panel. First, because the panel includes years before 2011, the average treated observation in the panel estimates has had a road for approximately 2 years, while the average treated observation in the RD sample has had a road for 3.5 years.[30] Second, the regression discontinuity specification estimates a local average treatment effect on villages that are both smaller and have less baseline enrollment. This is because the panel includes only villages treated before 2011, while the regression discontinuity sample includes villages that received roads after 2011 or did not receive roads under the program at all. The RD estimates reinforce the panel estimates using a different sample of data and a different empirical strategy, mitigating concern that the panel estimates reflect other factors occurring in villages at precisely the time of road construction.

6.3.2 Migration, School Quality and Other Spatial Effects

A possible concern with the results above is that they could be driven by a reduction in outmigration from treated villages. We do not have data on the number of school age children per village, hence we cannot directly test for changing migration patterns of families with middle-school-age children. However, we present two other results suggesting that there is no outmigration. The lower panel of Appendix Figure A5 shows the regression discontinuity estimate of road treatment on village population. There is clearly no discontinuity at the

[30] The median road in the treated group in the regression discontinuity sample was built in early 2008.

threshold; we can rule out the exit of more than four people from a treated village, and the point estimate is close to zero. Second, changes in migration would be expected to equally affect families with primary school age children. Appendix Tables A6 and A7 respectively present panel and regression discontinuity estimates of the impact of roads on log *primary* school enrollment. The estimates are zero or slightly negative in some specifications; there is no evidence suggesting reduced outmigration in treated villages.[31]

A second question is whether the results could be driven by changes in school quality or in the number of schools available. Appendix Table A8 shows panel and regression discontinuity estimates of the effects of road completion on school quality, as measured by a series of school infrastructure measures included in the DISE data. The final row shows estimates on the number of schools reported in DISE. While some specifications show positive effects on school infrastructure, the effects are less than one fifth the size of treatment effects on enrollment, and none are statistically significant in more than two out of five of the specifications.[32] We do not adjust for multiple testing, as our objective here is to demonstrate the lack of large effects of roads—multiple comparison adjustments would strengthen the case that there are no effects on school infrastructure or quantity. The evidence thus rejects the possibility that the school enrollment effects are driven by changes in school quantity of quality.[33]

Next, we test whether our results could be driven by displacement effects. That is, could the increase in enrollment in villages with new roads just reflect students exiting schools in nearby villages in order to enter the treated village? This seems implausible, as villages with no road by 2001 were among the poorest and most remote villages in India. It is thus unlikely

[31]Since the dependent variable is gross enrollment (rather than an enrollment rate), outmigration of students with a high propensity to drop out could not drive our estimates. Students not in school do not affect enrollment figures, whether they stay in the village or not.

[32]The exception is the presence of school computers, which if anything are lower in villages that received new roads.

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

that schools in these villages would be particularly desirable for anyone. Nevertheless, we tested whether there is reduction (or increase) in enrollment in villages *close* to treated villages. We calculated total annual middle school enrollment for all villages within a 3km and 5km radius of each village that received a new road. Columns 1 and 2 of Appendix Table A9 report panel estimates of the impact of roads on log middle school enrollment in surrounding villages; we estimate precise zeroes, ruling out displacement effects.

Finally, we examine the possibility that the mechanism for the impact of the road is increased accessibility to the school itself, rather than to outside labor markets. This is suggested by Muralidharan and Prakash (2013), who find the provision of bicycles made girls more likely to attend middle and high school. If schools are unreachable due to flooding and muddy trails at certain times of year, a village feeder road could thus increase enrollment by making it easier for children to walk to school. We tested two proxies for this narrative. First, we estimated the impact of roads on schooling in villages with below- and above-median surface areas. Children living in dispersed villages have further to walk to school, and thus would be expected to benefit more from a new road. Columns 3 and 4 of Appendix Table A9 show that treatment effects are almost identical in dispersed and dense villages; the result holds whether we use gross or per capita area. If our treatment effects are driven by ease of access to village schools, we should also see larger effects in villages where there are many nearby children who do not have access to middle schools. To test this, we counted the number of school age children within a 5km radius of treated villages, who were living in *villages without middle schools*.[34] Columns 5 and 6 of Appendix Table A9 show that treatment effects are virtually identical in villages close to more or less underserved children. The evidence therefore does not support easier walking access to school as a primary mechanism for our findings.

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

7 Conclusion

Poor access to markets, even within their own country or region, is a central feature of the lives of the very poor. Connecting remote villages to high quality transportation networks is a major goal of both governments of developing countries and development agencies. These roads will bring access to new opportunities; however, policy-makers may be rightly concerned that access to opportunity can paradoxically cause disinvestment in the human capital accumulation that is central to long-run growth.

We shed light on this issue by studying the impact of India's flagship rural road program, which has built feeder roads to 100,000 villages in India in the last 10 years. We show that the building of these roads had large positive effects on adolescent school enrollment. The immediate draw of labor markets is nevertheless important: in the (small) subset of villages where the appeal of outside labor markets is strongest relative to the returns to education, we find that adolescents may in fact exit school when a new road is built. The study takes place in the context of a nationally increasing trend in middle school enrollment; average enrollment is thus rising even in the newly connected villages where the opportunity cost effects are the largest. More broadly, our results provide support for a standard human capital investment model: individual schooling decisions respond to shocks to local returns to education.

This paper also highlights an understudied but important impact of rural infrastructure. Such investments are usually premised on their potential to bring economic opportunities and growth to rural areas. If road construction leads to increased human capital accumulation, then its long run impact could be much larger than previous studies suggest.

References

Adukia, Anjali, "Sanitation and education," *forthcoming: American Economic Journal: Applied Economics*, 2016, (May), 1–64.

Afridi, Farzana, Abhiroop Mukhopadhyay, and Soham Sahoo, *Female Labour-Force Participation and Child Education in India: The Effect of the National Rural Employment Guarantee Scheme* number 95 2013.

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

Altonji, Joseph G, Todd E Elder, and Christopher R Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

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

Asher, Sam and Paul Novosad, "Market Access and Structural Transformation: Evidence from Rural Roads in India," 2016.

____, Karan Nagpal, and Paul Novosad, "The Tyranny of Distance: Remoteness and Welfare in India," 2016.

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

____ and Dave Donaldson, "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs," 2015.

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

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

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.

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

Cascio, EU and A Narayan, "Who needs a fracking education? the educational response to low-skill biased technological change," 2015.

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, pp. 1–37.

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

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

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

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.

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

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.

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

Hernández, Sara, "A Rose to Success: Shocks to the Flower Industry and Education in Colombia." PhD dissertation, Massachusetts Institute of Technology 2015.

Hine, J, Abedin M, RJ Stevens, T Airey, and T 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.

Islam, Mahnaz and Anitha Sivasankaran, "How does Child Labor respond to changes in Adult Work Opportunities ? Evidence from NREGA," 2014.

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

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

Jensen, R., "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.

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

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

Li, Tianshi and Sekhri, Sheetal, "The Unintended Consequences of Employment-Based Safety Net Programs," 2015, (September), 1–40.

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

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

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti, "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 2004, *112* (4).

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

Muralidharan, K and N Prakash, "Cycling to School: Increasing Secondary School Enrollment for Girls in India," 2013.

Oster, Emily and Bryce Millett, "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, "Drought of Opportunities: Contemporaneous and Long Term Impacts of Rainfall Shocks on Human Capital," *National Bureau of Economic Research Working Paper Series*, 2013, *No. 19140.*

____ and ____, "Workfare and Human Capital Investment: Evidence from India," 2015.

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

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

UNESCO, "UNESCO Institute for Statistics: Education statistics.," Technical Report, UNESCO Institute for Statistics 2016.

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

	No Paved Road	Paved Road	Total
Population (2001 Census)	1582.1	983.4	1326.8
	(1084.1)	(777.8)	(1009.8)
Total Employment (1998 Economic Census)	81.61	31.10	60.08
	(218.3)	(81.72)	(175.5)
Number of Schools	1.905	1.436	1.705
	(2.425)	(1.299)	(2.036)
Total Enrollment (grades 1-8)	266.8	169.9	225.5
	(498.3)	(247.3)	(413.3)
Total Primary Enrollment (grades 1-5)	212.2	145.0	183.6
	(372.5)	(183.3)	(308.3)
Total Middle Enrollment (grades 6-8)	54.57	24.85	41.90
	(153.1)	(87.78)	(130.2)
Middle School Exam Passers	19.13	7.042	13.98
	(39.95)	(30.99)	(36.89)
Exam Passers with Distinction	7.530	2.077	5.205
	(18.06)	(9.552)	(15.27)

The table shows means and standard deviations of key variables at baseline, in the sample of villages that were matched across all analysis datasets. Unless otherwise indicated, the data source is the District Information System for Education, 2002. Results are separated by a village's road status according to the village directory of the 2001 Population Census.

Table 2

Panel Estimates of Impact of Road on Middle School Enrollment Growth

	(1)	(2)	(3)	(4)	(5)
New Road	0.073	0.060	0.049	0.070	0.042
	(0.016)***	(0.013)***	(0.015)***	(0.013)***	(0.017)**
State-Year F.E.	Yes	Yes	Yes	Yes	Yes
Village F.E.	Yes	Yes	Yes	Yes	Yes
Village Time Trends	No	Yes	No	No	No
Panel Sample	Balanced	Balanced	Balanced Post-2004	Unbalanced	3 Years Pre/Post
N	119050	119050	83335	178112	42609
r2	0.80	0.91	0.88	0.76	0.85

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

The table reports panel estimates of the effect of new road construction on village-level middle school enrollment, estimated with Equation 3. Column 1 presents our main balanced panel specification. Column 2 adds a separate linear time trend for each village. Column 3 restricts the sample to years 2005 or later. Column 4 presents estimates from an unbalanced panel, which brings additional villages that do not have data in all years. Column 5 presents a specification which includes data only for three years before each road is built and three years after. Different years are thus included for different villages, but each village has seven observations. Due to data availability, the Column 5 sample only includes roads built between 2005 and 2008. All estimations have state-year fixed effects, and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table 3

Panel Estimates of Impact of Road on Middle School Completion Examinations

	Exam Taken	Exam Passed	High Exam Score
	(1)	(2)	(3)
New Road	0.060	0.062	0.032
	(0.022)***	(0.021)***	(0.015)**
State-Year F.E.	Yes	Yes	Yes
Village F.E.	Yes	Yes	Yes
Panel Sample	Balanced	Balanced	Balanced
N	32240	32475	32721
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 3. All columns use a balanced panel specification, analogous to Column 1 in Table 2. The dependent variable in Columns 1 through 3 is, respectively: (1) the log of the number of students sitting for the middle school completion examination; (2) the log number of students who pass this exam; (3) the log of the number of students who pass this exam with distinction. All estimations have state-year fixed effects, and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table 4

Panel Interaction Estimates of Impact of Road on Middle School Enrollment Growth

	(1)	(2)	(3)	(4)
New Road	0.072	-0.032	-0.074	-0.033
	(0.020)***	(0.047)	(0.050)	(0.048)
New Road * High Rural-Urban Wage Gap		-0.042	-0.034	-0.039
		(0.035)	(0.036)	(0.036)
New Road * Low Assets		0.130	0.143	0.123
		(0.044)***	(0.045)***	(0.046)***
New Road * High Rural-Urban Returns Gap (district)		0.068		
		(0.034)**		
New Road * High Paved Rural Returns Gap (district)			0.104	
			(0.034)***	
New Road * High Rural-Urban Returns Gap (village)				0.062
				(0.035)*
State-Year F.E.	Yes	Yes	Yes	Yes
Village F.E.	Yes	Yes	Yes	Yes
Panel Sample	Balanced	Balanced	Balanced	Balanced
N	76620	81960	76620	77410
r2	0.81	0.82	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-level school enrollment, interacted with district-level measures of rural-urban wage gaps, household assets, and a measure of the skill premium associated with a new road, estimated with Equation 3. 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 results with different measures of the change in skill premium associated a new road. Column 2 uses the district-level difference between the rural and urban skill premia. Column 3 uses the district-level difference between the skill premium in villages with roads and in villages without roads. Column 4 uses the difference between the district-level urban skill premium, and the village-level rural skill premium. All estimations have state-year fixed effects, and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table 5

Regression Discontinuity Estimates of Impact of Road on Middle School Enrollment

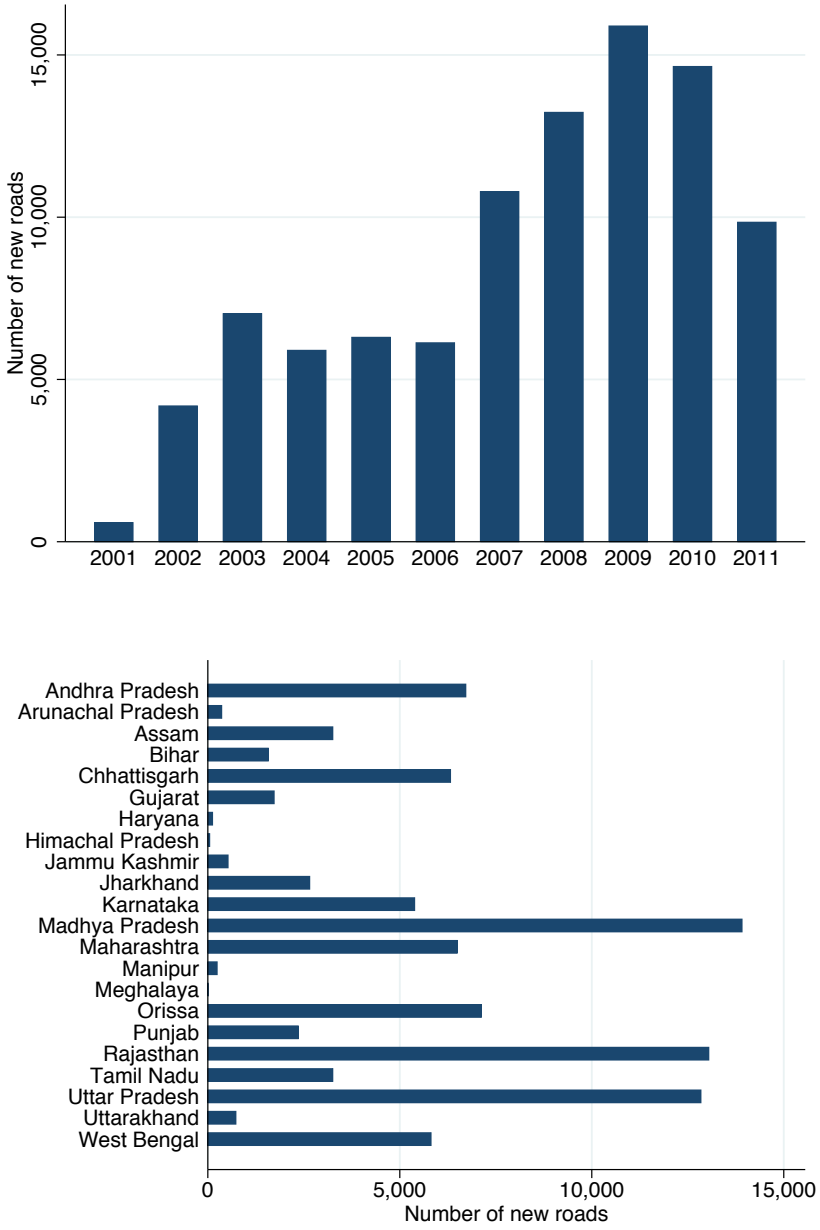
	First Stage	Reduced Form		IV	
	(1)	(2)	(3)	(4)	(5)
Above Population Threshold	0.238	0.104	0.096		
	(0.012)***	(0.045)**	(0.045)**		
New Road by 2011				0.417	0.405
				(0.183)**	(0.189)**
Population * 1(Pop $\leq$ Threshold)	0.305	1.889	1.916	1.777	1.792
	(0.060)***	(0.222)***	(0.223)***	(0.255)***	(0.259)***
Population * 1(Pop $\geq$ Threshold)	0.049	1.654	1.787	1.653	1.767
	(0.065)	(0.242)***	(0.243)***	(0.244)***	(0.246)***
Fixed Effects	Subdistrict	District	Subdistrict	District	Subdistrict
N	16463	16463	16463	16463	16463
r2	0.31	0.50	0.53	0.50	0.52

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

The table shows regression discontinuity estimates of the impat of new road construction on village-level middle school enrollment, estimated with Equation 2. Column 1 reports first stage estimates of the effect of being above the state-specific population threshold that defines eligibility for a PMGSY road on the probability of receiving a new PMGSY road before 2011. Columns 2 and 3 show reduced form regression discontinuity estimates of the impact of being above the population eligibility threshold on log middle school enrollment, respectively with district and subdistrict fixed effects. Columns 4 and 5 show instrumental variable estimates of the impact of a new PMGSY road on village level middle school enrollment, respectively with district and subdistrict fixed effects. All specifications also control for log baseline school enrollment, so the point estimates can be interpreted as log growth estimates.

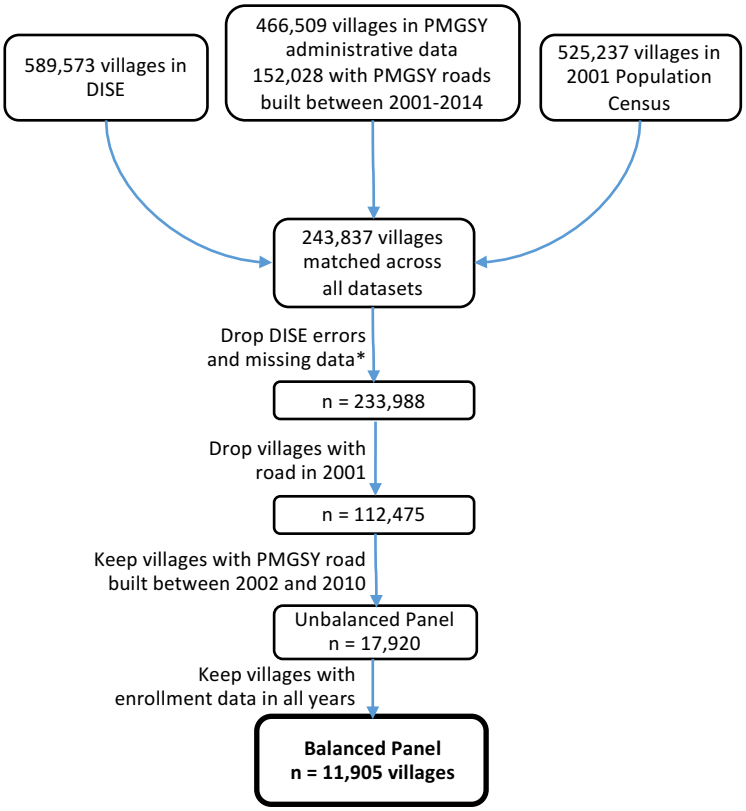
Figure 1

PMGSY Road Construction Summary Statistics



The figure describes the distribution of new roads built under the PMGSY program between 2001 and 2011, across years and states. Graphs show new roads according to their registered completion dates. Data source: PMGSY Online Monitoring and Management System.

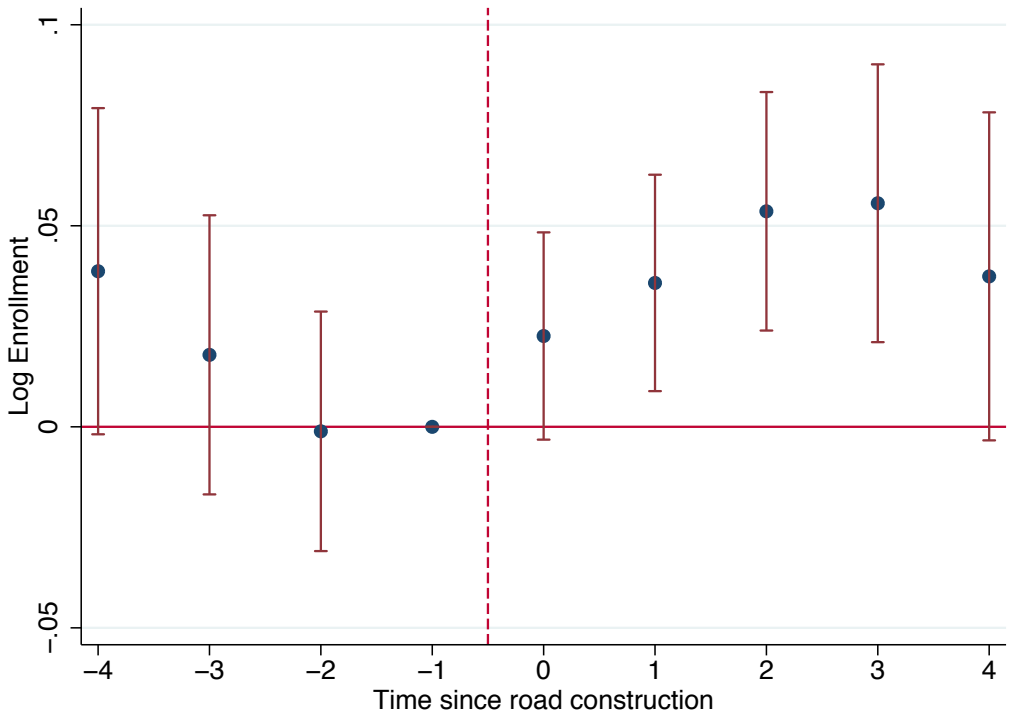
Figure 2
Sample Construction



The figure shows how we arrived at our final number of observations from the original datasets. DISE = District Information System for Education. PMGSY = Prime Minister's Road Building Program. All observation counts indicate number of villages at each stage.
*Observations were dropped if DISE reported grade one to eight enrollment greater than 60% of village population (99th percentile), year-on-year enrollment growth outside of interval (-73%, +270%) (99th percentile), or zero enrollment in all years. State-years were dropped if DISE reported enrollment for fewer than 25% of villages (Jharkhand 2005, Karnataka 2005, Uttarakhand 2006).

Figure 3

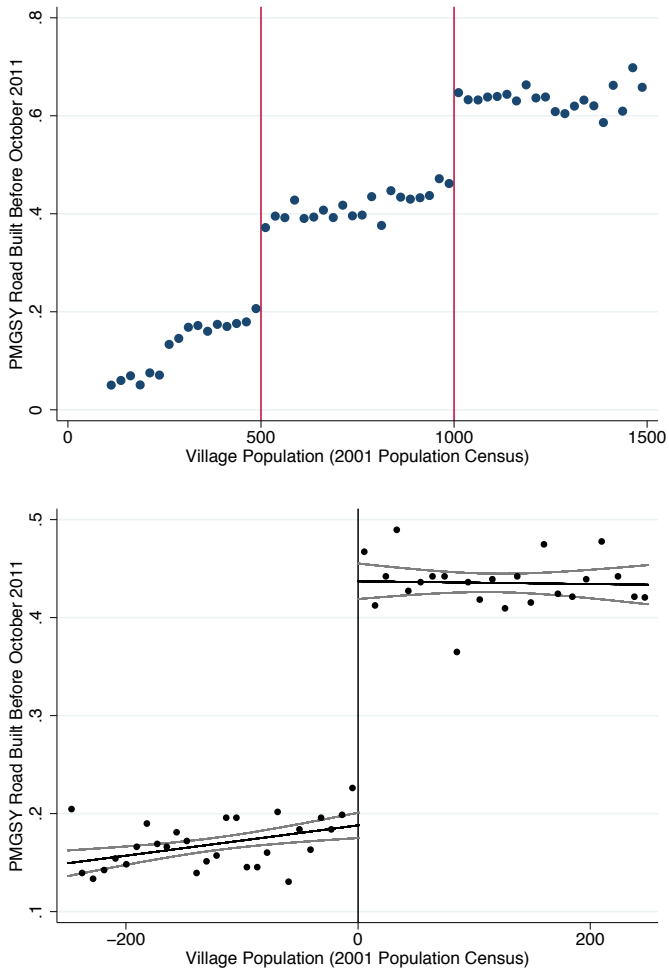
Panel Estimates of Effect of Roads on Middle School Enrollment



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 PMGSY road was constructed, along with a set of state*year fixed effects and village fixed effects. Year 0 is the first year in which a road was present when enrollment data were collected in October. The year before road completion (t = -1) is the omitted indicator. 95% confidence intervals are displayed around each point estimate. Standard errors are clustered at the village level.

Figure 4

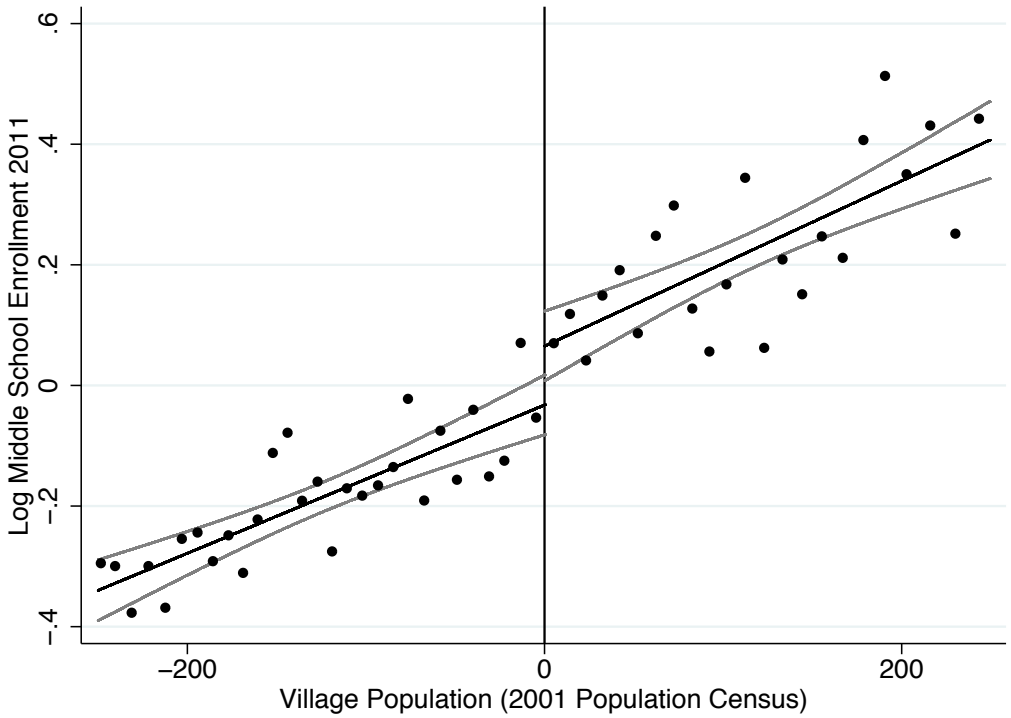
Regression Discontinuity First Stage: Population and Share of Villages Treated



The two graphs plot the conditional expectation function of a dummy variable indicator that a village has received a PMGSY road before 2011, the last year in our school enrollment sample period, conditioning on the village population, as reported in the 2001 Population Census of India. Each point represents the mean of approximately all villages in the given bin defined by population (1,117 villages per bin in the top panel, and 328 villages per bin in the bottom panel). The top graph shows raw population on the X axis. The bottom panel shows a population variable that has been centered around the state-specific threshold used for PMGSY eligibility, which is either 500 or 1000. Points to the right of the center line represent villages with a higher likelihood of treatment under PMGSY, according to program rules.

Figure 5

Regression Discontinuity Reduced Form: Population and Middle School Enrollment



The figure plots the conditional expectation function of the mean of village-level log enrollment in middle school (grades 6-8) in 2011, the last year in our enrollment sample period, conditioning on the village population, as reported in the 2001 Population Census of India. The Y variable is the residual of a regression of log middle school enrollment on subdistrict fixed effects and beginning of sample (2002) enrollment. Population is normalized to be centered around the state-specific threshold used for the road program (PMGSY) eligibility, which is either 500 or 1000. Each point represents the mean of approximately 328 villages in the given bin defined by population.

A Appendix: Additional Figures and Tables

Table A1

Panel Estimates of Impact of Road on Middle School Enrollment: Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
New Road	0.030	0.050	0.023	3.275	1.068	2.238
	(0.009)***	(0.012)***	(0.008)***	(0.570)***	(0.498)**	(0.576)***
Lagged Log Enrollment	0.600	0.252	0.638			
	(0.004)***	(0.006)***	(0.004)***			
State-Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Village F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Village Time Trends	No	Yes	No	No	Yes	No
Sample	Balanced	Balanced	Unbalanced	Balanced	Balanced	Unbalanced
N	115077	115077	150274	119050	119050	178112
r2	0.88	0.92	0.88	0.79	0.90	0.70

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

The table reports panel estimates of the effect of new road construction on village-level middle school enrollment, estimated with Equation 3. Estimates are analogous to those in Table 2, with the following modifications. Columns 1 through 3 include a lagged dependent variable; in columns 4 through 6, the dependent variable is the level of middle enrollment, rather than the log of middle school enrollment. Columns 2 and 5 add linear village time trends, and Columns 3 and 6 show results for the unbalanced panel. All estimations have state-year and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table A2

Treatment Heterogeneity in Panel Estimates of Road Impacts: Continuous Interaction Terms

	(1)	(2)	(3)	(4)
New Road	0.072	-0.394	-0.449	-0.362
	(0.020)***	(0.106)***	(0.111)***	(0.111)***
New Road * Rural-Urban Wage Gap		-0.018	-0.029	-0.022
		(0.044)	(0.045)	(0.045)
New Road * Low Assets		0.607	0.591	0.537
		(0.134)***	(0.135)***	(0.140)***
New Road * Rural-Urban Returns Gap (district)		-0.063		
		(0.128)		
New Road * Paved Rural Returns Gap (district)			4.126	
			(1.368)***	
New Road * Rural-Urban Returns Gap (village)				0.046
				(0.129)
State-Year F.E.	Yes	Yes	Yes	Yes
Village F.E.	Yes	Yes	Yes	Yes
Panel Sample	Balanced	Balanced	Balanced	Balanced
N	76620	81960	76620	77410
r2	0.81	0.82	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-level school enrollment, interacted with district-level measures of rural-urban wage gaps, household assets, and a measure of the skill premium associated with a new road, estimated with Equation 3. All columns are analogous to those in Table 4, but this table uses continuous interaction terms rather than discrete interaction terms. All columns use a balanced panel, state-year fixed effects, and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table A3

Treatment Heterogeneity in Panel Estimates of Road Impacts: Continuous Interaction Terms

	(1)	(2)	(3)	(4)
New Road	0.096	0.031	-0.011	0.045
	(0.017)***	(0.039)	(0.042)	(0.041)
New Road * High Rural-Urban Wage Gap		-0.122	-0.119	-0.117
		(0.030)***	(0.031)***	(0.031)***
New Road * Low Assets		0.131	0.154	0.150
		(0.038)***	(0.039)***	(0.041)***
New Road * High Rural-Urban Returns Gap (district)		0.055		
		(0.030)*		
New Road * High Paved Rural Returns Gap (district)			0.114	
			(0.030)***	
New Road * High Rural-Urban Returns Gap (village)				0.013
				(0.032)
State-Year F.E.	Yes	Yes	Yes	Yes
Village F.E.	Yes	Yes	Yes	Yes
Panel Sample	Balanced	Balanced	Balanced	Balanced
N	105487	113129	105487	107103
r2	0.77	0.77	0.77	0.76

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

The table reports panel estimates of the effect of new road construction on village-level school enrollment, interacted with district-level measures of rural-urban wage gaps, household assets, and a measure of the skill premium associated with a new road, estimated with Equation 3. All columns are analogous to those in Table 4, but this table uses the unbalanced panel rather than the balanced panel. All estimations include state-year fixed effects, and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table A4

Treatment Heterogeneity in Panel Estimates of Road Impacts: Triple Interaction

Wage gap	Poverty	Returns to Education Gap	Treatment Effect	Number of Observations
Low	Low	Low	0.147	1914
			(0.083)*	
Low	Low	High	-0.244	65
			(0.172)	
Low	High	Low	0.063	1255
			(0.038)	
Low	High	High	0.147	1540
			(0.037)***	
High	Low	Low	-0.110	745
			(0.050)**	
High	Low	High	-0.039	465
			(0.071)	
High	High	Low	0.053	843
			(0.045)	
High	High	High	0.161	914
			(0.044)***	

The table reports panel estimates of the effect of new road construction on village-level school enrollment, fully interacted with district-level measures of rural-urban wage gaps, household assets, and a measure of the skill premium associated with a new road, estimated with Equation 3. The table shows the estimated treatment group in each subgroup, defined by the variables above. The number of observations varies across the bins because the categorical variables are correlated; for example, there are very few districts in our sample with below-median poverty, an above-median rural-urban skill premium, and a below-median urban-rural wage gap. We use the balanced panel, state-year fixed effects, and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table A5

Regression Discontinuity Baseline Tests

Variable	Baseline Mean	RD Estimate
Number of schools (DISE)	1.57	0.015
		(0.031)
Enrollment Divided by Population	0.21	0.002
		(0.005)
Log Total Enrollment (grades 1-8)	4.65	0.005
		(0.020)
Log Primary Enrollment (grades 1-5)	4.51	0.024
		(0.020)
Log Middle Enrollment (grades 6-8)	1.22	-0.003
		(0.070)
Log Students Passing Exam	0.47	0.018
		(0.040)
Log Students with Distinction on Exam	0.13	-0.012
		(0.019)
Literacy rate (2001)	0.45	0.001
		(0.005)
Scheduled Caste Population Share (2001)	0.14	0.001
		(0.006)
Distance to Nearest Town (km)	25.58	0.779
		(0.545)
Approach - Paved Road (2001)	0.14	0.025
		(0.014)
Number of Observations	17233	

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

The table reports regression discontinuity estimates of the change in baseline variables across the PMGSY eligibility threshold, using Equation2. All variables are measured in 2002 unless otherwise specified. All specifications include subdistrict fixed effects, and control linearly for population (the running variable) on each side of the treatment threshold. Standard errors are in parentheses. The data source for all school-related variables is the District Information System for Education; other variables are from the 2001 Population Census.

Table A6

Panel Estimates of Impact of Road on Primary School Enrollment

	(1)	(2)	(3)	(4)	(5)
New Road	-0.007	-0.001	-0.006	-0.017	0.009
	(0.004)	(0.004)	(0.004)	(0.005)***	(0.005)*
N	119050	119050	74795	178112	42609
r2	0.85	0.92	0.91	0.73	0.88

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

The table reports panel estimates of the effect of new road construction on village-level middle school enrollment, estimated with Equation 3. Column 1 presents our main balanced panel specification. Column 2 adds a separate linear time trend for each village. Column 3 restricts the sample to years 2005 or later. Column 4 presents estimates from an unbalanced panel, which brings additional villages that do not have data in all years. Column 5 presents a specification which includes data only for three years before each road is built and three years after. Different years are thus included for different villages, but each village has seven observations. Due to data availability, the Column 5 sample only includes roads built between 2005 and 2008. All estimations have state-year fixed effects, and village fixed effects, so constant terms are not displayed. Standard errors are clustered at the village level.

Table A7

Regression Discontinuity Estimates of Impact of Road on Primary School Enrollment Growth

	Reduced Form		IV	
	(1)	(2)	(3)	(4)
Above Population Threshold	0.008	0.003		
	(0.017)	(0.017)		
New Road			0.034	0.014
			(0.070)	(0.071)
Population * 1(Pop $\leq$ Threshold)	1.823	1.771	1.815	1.767
	(0.085)***	(0.082)***	(0.095)***	(0.095)***
Population * 1(Pop $\geq$ Threshold)	1.165	1.218	1.166	1.218
	(0.093)***	(0.090)***	(0.092)***	(0.090)***
Fixed Effects	District	Subdistrict	District	Subdistrict
N	16890	16890	16890	16890
r2	0.38	0.44	0.38	0.44

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

The table shows regression discontinuity estimates of the impat of new road construction on village-level primary school enrollment, estimated with Equation 2. Columns 1 and 2 show reduced form regression discontinuity estimates of the impact of being above the population eligibility threshold on log *primary* school enrollment, respectively with district and subdistrict fixed effects. Columns 3 and 4 show instrumental variable estimates of the impact of a new PMGSY road on village level primary school enrollment, respectively with district and subdistrict fixed effects. All specifications also control for log baseline school enrollment, so the point estimates can be interpreted as log growth estimates.

Table A8

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

	Panel Sample				
Dependent Variable	Balanced	Year $\geq$ 2005	Unbalanced	+/- 3 Years	RD
Piped water	0.001	-0.004	0.002	-0.006	0.002
	(0.003)	(0.003)	(0.003)	(0.005)	(0.007)
Toilet	0.003	-0.002	0.016	0.006	-0.002
	(0.003)	(0.004)	(0.003)***	(0.007)	(0.008)
Electricity	0.003	-0.002	0.004	-0.001	-0.004
	(0.002)	(0.002)	(0.002)**	(0.004)	(0.007)
Library	0.000	0.000	0.006	-0.008	0.004
	(0.004)	(0.004)	(0.003)*	(0.007)	(0.009)
Computer	-0.004	-0.005	-0.002	-0.008	0.001
	(0.002)***	(0.002)***	(0.001)	(0.003)**	(0.004)
Perimiter Wall	0.001	0.003	0.002	0.002	0.001
	(0.003)	(0.004)	(0.003)	(0.006)	(0.010)
Play area	0.009	0.003	0.007	0.008	0.006
	(0.003)***	(0.004)	(0.003)**	(0.007)	(0.010)
Log Number of Schools	0.008	0.002	0.001	-0.000	0.008
	(0.002)***	(0.002)	(0.002)	(0.003)	(0.006)

The table reports panel estimates of the effect of new road construction on village-level school infrastructure, estimated with Equation 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 restricts the sample to years 2005 or later. Column 3 presents estimates from an unbalanced panel, which brings additional villages that do not have data in all years. Column 4 presents a specification which includes data only for three years before each road is built and three years after. Columns 1 through 4 include state-year fixed effects, and village fixed effects. Column 5 presents reduced form regression discontinuity estimates of the impact on the infrastructure variable of being in a village just above the treatment threshold. Standard errors are clustered at the village level.

Table A9

Panel Estimates of Impact of Road on Middle School Enrollment: Spatial Effects

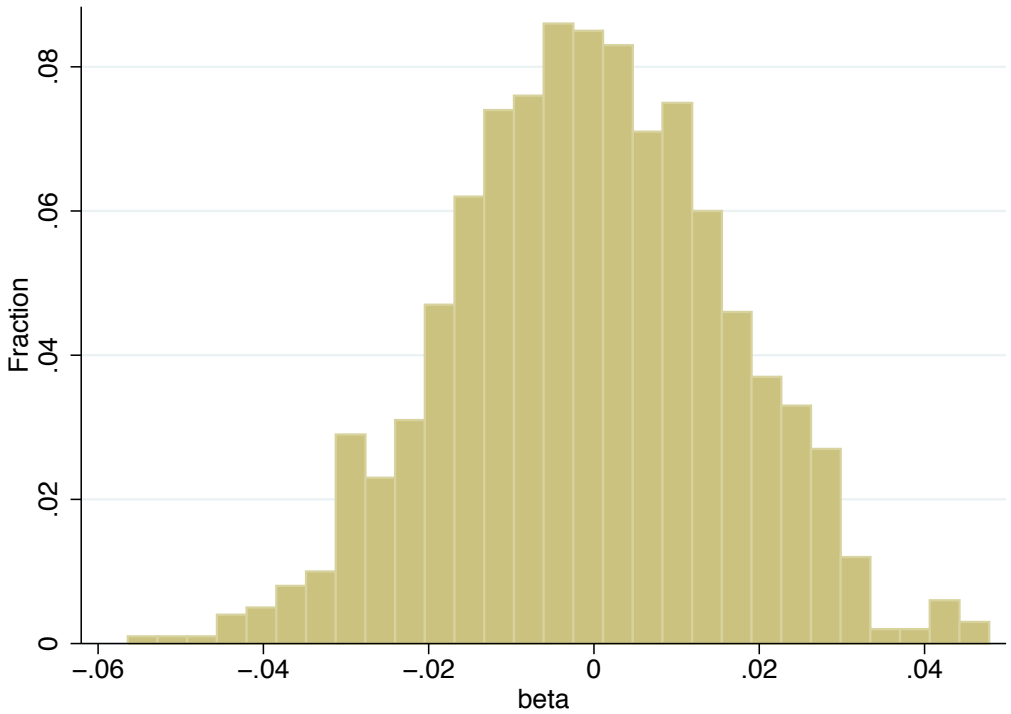
	Spillovers		Village Area		Nearby Eligible Kids	
	3km	5km	Low	High	Low	High
New Road	-0.008	-0.000	0.071	0.064	0.050	0.068
	(0.012)	(0.011)	(0.015)***	(0.014)***	(0.016)***	(0.016)***
State-Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
N	169653	174888	86777	86778	74196	74188
r2	0.80	0.80	0.76	0.76	0.75	0.76

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

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

Figure A1

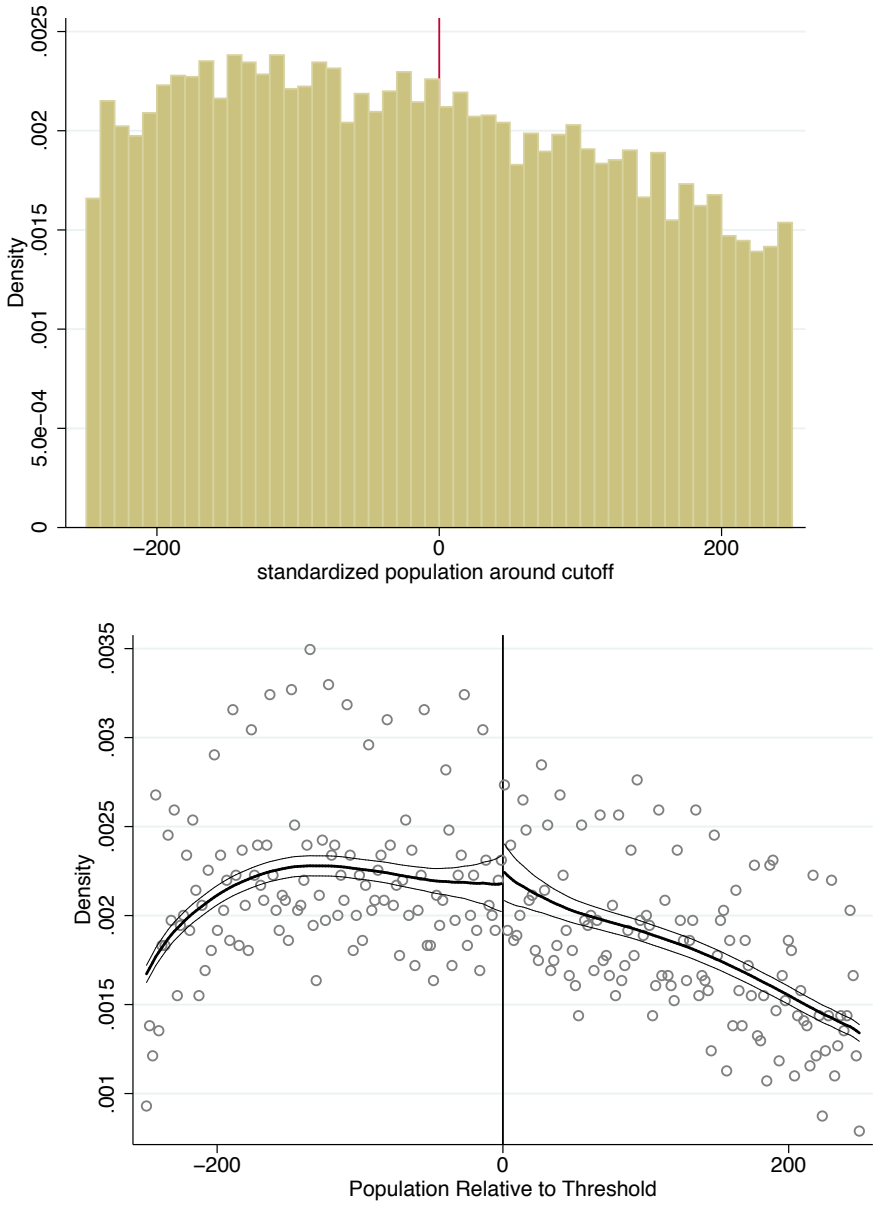
Panel Estimates of Effect of Roads on Middle School Enrollment: Permutation Test



The table 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 3. We ran this estimation 1000 times; the graph shows the distribution of estimates of β_1, which would be the impact of a new road on log middle school enrollment growth. The Table 2 estimate of 0.07 is beyond the right tail of the randomly generated estimates, which is consistent with our estimated p-value of less than 0.001, and suggests that the very low p-values are reasonable.

Figure A2

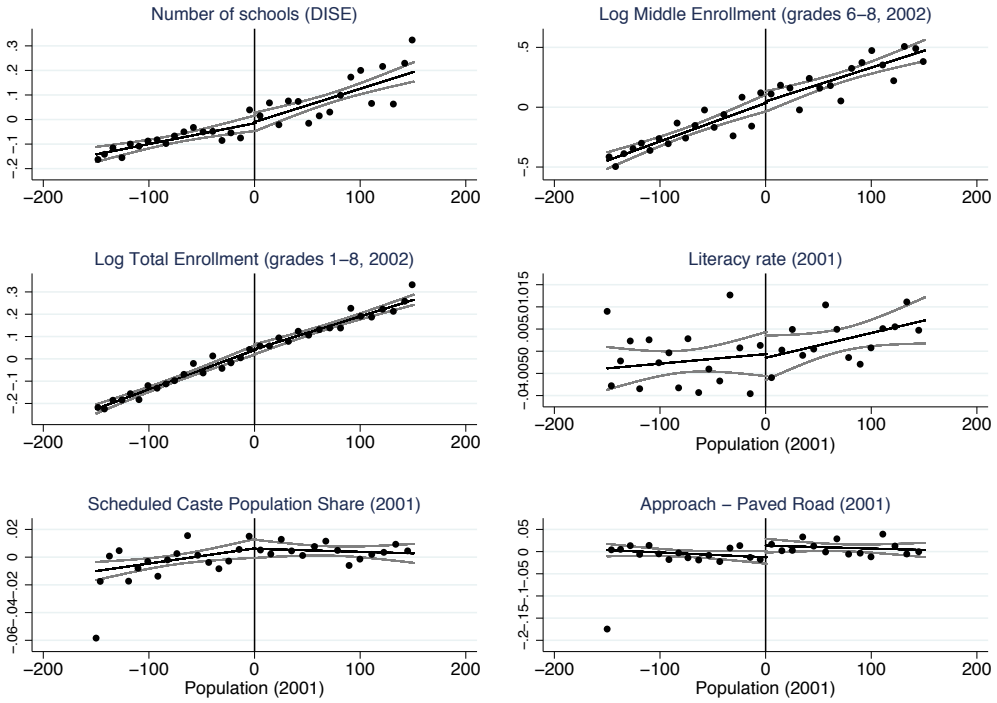
Regression Discontinuity: Continuity of Running Variable



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. 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. The population variable has been transformed so that the treatment threshold is zero.

Figure A3

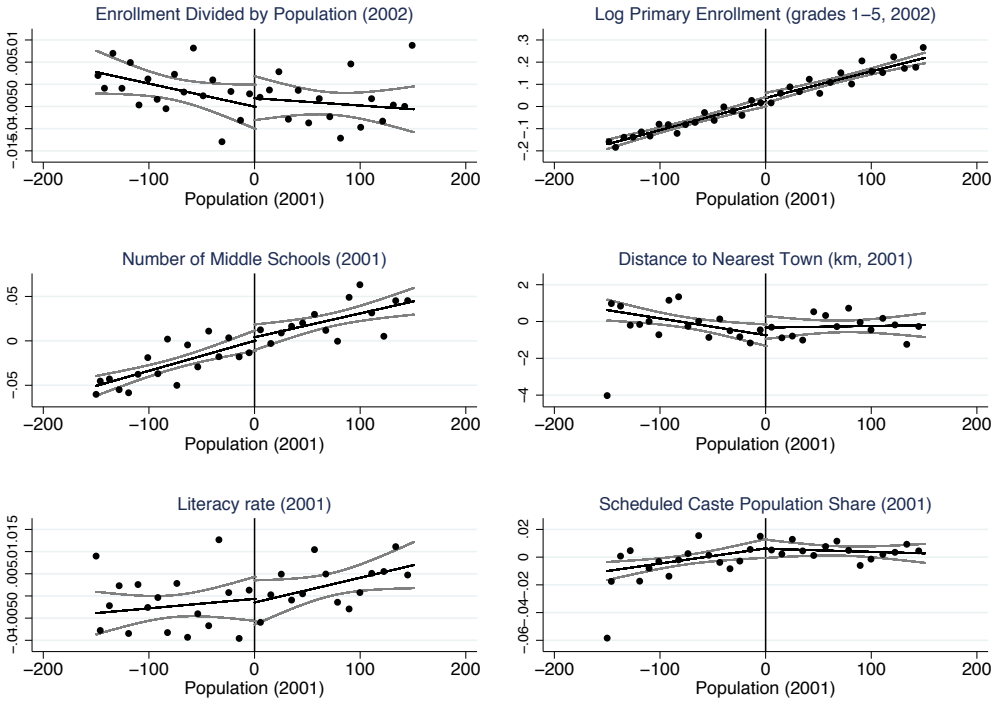
Regression Discontinuity: Continuity of Baseline Variables



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 A4

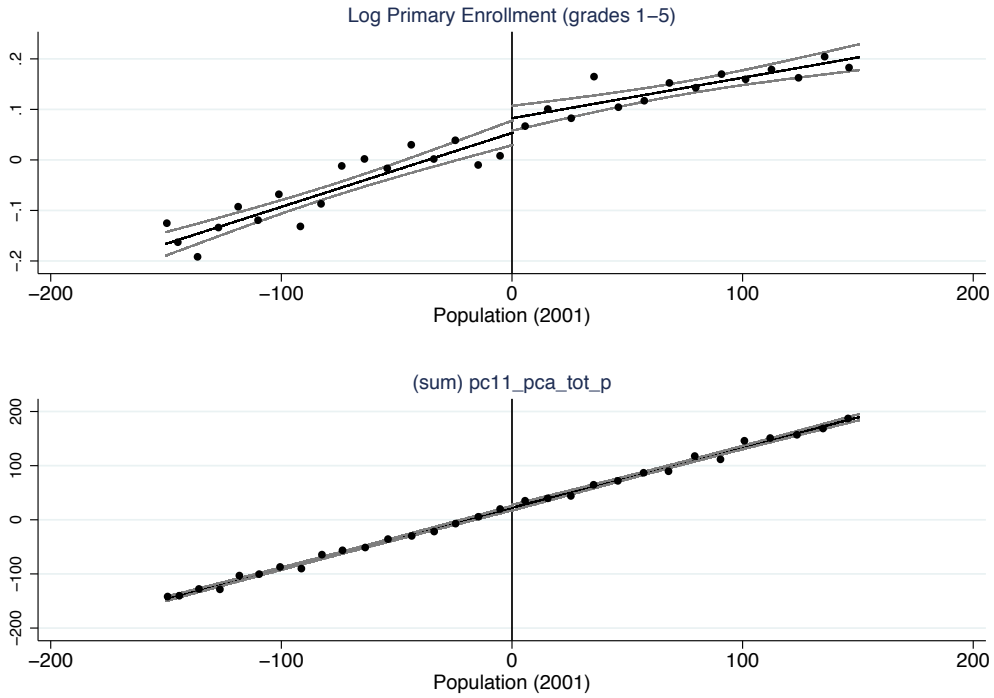
Regression Discontinuity: Continuity of Additional Baseline Variables



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

Regression Discontinuity Reduced Form: Population and Primary School Enrollment



The figure plots the conditional expectation function of the mean of village-level log enrollment in *primary* school (grades 1-5) in 2011, the last year in our enrollment sample period, conditioning on the village population, as reported in the 2001 Population Census of India. The Y variable is the residual of a regression of log primary school enrollment on subdistrict fixed effects and beginning of sample (2002) enrollment. Population is normalized to be centered around the state-specific threshold used for the road program (PMGSY) eligibility, which is either 500 or 1000. Each point represents the mean of approximately 328 villages in the given bin defined by population.