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

Anjali Adukia*

Sam Asher†

Paul Novosad‡

March 2019

*University of Chicago, 1307 East 60th Street, Chicago, IL 60637, adukia@uchicago.edu

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

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

Appendix: Additional Figures and Tables for Online Publication Only

Table A1

Urban vs. Rural Wages and Mincerian Returns to Education

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

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

Table A2
Impact of New Roads on Middle School Enrollment:
District-Year Fixed Effects

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

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

The table reports panel estimates of the effect of new road construction on village-level log middle school enrollment, estimated with Equation 1. Specifications are identical to Table 2, but with district-by-year fixed effects instead of state-by-year fixed effects. Column 1 presents the primary balanced panel specification. The dependent variable in columns 2 and 3 is log middle school enrollment for boys and girls respectively. Column 4 estimates the same regression with the level of middle school enrollment as the dependent variable, and columns 5 and 6 do the same for boys and girls respectively. All specifications include district-year fixed effects and village fixed effects. Standard errors are clustered at the village level.

Table A3
Regression Discontinuity Baseline Tests

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

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

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

Table A4
Panel Estimates in Regression Discontinuity Sample

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

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

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

Table A5
Treatment Heterogeneity in Road Impacts: Quartile Results

Panel A: Opportunity Cost Effect Quartiles

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

Panel B: Returns to Education Effect Quartiles

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

Panel C: Income/Liquidity Effect Quartiles

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

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

The table reports panel estimates of the effect of new road construction on village log middle school enrollment. The estimates are calculated for separate samples defined by quartiles of the mechanism proxies for the opportunity cost effect (panel A), the returns to education effect (panel B), and the income/liquidity effects (panel C). The size of the opportunity cost effect is proxied by the district-level mean low-skill urban wage minus the mean low-skill rural wage. The size of the returns to education effect is proxied by the difference between the urban and rural Mincerian returns to one additional year of education. The size of income and liquidity effects are proxied by the share of households in a village reporting zero assets in 2002. The estimating equation is Equation 1. All specifications include state-year fixed effects and village fixed effects. Standard errors are clustered at the village level.

Table A6

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

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

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

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

Table A7
Regression Discontinuity Estimates:
Other Public Goods

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

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

The table shows reduced-form regression discontinuity estimates of the difference in other public goods across the PMGSY population treatment threshold, using Equation 3. The dependent variable, column by column, is: (1) presence of primary school; (2) presence of middle school; (3) presence of secondary school; (4) village access to electric power; (5) presence of a primary health center; and (6) presence of a commercial bank. All specifications include district fixed effects and control for baseline log middle school enrollment, literacy rate, number of primary schools, number of middle schools, and the log number of non-farm jobs in the village.

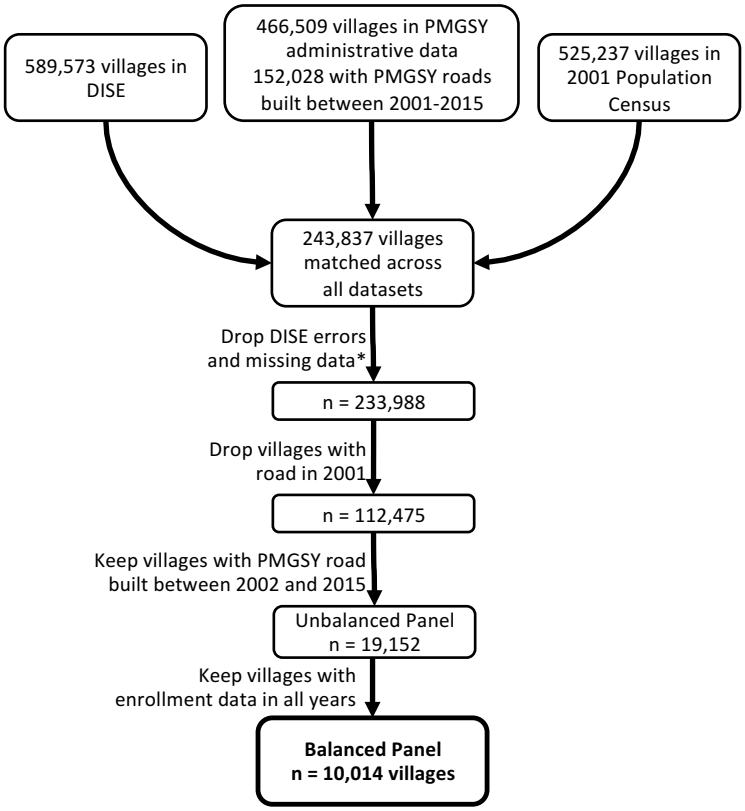
Table A8
Impact of Roads on Middle School Enrollment:
Spatial Effects

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

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

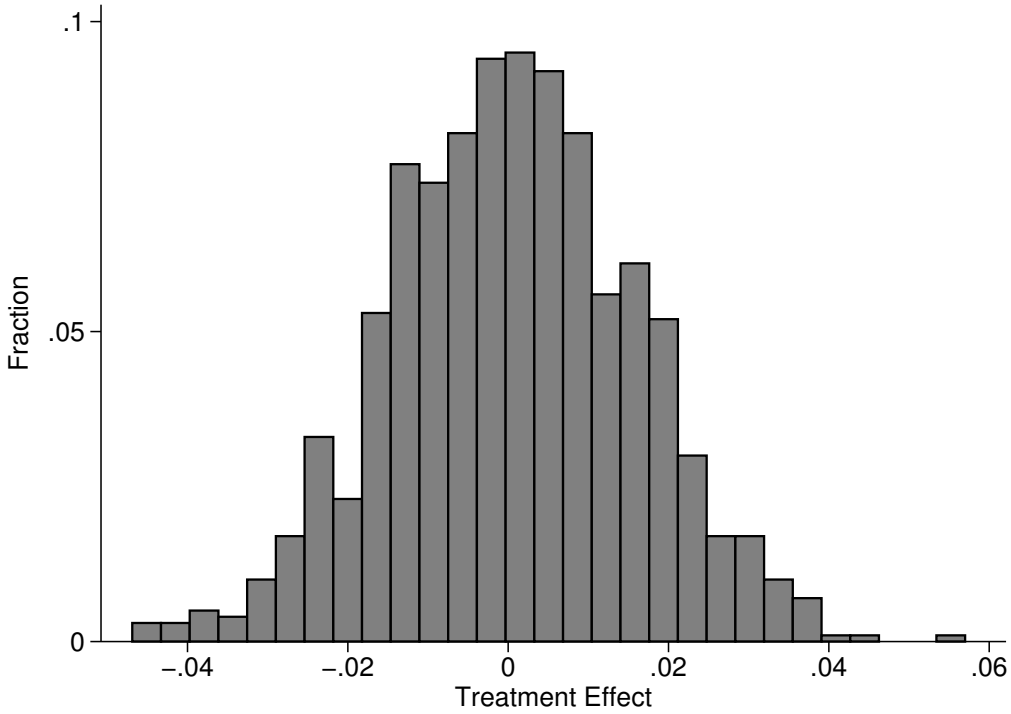
This table shows panel estimates of the impact of road construction on log middle school enrollment. Columns 1 and 2 show the impact of a new road on middle school enrollment in nearby villages, measured as those villages within a 3 km or 5 km radius, respectively. Columns 3 and 4 divide the sample into villages with above-median land area per capita and below-median land area per capita, and report effects separately. Columns 5 and 6 divide the sample into villages according to the number of children in nearby villages without middle schools. Column 5 shows the effect of new roads on middle school enrollment in villages with few nearby children in villages without middle schools; Column 6 shows estimates in villages where there are many nearby under-served children. All specifications include state-year fixed effects and village fixed effects. Standard errors are clustered at the village level.

Figure A1
Sample Construction



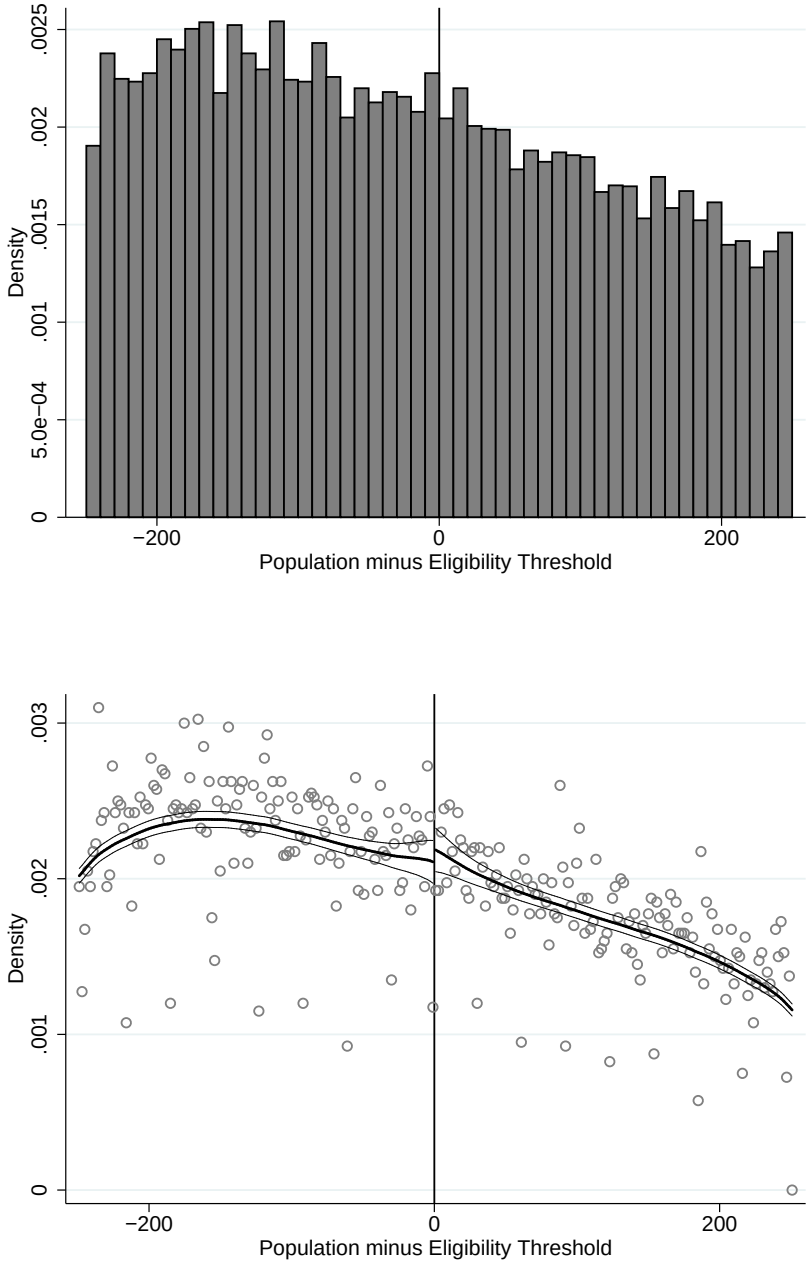
The figure shows how the sample was constructed from the original datasets. DISE refers to the District Information System for Education. PMGSY refers to the Prime Minister's Village Road Program. Observation counts indicate the number of villages at each stage. * Observations were dropped if DISE reported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

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



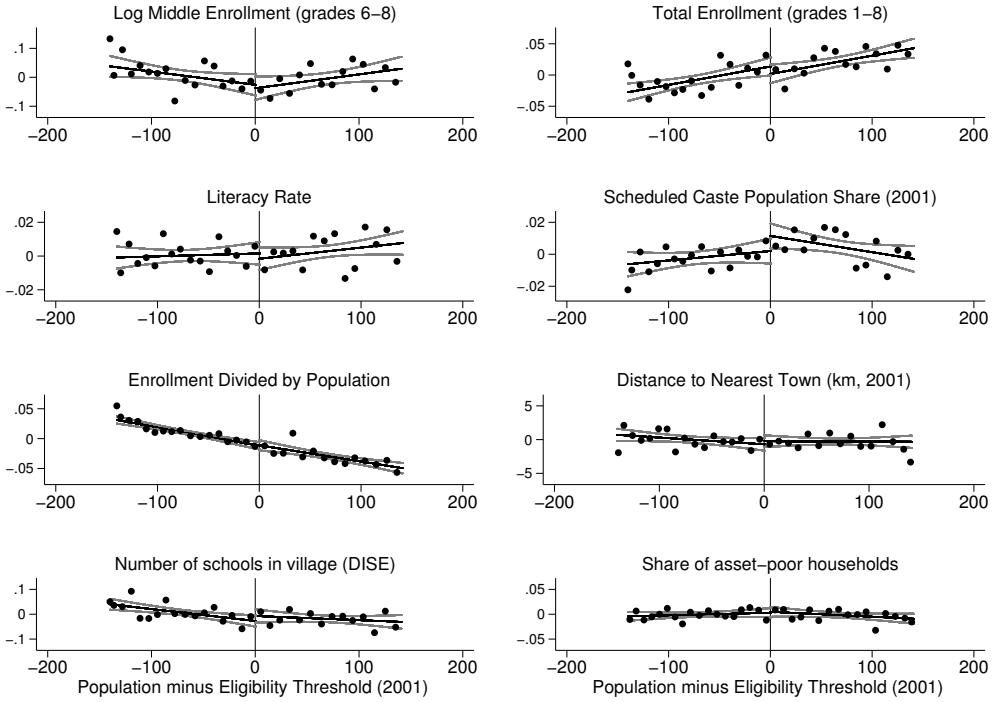
The figure shows the distribution of estimates from a placebo permutation test of the main panel specification presented in column 1 of Table 2. For each village in the main sample, we randomly generated a placebo year of road completion, and then estimated Equation 1. We ran this estimation 1000 times;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

Figure A3
Regression Discontinuity: Continuity of Running Variable



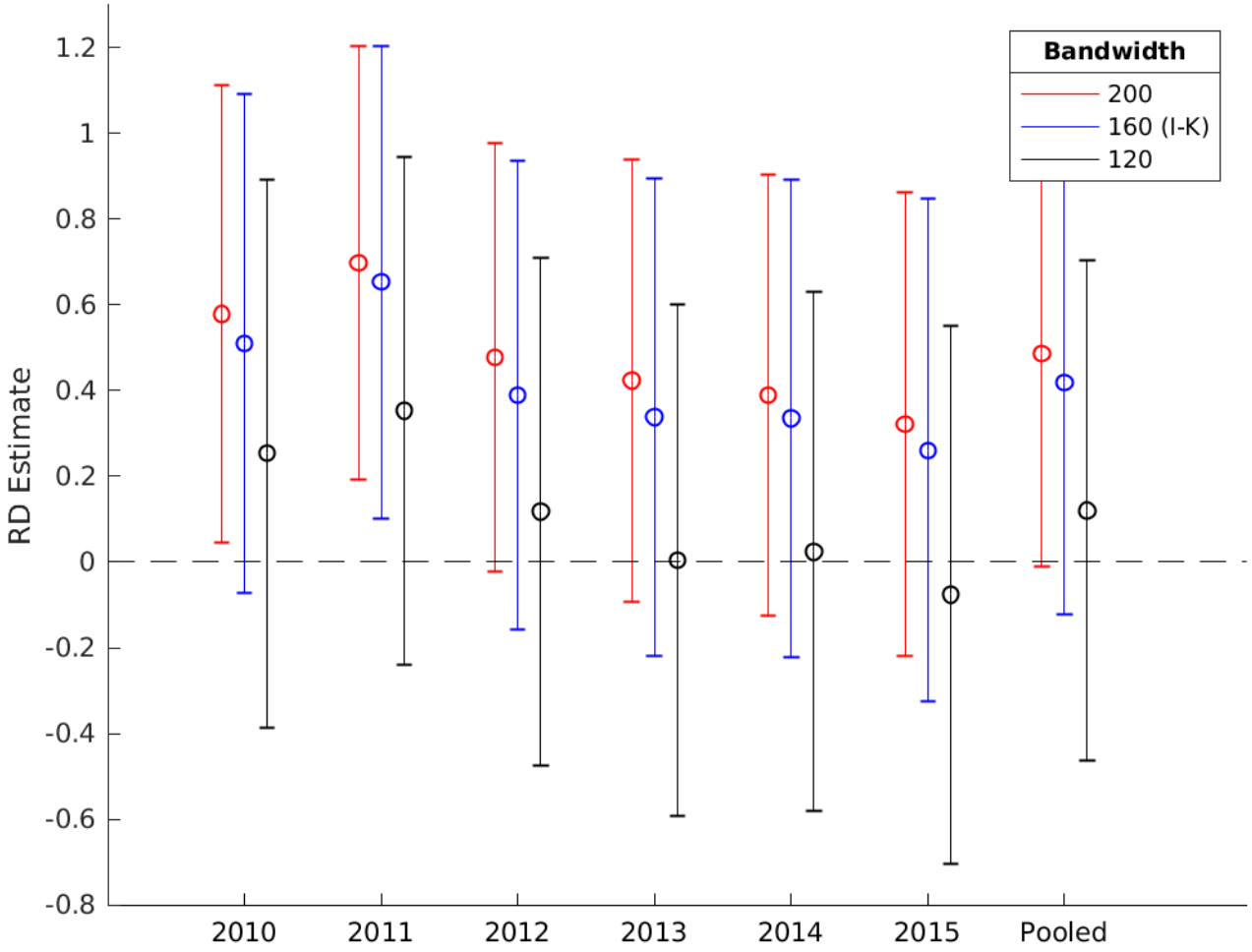
The figures show the distribution of village population in the set of villages in our sample. The top panel shows a histogram of village population, centered around the treatment threshold. In the bottom panel, we plot a non-parametric regression to each half of the distribution following McCrary (2008),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

Figure A4
Regression Discontinuity: Continuity of Baseline Variables



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

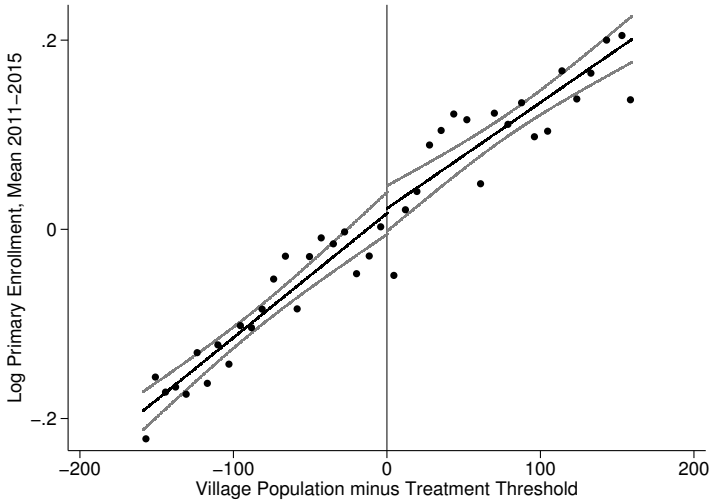
Figure A5
Impacts of New Roads on Middle School Enrollment:
Regression Discontinuity Estimates by Year and Bandwidth



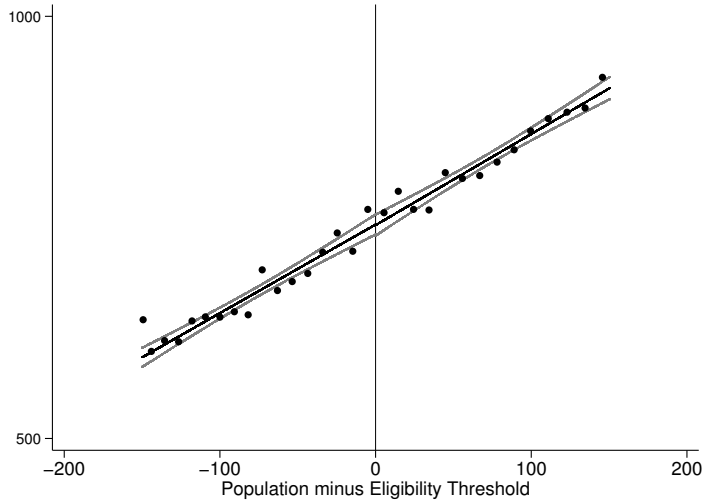
The figure shows IV estimates from Equation 3, estimated on different sample years, and at bandwidths 25% higher and lower than the optimal bandwidth of 160 selected with the algorithm of Imbens and Kalyanaraman (2012). Each point represents a single regression discontinuity estimate of the impact of new roads on log middle school enrollment. Error bars show 95% confidence intervals. The pooled estimate corresponds to that from Table 4, and pools years 2011-2015, clustering standard errors at the village level. All specifications control for baseline log middle school enrollment, literacy rate, number of primary schools, number of middle schools, and the log number of non-farm jobs in the village.

Figure A6
Regression Discontinuity Reduced Form:
Population and Primary School

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



Panel B: Log Population (2011)



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