Do Better Roads Increase School Enrollment? Evidence from a Unique Road Policy in India*

Mukta Mukherjee
Department of Economics
Syracuse University
110 Eggers Hall, NY-13244.
mumukher@maxwell.syr.edu
+1(518)364 1707

December 2, 2011

Abstract

In 2001, a national infrastructure development program was initiated in India to construct new all-season roads (roads that can be used in all-weather especially monsoons) in villages that previously had only had dry-season roads (roads that are difficult to use in monsoons). The eligibility rule that was used for undertaking construction of new all-season roads in these villages was that a minimum population of 500 had to be benefit from this road. This eligibility rule induced a nonlinear relationship between the population and the number of new all season roads in the villages of India today. In order to control for factors like communities' collective action ability that simultaneously determine the timing of completion of roads and students' enrollment in the schools, I instrument new roads with a population eligibility criteria. Exploiting the exogenous nature of the program eligibility criteria I compare students' enrollment in schools between villages on either side of the population cutoff. The most conservative estimates show that an improved access to school by better roads increases school enrollment by 22 percent in 2009. I find no

*I thank Jeffrey D.Kubik, Devashish Mitra and seminar participants of Syracuse University for their valuable comments. I have also benefited from helpful conversations with Debratta Dutta, Fariha Kamal and Sohini Sahu. This research was funded by Roscoe Martin Grant Award. I am grateful to NUEPA, PMGSY and GIS India for granting me to use their data. A special thanks to Naveen Bhatia (NUEPA),Vandana Sharma and Sangeeta (GIS) and Vignan (PMGSY) for their immense help.

spurious effects in 2002 when the same villages did not have a road. The effect of better access to roads on students' enrollment is heterogeneous depending on the age cohort and the caste (social background) to which they belong.

JEL No: H54, J24, I25, C14, C21, I21

Keywords: Infrastructure; Human Capital; Regression Discontinuity; India

1 Introduction

Physical distance to school is cited as a major barrier to participation for rural children in India (UNICEF, 2006; Ward, 2007). Similarly, in many other developing countries schools are not easily accessible[1], thus social scientists and policy makers are interested considerably in whether better access to schools increases students' enrollment (Duflo, 2001; Filmer, 2007; Handa, 2002). For example, in India, on average in most villages primary schools are one km away, middle schools are at three km away and secondary and higher secondary are five km away from the village center (Census, 2001; Ward, 2007). A considerable travel time is involved in accessing these schools. This time lost in travelling cannot be used either for productive activities or for leisure. It is just the additional cost that has to be borne to acquire education and is not used in actual learning. In many instances, the distances have to be covered on foot which leads to physical discomfort especially in hot summers and monsoons. The time lost is a major implicit cost in schooling decision.

Irrespective of this considerable interest in consequences of better access for school enrollment, measuring casual effects of access to school on schools' enrollment has proved to be very difficult. One of the common measures used is the new school availability in the locality (Duflo, 2001; Foster and Rosenweig, 1996; Jalan and Glinskaya, 2003). However, the placement of schools is not random. Usually the new schools are constructed in localities which previously suffered from low enrollment. This will lead to under estimation of the impact of the benefits of an improved access to schools on enrollment. On the other hand, if families who value schooling move towards localities with better schooling or schools are constructed in areas where the people value more education, the impact on enrollment will be over estimated. Another measure that has been used in cross-country studies is the average distance to the nearest primary or secondary schools or travel time on enrollment (Filmer, 2007; Bommier and Lambert, 2000; Handa, 2002). A robust pattern observed in most of these studies (Duflo, 2001; Jalan and Glinskaya, 2003; Filmer, 2007) is that the impact is highest for those who are interiorly located; usually these are the poorer sections.

In this study I use the regression discontinuity technique to surmount the fundamental problem of identification that the previous literature suffered from. This study uses a new measure for better access to schools by exploiting the provision of new all-season roads (roads that can be used in all weather especially monsoons) as a part of a national infrastructural program which was initiated in 2001 in India. The provision of new all-season roads (roads that can be used in all weather especially monsoons) in villages which

[1] The average distance required by a child to travel to reach the nearest primary schools ranges from 0.2 km in Bangladesh to 7.5 km in Chad. The distance to the nearest secondary schools ranges from 2 km in Bangladesh to 71 km in Mali (Filmer, 2007).

previously only had dry-season roads (roads that are difficult to use in monsoons) were done on the basis of a population eligibility criteria. This rule generates a potentially exogenous source of variation in the provision of new all-season roads which can be used to estimate the effects of better access to schools on the students' enrollment in rural India.

The importance of the population criteria rule in this study is that it can be used to determine the provision of new all-season roads in villages on the basis of their villages' population as in 2001. The implementation of the policy is as follows. According to the population criteria rule, there is no provision of all-season roads in villages when the population recorded is between 100 and 499, but when the population recorded is more than 500 then there is a sharp increase in the provision of all-season roads. In this study, villages with more than 500 populations as recorded in 2001 have a twenty five higher probability of being provided with a new all-season road. This study uses a novel dataset compiled from administrartive reports linking school enrollment at the village level to the number of new all-season roads.

Usually, in India the provision of public goods is simultaneously determined by many other factors. According to Banerjee, Iyer and Somanathan (2008) the communities' collective action ability is a major factor determining both the provision of public goods like roads and facilities in schools, affecting students' enrollment in India. In this study, I use the population rule to construct instrumental variable estimates of better roads effects. The most conservative estimates show that an improved access to school by better roads increases school enrollment by 22 percent in 2009. Further, the effect of better access to roads on students' enrollment is heterogeneous depending on the age cohort and social background to which they belong. The development of roads brings forth both intended and unintended effects on students' enrollment in school. Whereas for the younger students' participation rate in school increases, for the older students' participation rate in school decreases. The participation response of enrollment to a development of better roads is much higher for students from higher caste (students belonging to the higher social hierarchy scale) compared to those from backward caste (students belonging to the lower social hierarchy scale).

A possible explanation of the heterogeneous effects of roads on students on the basis of their age and social background is the following. The first phenomena can be explained in a child economy context with outside job opportunities. In comparison to the younger age cohort the higher age cohort students are physically able to reap the benefit of better job opportunities that a better connectivity brings forth; therefore their incentive to participate in school decreases with a development of a road. A possible explanation of the second phenomena is that the backward caste students in one hand benefits most from the development of better roads as they stay in the vicinity of the village, but in the other hand these students reaps less from an investment in schooling as the social restrictions they face constraints their future wage earnings.

The contribution of this paper is manifold. This study provides new insight into the question of whether students' enrollment increases with better access to school which has always interested economist and policy makers simultaneously. The study also contributes to a scanty body of literature on the measurement of economic benefits that comes with a better connectivity (Jacoby, 1998; Jacoby, 2008). Usually the benefits of an improved connectivity are calculated on the basis of hypothetical projects in countries like Nepal and Madagascar due to the lack of real instances. This unique road scheme allows study of the economic benefits from a wide scale real construction of roads. This paper overcomes limitations in the other past studies as it uses a novel dataset constructed from the administrative reports on eligible villages and their new roads completion information.

The remainder of the paper proceeds as follows: Section 2 lays down the details of the national infrastructure road development program introduced in India. Section 3 provides a brief theoretical framework of human capital investment and time lost in travelling. Section 4 discusses the identification technique and the data used for the analysis. Section 5 explores the empirical findings, with some robustness checks. Section 6 summarizes our findings and some limitations of the study.

2 Background

In India, roads form the life-line of villages as the access to schools, market and health centers is dependent on it. According to Bell (2010) the benefits of better roads in villages is manifold. First, as consumers they enjoy a reduced prices and as producers they can negotiate for higher price for their marketable surplus. Second, as students they can access schools located outside village. Finally, with a better access to medical amenities and crucial drugs not only can their health condition improve but it actually can make a difference between life and death in several scenarios. However, forty percent of all villages in 2000 were still unconnected by roads.

In 2001,[2] a national infrastructure development program PMGSY was initiated to fulfill the gap of roads in villages. The primary objective of this policy was to provide all-season roads to hamlets[3] that previously did not have any all-season roads within 0.5 km[4]. A secondary objective of this scheme was to upgrade the already existing all-weather

[2]The policy was declared on 25th December, 2000.

[3]A hamlet is a cluster of population, living in an area, the location of which does not change over time.

[4]The eligibility condition for new all weather roads for hilly states (North-East, Sikkim, Himachal Pradesh, Jammu & Kashmir, Uttaranchal), desert areas (districts, blocks eligible for Desert Development Projects) and Schedule 5 areas (constitution prescribed districts, block and villages with high populations of schedule castes and tribes) is that within a radius of 1.5km no all-weather roads exist and the minimum

roads based on the roads' deterioration. This gave emphasis to more populated areas, however only 20 percent of the funds were to be allocated towards this goal and the other 80 percent of the funds would be allocated towards the fulfillment of the primary objective.

The provision of new all-season roads followed an eligibility rule. The eligibility rule for a new all-season road (NR) linking several unconnected[5] hamlets ($h_1, h_2, ...h_n$) whose populations as recorded in Census 2001 ($p_1, p_2, ...p_n$) that would be considered for construction is given in equation (1). Equation (1) captures the fact that the PMGSY rule allows the unconnected hamlets with a population greater than 500 to receive new roads but hamlets eligible by virtue of distance with a population of 100-499 do not receive a new road. The discontinuity created by the 500 population threshold will be used as an identification strategy in this study.

$$\begin{aligned} NR &= 1 \; if \;\; \max \; (p_1, p_2, ...p_n) >= 500 \\ &= 0 \; if \;\; \max \; (p_1, p_2, ...p_n) < 500 \end{aligned} \tag{1}$$

The PMGSY was a national policy with an estimated cost of \$14 billion (Rs 600 billion) and more emphasis was given to states that had eligible unconnected hamlets. However, there is a great deal of state-wise variation in the progress of the construction and completion of new roads, depending on the local government initiative. In some states the federal grants were a part of a loan sanctioned by the World Bank (Jharkhand, Rajasthan, Uttar Pradesh, Himachal Pradesh) and Asian Development Bank (Assam, West Bengal, Orissa, Madhya Pradesh, Chhattisgarh). In return these states had to meet some conditions for safeguarding local environment and communities[6]. However, the eligibility rule for new all-season roads did not vary between the World Bank and ADB sponsored states and other non-sponsored states.

3 Human Capital Investment with a Travel Cost

This section presents a simple human capital investment model that informs the empirical analysis.

Economics. The human capital

Optimal level of Schooling Given that there is no productivity growth, the present

benefited populations has to be 250 at least.

[5]The hamlet would be considered unconnected if it does not have an all-weather road within a 0.5km radius.

[6]The existing standards were appraised and modified to minimize impacts on communities and environment.

value of life-time earnings of a "representative" individual with s years of education, evaluated at the age of school entry is

$$V(s,r) = \int_0^n y(s,\tau)e^{-r(s+x)}dx - C(s,\tau), \tag{2a}$$

where $\{y(s,\tau)\}$ is based on the estimated statistical earnings and r is the discount rate. The life-time earnings in (2a) can be expressed as

$$V(s,r) = \frac{\{y(s,\tau)\}\, e^{-rs}\left(1-e^{-rn}\right)}{r} - C(s,\tau), \tag{2b}$$

The individual maximizes the life-time earnings in (2b) w.r.t. choice of schooling s. The first order condition is given by

$$\frac{\{y_s - ry\}e^{-rs}\left(1-e^{-rn}\right)}{r} - C_s = 0, \tag{3}$$

With an improvement in the state of connectivity the equilibrium choice of schooling is (refer Theory Appendix)

$$\frac{ds}{d\tau} = -\frac{e^{-rs}\left(1-e^{-rn}\right)[y_{s\tau} - ry_\tau] - rC_{s\tau}}{e^{-rs}\left(1-e^{-rn}\right)[y_{ss} - ry_s + ry^2] - rC_{ss}}, \tag{4}$$

In (4) the denominator would always be negative to satisfy the second order condition for maximization. Therefore, the equilibrium choice of schooling would depend on the numerator. The first term $[y_{s\tau} - ry_\tau]$ denotes the present value of net gains from additional schooling after adjusting for forgone earnings. With an improved state of connectivity there would be two effects operating in the opposite direction ; $(y_{s\tau})$ the benefits from earnings from an additional period of schooling and ry_τ the forgone earnings adjusted for the interest rate. With a better access to the employment opportunities the potential wage earnings y_τ would also increase, thus the latter term also increases. Hypothetically, $y_{s\tau} \gtreqless 0$, but in reality for a short span of time $y_{s\tau} < 0$ can be ruled away. This implies that better connectivity leads to higher competition and that the returns from schooling decreases. But, such a scenario would take considerable time even after development in the level of transport. The second term $rC_{s\tau}$ denotes the present value of cost of travelling to school after there is an improvement in connectivity. This term is always negative on the basis of assumption that travel cost to school decreases with an improvement in connectivity. The schooling choice of the individual would depend upon which of the two effects dominate as is given in (5)

$$\frac{ds}{d\tau} \gtreqless 0 \; if \;\; [y_{s\tau} - ry_\tau] \gtreqless 0. \tag{5}$$

There are three possible cases depicted in Figure 1, but with each of them, with an improved state of connectivity the individual can reach a higher iso-wealth curve and is always better off than in the earlier state. The individual would choose a higher (lower) schooling level, when the net earnings gained from an additional period of schooling dominates (is dominated by) the loss of forgone earnings. If the two effects exactly offset each other the schooling level will remain unchanged.

The model yields the following testable hypotheses.

Hypothesis 1 In an improved state of connectivity, the lower the outside wage earnings potential the higher is, the higher the incentive to remain in school. Thus in an economy with child-labor opportunities increasing with an improved connectivity, the lower age group who are not physically fit to work have higher incentive to participate in school than a higher age group who can work (Basu, 1999). This leads to a following ranking of the schooling choice among different age groups less than 5 years, 5 to 9 years, 10 to 14 years and more than 14[7].

$$\frac{ds}{d\tau}_{less\ than\ 5yrs} > \frac{ds}{d\tau}_{5-9yrs} > \frac{ds}{d\tau}_{10-14yrs} > \frac{ds}{d\tau}_{15-18yrs}. \tag{6}$$

Hypothesis 2 In an economy where there is heterogeneity among sections by economical and social background, the benefits of an improved connectivity on different sections' students is ambiguous. In India, the schools are usually spatially clustered in the high caste[8] dominant areas and the low castes reside outside the villages. Thus, a village school would still be a considerable walking distance for these students. Therefore, the lower caste students would reap the benefit of an improved connectivity more than the higher caste students who usually reside within the village. However, the low castes are usually economically disadvantaged thus outside wage opportunities poses for more incentive to drop out of school. Henceforth, there are two effects that are simultaneously operating in opposite directions, and the decision to participate in school for different castes depends on which of these effects are dominant.

$$\frac{ds}{d\tau}_{low\ caste} \lesseqgtr \frac{ds}{d\tau}_{high\ caste}.$$

I test this by taking these hypotheses to the data accounting for various econometric challenges outlined below.

[7]According to the Child Labor Act, work by children less than 15 is considered as child labour. The Age group 5-14 years is the most vulnerable and in the age group 10-14 years the prominence of child labor is highest.

[8]In India, the Hindu caste system the society is divided into different caste on the basis of the profession they used to perform.

4 Estimation Strategy and Identification Assumptions

The identification strategy is intended to exploit the quasi-experiment created by the Indian rural road scheme using a fuzzy regression discontinuity (FRD) approach. [9]

The estimating equation at the village-level is the following:

$$(y)_{i,2009} = \alpha + \pi(NR)_{i,2009} + \varepsilon_{i,2009}; i = village \tag{7}$$

where $(y)_{i,2009}$ is village-level total school enrollment of students in year 2009. The regressor of interest $(NR)_{i,2009}$ is a dummy variable which equals 1 if a new all weather road has been constructed in the village as of year 2009 and 0 if it is either under-construction or no-construction have been undertaken ; $(\varepsilon_{i,2009})$ is the error term.

The coefficient of interest (π) indicates the causal effect of an improved access to schools by the construction of all weather roads in a village on its total school enrollment. The problem of inference is that provision of new all-season roads is non-random. Usually political factors like caste, community and collective actions determine public good provisions in devloping countries (Banerjee, Iyer and Somanathan, 2008). These factors will affect both the provisions of new roads in villages and enrollment of students in school causing them to be spuriously correlated and suffer from political endogenity. The power of collective action at a local level will both determine the timing of completion of the road construction and provision of school facilities and thus enrollment, causing them to be correlated. Thus, an ordinary OLS estimate will be inconsistent. This problem can be overcome by a fuzzy regression discontinuity (FRD) approach. The regression discontinuity technique which compares individuals to the left and right of an exogenous cutoff has gained popularity in recent emperical literature (Angrist and Lavy, 1999; Card, Chetty and Weber, 2006; Carrell, Hoekstra and West, 2010) as it is closer to "gold standard" randomized experiments than other program evaluation methods (Lee and Lemieux, 2009).

The fuzzy regression discontinuity approach exploits the fact that the regressor of interest (new road) is partly determined by a known discontinuous function of an observed covariate ($X_{i,2001}$−the population of the hamlets as recorded in census 2001. It is also the village population for this sample). The other observable characteristics are smooth around this threshold. The fuzzy regression discontinuity can be analyzed in an instrumental variables framework (Lee and Lemieux, 2009; Imbens and Lemieux, 2007). In this case, instrumental variables estimates of equation (7) use discontinuities or nonlinearities in the relationship between new roads and village population; while at the same time, any

[9] The regression discontinuity with perfect compliance i.e. treated=1, others=0 is called sharp RD. Whereas with imperfect compliance, i.e. difference in the probability of treatment between treated and others, is known as fuzzy RD.

other relationship between the population of the village and total enrollment is controlled by including smooth functions of village population as a covariate as in equation (8). Equation (8) represents the first stage where the new roads are being instrumented by the population threshold criteria and equation (9) represents the second stage.

$$(NR)_{i,2009} = \gamma + \delta T + g(X_{i,2001} - 500) + \nu_{i,2009}, \tag{8}$$

$$(y)_{i,2009} = \alpha + \pi(\widehat{NR})_{i,2009} + f(X_{i,2001} - 500) + \eta_{i,2009}. \tag{9}$$

where $T = 1\left[X_{i,2001} \succeq 500\right]$ indicates whether the population of the villages exceeds the eligibility threshold 500 ; $f(X_{i,2001} - 500)$ and $g(X_{i,2001} - 500)$ are respectively control functions[10] and $\eta_{i,2009}$ is a stochastic error term. As the control functions are smooth at the population threshold of 500 whereas the new road is discontinuous, this allows the coefficient π to be identified. In practice, the control function is unknown and has to be approximated by a smooth flexible function, such as a lower order village population polynomial term centered at the threshold 500.

Selection around Discontinuity. A concern for the validity of this technique is that, the eligibility threshold for new roads is common knowledge; this criterion can be manipulated by the local administration those are also aware about the benefits of the roads. Such self-selection or sorting will invalidate the FRD approach estimate π as there will be discontinuous differences in the village's characteristics to the left and right of the cutoff. Lee and Lemieux (2009), prescribe two checks to validate the crucial assumption of absence of self-selection or sorting around the threshold.

First, the density of the population of the villages should be smooth around the 500 threshold. A test outlined in McCrary (2008) is used to test this assumption. As is evident from Figure 2, there is an absence of any jump at the 500 person threshold. Also the log difference in the density around the threshold is statistically insignificant (refer to Appendix 1). However, McCrary cautions that this test can only be useful for discerning manipulation provided manipulation is monotonic. It is reasonable to assume that manipulation would occur only in the upward direction in this case as none of the villages that are legitimately eligible would like to drop out. Further, this is a rare scenario where for the entire sample the true population as collected and reported by the Census 2001 and the administrative reported village population information are both simultaneously available. This makes it possible to control for manipulation additionally. In the entire sample the true population as reported in the Census 2001 has been considered (refer Data Appendix -Step 6).

[10] The control functions are smooth functions of village population controlling for any relationship between population and enrollment of the village. It is depicted as $E(\varepsilon|X_{2001}) = f(X_{2001})$; $\varepsilon = error$

Second, Lee and Lemieux (2009) pointed out that the observable baseline covariates should trend smoothly around the given threshold of 500. As a second check, I test some of the baseline covariates along with the total enrollment in academic year 2002 for discontinuity. A requirement of the regression discontinuity technique is that the baseline covariates have to be assigned before the Census 2001 village population information has been collected. In order to fulfill this requirement, I was limited to consider only those covariates which by construction had been assigned before 2001. As is evident from Table 2 there are negligible differences in the number of primary schools, middle schools, bank facilities and electric facilities between eligible and ineligible villages around the 500 person threshold. Although the eligible villages are on average located 2 km further to the interior than the ineligible villages, the difference in the distance to the nearest town between the eligible and ineligible villages is statistically insignificant. This indicates that there is no systematic bias around the 500 person threshold. This lends confidence that the estimate of π should be purged of any endogeneity.

4.1 Data Characteristic: PMGSY, Census Village Directory and NUEPA

The data used in this paper has been drawn from three main sources: the primary data has been constructed from the assorted reports of Pradhan Mantri Gram Sadak Yojana (PMGSY), the 2001 Census Village Directory and the School Report Cards of National University of Educational Planning and Administration (NUEPA) (refer to data appendix for details on the data construction method and definitions of variables). All three sources provide information at the village level.

The PMGSY provides information on the population size and connectivity status of hamlets which are eligible by distance to get a new all-season road. It also provides their respective census villages, new roads considered for construction, their progress and completion status (refer data appendix).

The 2001 Census Village Directory provides the population demographic information on all the villages in India collected as a part of the decennial census between 9th February, 2001 and 28th February, 2001. It also provides information on different facilities available in these villages like primary schools, middle schools, banking facilities and electricity.

The School Report Cards published by NUEPA provide comprehensive information on enrollment for more than 1.3 million primary and upper-primary schools located across India, for academic year[11] 2002-2009. In addition to total enrollment NUEPA also collects disaggregated information enrollment based by gender, age, grade and caste. The same

[11] The academic year is between September to August.

database is used to derive official enrollment statistics in India (DISE)[12]. The disaggregated information of enrollment based on age and caste, is available only for academic year 2005-2009.

I utilize a subset of the original data. First, the entire analysis is conducted for all unconnected villages, which otherwise satisfy the distance criteria and have a population between 350 and 650. Second, to avoid self-selection problems in this study only those states were considered that individually satisfied the McCrary test (2008) limiting the sample to four states (Rajasthan, Madhya Pradesh, Andhra Pradesh and Kerala). Third, only those villages are considered that are simultaneously eligible by distance and also have at least a primary or middle school[13]. Fourth, only those villages are considered where the entire village population[14] is eligible for the new road. Lastly, to avoid results being driven by changes in sample composition only those villages are considered in the analysis that are tracked consecutively for 8 years, resulting in 3326 villages for each cross-section from 2002-2009.

Table1 summarizes the total enrollment, completed new roads and other facilities like the number of primary and middle schools.for the entire sample of villages considered in this study. Although, the average total enrollment in academic year 2009 is 98 students, the sample exhibits a wide dispersion indicating very small and large schools. Only seventeen percent of the entire sample of villages has a new road as of 2009. However, half of them actually satisfy the population criteria; fifty percent of the villages that were eligible for an all-season road had construction completed as of 2009. The majority of the construction was completed by 2007[15]. Almost every village had a primary school (grade 1-grade 5), and very few had a middle school (grade 6-grade 8) in 2001, when the decennial census survey was conducted[16]. This indicates that for higher education a student has to travel outside the village, validating the relevance of the issue for this study. The spatial dispersion of villages is very high, some being close to the town while others are located further inland. The average village population is approximately near the 500 threshold. Therefore, the entire sample is equally divided on either side of the threshold and there is no systematic bias around the threshold as discussed previously.

[12] This figures is suspected to suffer from upward bias and manipulation (Dre‘ze and Kingdon,1998) as the information is based on a self-reported survey of schools. This tendency is curbed to a certain extent by conducting a cross-verification of school's information of five percentage sample drawn randomly.

[13] Ideally one would like to consider all the schools which are within 3 km and 5 km radius of the eligible villages. But due to data limitation this strategy could not adopted.

[14] The PMGSY rule considers the hamlet population. In order to avoid cases eligibility occurring due to distribution of hamlets I restrict only to those cases where the hamlet population is equivalent to the entire village population.

[15] The majority of construction was completed in Rajasthan for the entire sample.

[16] In 2002, as a part of a new scheme of Sarva Shiksha Abhiyan (SSA) new schools were built in villages which previously did not have schools. In this sample more schools were built on average in the ineligible villages.

5 Effects of Improved Access on School Enrollment

5.1 Basic Results

This section presents results on the effect of new all weather roads on school enrollment. I begin with a graphic overview and then provide a numerical estimate. The core analysis draws on the total enrollment and on disintegrated information of enrollment based on gender for the academic years 2002 and 2009. However, for the entire analysis only the latest cross-section of 2009 is used and 2002 only serves as a base year.

5.1.1 Graphical Analysis

Figure 3 depicts the PMGSY rule stated in equation (1) that governs the provision of new all-season roads. The vertical axis measures the probability[17] of a new all-weather road in a village that was previously unconnected and the horizontal axis measures the population [18] of all villages in the sample that are otherwise eligible by distance. For visual reference, I superimpose a local polynomial regression model fit separately to points on the right and left of the 500 eligibility threshold. Although there can be other factors affecting a village obtaining a new road in India, there is a high correlation between the PMGSY rule and the new roads. There is a significant jump in the probability of gaining a new road for villages that have crossed the 500 person threshold and for those which have failed to do so. The villages that marginally satisfy the threshold have an approximately 20 percent higher probability of getting a new road than those that marginally fail.

Figure 4 plots the mean total enrollment for all grades versus the village population. For visual reference, I superimpose a local polynomial regression model fit separately to points on the right and left of the 500 person eligibility threshold. We observe a discrete jump in the mean total enrollment at the threshold. However, as is evident from the figure the (mean) total enrollment suffers from noise or high fluctuations. The increase in total enrollment at the threshold is not discernible by the naked eye. A concern in this regard is that the increase in total enrollment is driven by a higher base enrollment prior to the construction of new roads.

Figure 5 plots the trends in total enrollment of eligible (exceeds the 500 person population threshold) and ineligible (falls below the 500 population threshold) villages over the time. As is evident from figure 6, the eligible villages as of 2002 had marginally lower total enrollment than that of the ineligible villages. However, as of 2009 there is a significant increase in total enrollment. Further, in general the trends for total enrollment in the economy are similar indicating that the other factors are smooth at the 500 threshold.

[17] Average number of new roads completed as on 2009.

[18] The population of the village as recorded in the Census 2001.

5.1.2 Empirical Analysis

To formally identify the impacts of new roads on total enrollment, I estimate equation (9) using the fuzzy regression discontinuity approach. The results of the analysis are reported in Tables 3-5.

Table 3 provides the first stage results from equation (8). There are five columns for different discontinuity samples. The dependent variable in each regression is the new road, which is instrumented by population at the 500 threshold point. The additional regressors are distance to town and the number of primary, middle, secondary and higher secondary schools in the village. The standard errors have been clustered both at the district and block level. The results across the five samples display a robust pattern. The villages that are marginally above the 500 population threshold have an additional twenty five percent point probability of obtaining the new road, compared to those that are marginally below the threshold. Since the standard errors are clustered and therefore non-i.i.d,Kleibergen-Paap rk statistic[19] is calculated. A statistic above 10 indicates a strong instrument.

Table 4 provides second stage results from equation (9). There are nine columns, showing the effects of new road on three alternative measures of total school enrollment[20] (all, boys and girls) in academic year 2009-10 using three different specifications. Specification 1 of Table 4 includes only the linear polynomial[21] and district fixed effects. In Specification 2, base line covariates are added: distance to the town, number of primary, middle, secondary and higher secondary schools in the village, bank, electricity and newspapers facilities. In Specification 3, the total enrollment as of base year 2002 is included. Irrespective of the specification a robust pattern is observed. Better access to schools by new all season roads leads to an additional total enrollment of approximately 29 students in the schools per eligible village. This effect is equally contributed by both boys and girls: Per eligible village there was an additional total school enrollment of 14 students of each gender[22].

The empirical estimates validate the fact that improved access to schools through the construction of new roads has increased students' enrollment in schools in rural India. Further, the benefits of better access to schools within villages can be reaped equivalently across different genders. This is an important finding because usually in a country like India there are social taboos for girls travelling far, since it may impose personal safety

[19] When the errors are non- i.i.d the critical values complied by Stock and Yogo (2005) can not be used, then the Kleibergen-Paap rk statistic is calculated (Baum, Schaffer and Stillman ,2007).

[20] The total enrollment for all, boys and girls are individually winzorized at one percent to adjust for extreme outliers at both the tails of the distribution.

[21] The order of the polynomial is chosen by the minimum AIC. $AIC = N \ln \widehat{\sigma}^2 + 2p, where\ \widehat{\sigma}^2 = MSE$ & $p = \#\ of\ regressors$

[22] The discrepancy in increase of all students with that addition of each gender individually is caused by winzorization.

issues for them (Holmes, 2003). This may impose a hindrance for the higher education of girls compared to boys when the schools are located outside villages. This implies that girls' enrollment for schools outside the village would increase more than the boys, but due to data restriction this phenomena cannot be observed here.

A concern for this analysis is that it could be sensitive to the size of the discontinuity sample whereas it should be robust even if half of the sample has been discarded (Lee and Lemieux, 2009). Table 5 provides estimates of Specification 3 for different discontinuity samples. As is evident, irrespective of the discontinuity sample, the impact of roads on total school enrollment is robust. The discontinuity sample +/- 80 represents the case when only half of the entire sample has been used. Using the most conservative measure, I find that the villages which obtain a new all-weather road experience an additional total school enrollment of 21 students in the academic year 2009-10. This is equivalent to a twenty two percent increase in total enrollment compared to the mean enrollment per village in that year. Further, goodness of fit of the model is also tested. The goodness of the fit model compares the parametric model with that of a general non-parametric alternative (unrestricted graph). In this test a set of bin dummies is added to the polynomial regression and the joint significance of the bin dummies is tested (Lee and Lemieux, 2009). The square bracket represents the P-value for the joint significance tests. As one moves across columns left (full sample) to right (very restricted sample), non-parametric becomes a better alternative compared to the parametric model. Also, the optimal order of polynomial is tested for different samples. The optimal order of the polynomial was chosen by the minimum AIC after comparing up to fourth order polynomials. For all the alternatives the optimal order of polynomials is linear.

Another concern is that some other factors like mid day meal schemes[23] or new schools opening could also be potentially driving this result. First, except for the unaided private schools (which is a small percentage of the entire composition) all schools in this analysis are eligible for the mid day meal scheme (refer Appendix 2). As the mid day meal is smooth at the 500 threshold point or in other words there is no kink in the mid day meal scheme for the 500 population villages, therefore the trend should be similar for schools in villages both marginally above and below the 500 threshold point. Secondly, new schools that have opened since the 2001Census was conducted were on average in villages with marginally less than 500 inhabitants. Therefore, we can reasonably rule out these factors as contributing to the increase in school enrollment.

Up to this point in the analysis, the inherent assumption was that the effect of better access to school on students' enrollment is homogenous for all age cohorts and different sections of the economy. However, these are very restrictive assumptions I relax these

[23] Mid-Day Meal (MDM) scheme was introduced in Indian schools to help students' nutrition and reduce the dropout rate.

assumptions in the following sections.

5.2 Impacts of New Roads on Students' Enrollment across Different Age Cohorts

In this section, I test the first hypothesis of the theoretical model that states in an economy with child labour[24] prevalence as the age increases the impact of new roads on higher cohorts' total school enrollment will decrease. The lower the outside wage earnings potentiality is, the higher is the incentive to stay in school. Therefore in an economy with child labour opportunity, the lower age group who are not physically fit to work always participate more than a higher age group who are physically able to work. Incorporating the child labour assumption is crucial as incidence of child labour is high in three of the four states considered in this study (refer to Appendix 3). Further, the children in the age group ten to fourteen the most vulnerable to child labour.

Table 6, provides the main results for various versions of Specification 3 in equation (9). There are four columns, with each column representing a cohort. Officially the age group five to fourteen is considered to be vulnerable to child labour. Depending on their degree of vulnerability to child labour I divide the entire sample into four cohorts. The four cohorts are less than five, five to nine years, ten to fourteen years and above 15 years. As one moves across the columns (1) to column (4) i.e. from cohorts less than five to more than fifteen, the impact of the new roads decreases. Statistical significance is observed for only the cohort five to nine, but the mean enrollment in some other cohorts is itself small. Although, the magnitude of the coefficient is small for certain cohorts, the estimate in terms of the percentage change compared to the mean enrollment for that cohort is large. Therefore, the cohorts' partial elasticity i.e. the percentage change in total enrollment in a cohort with a new road in the village compared to the mean enrollment is more meaningful for the analysis.

Consistent with the hypothesis of monotonically decreasing impact of new roads on total enrollment with an increase in age, I find the magnitude of partial elasticity decreases as the age of the cohort increases. As one moves from column (1) to column (4) respectively, the percentage change in total enrollment with a new road changes direction from being positive to negative. The largest impact of the new roads occurs on the two extreme cohorts, that is, less than five and more than fifteen age groups. For the five to nine and ten to fourteen age groups, the percentage change in total enrollment is thirty and nineteen respectively. This indicates the fact that with an improved connectivity

[24] According to Child Labour (Prohibition and Regulation) Act 1986 in India, a child less than 14 years of age employment is prohibited from employment in certain occupations. Further according to the International Labour Organization (ILO) all forms of work by children under the age of 12 should be considered as child labour.

both positive and negative effects on total enrollment exist as both the access to school and potential for work in the outside job market is increasing simultaneously.

The empirical estimates validate the fact that there are both intended and unintended consequences of an improved connectivity through the development of new roads on students' enrollment in schools. With new roads the students' access to schools is improved causing to an increase in enrollment. The flip side of the better connectivity is that it simultaneously increases the availability of outside job options for the students. However, in order to reap the benefit of outside job options, physical fitness is required. Henceforth, the higher age cohort students who are physically fit to work compared to the lower age cohorts have less incentive to participate in schools.

Until now in this analysis it has been assumed that the benefit of roads is homogenous across students from different social background. This assumption is actually invalid for India where access to schools and the potentiality of reaping future benefits from higher schooling is contingent on the social background of students. In the following section I relax this restrictive assumption.

5.3 Impacts of New Roads on Students' Enrollment across Different Social Background

In this section, I test the second hypothesis of the theoretical model which states that effects of new roads on students' enrollment is ambiguous and contingent on their social background. Investment in schooling is dependent on future benefits of schooling and the travel cost involved in acquiring schooling. Usually in India schools are spatially clustered in areas dominated with high castes[25]. In many states the backward castes (especially schedule tribes) are interiorly located or reside outside the vicinity of villages. Banerjee and Somanathan (2007), use data for the Indian parliamentary constituencies and find that in the early 1970s the population share of Brahmans in a constituency is positively correlated with access to primary, middle and secondary schools, to post offices and to piped water. Hence, even within villages the distance to schools varies for different castes. The backward caste students' access to schools will improve from the development of new roads as for them travel cost is higher, newroads therefore positively influence their participation in schooling. On the other hand, social mobility is an important determinant in predicting future earnings from higher levels of schooling. The backward caste have more restrictions on social mobility; thus they can reap less benefit from future earnings from schooling. This will negatively impact current investment decisions for students of backward castes; they have more incentive to drop out from the schools. Accordingly the impact of better connectivity on enrollment will be dependent which of the two effects

[25] In India an example of the high castes in the Hindu society are Brahmans.

dominate, and will be heterogeneous across castes.

Table 7 provides the main results for various versions of Specification 3 in equation (9). There are two columns, with each column representing different castes. The column (1) represents general castes and column (2) represents backward castes. The backward castes are comprised of schedule castes, schedule tribe and other backward class. The general castes are comprised of other castes beside backward castes. For both the castes the students' enrollment in schools increases with the development of road and improvement in connectivity. While the magnitude of the coefficient terms is only considered then with new roads the enrollment of students in schools from backward castes are higher compared to the general castes, but in terms of partial elasticity the results reverse. The percentage change in total enrollment compared to the mean for general caste students is more than hundred percent, whereas for the backward caste students it is only twenty one percent.

The empirical estimates indicate the fact that irrespective of caste, better access to schools increases students' enrollment in schools. However, the response of enrollment to an improvement in connectivity is much higher for general caste students compared to the backward caste students. This indicates the fact that restrictions in social mobility are an important factor in determining investment decisions of schooling. The general (backward) castes those face low (high) restrictions in social mobility and can reap high (low) future earnings from an investment in schooling. The social mobility restrictions of the different castes influence their current participation decisions on schooling.

5.4 Robustness Check: Placebo Test

A crucial assumption internal in this analysis is that the other factors affecting total enrollment are uncorrelated with the village population at the 500 threshold. If the other factors are correlated then the increase in total enrollment due to new-roads would be spuriously correlated. As a possible mechanism to discern the spurious from the casual effect I create a simulated threshold point at the left and right medians 422 and 573 threshold respectively. Lee and Lemieux, 2009 suggests that the power of the test is highest at the medians. If the effect is casual then at the left and right median threshold points there should not be any increase in total enrolment.

In Figure 6 the top panel represents the simulated threshold at the left median point 422 and the bottom panel represents the simulated threshold at the right median point 573. As is evident from the top panel, there is an increase in the total enrollment at the left median point. In the bottom panel we observe that at the right median point the increase in total enrollment is absent. This indicates that the increase in total enrollment at the left median point is due to the high fluctuations in the total enrollment caused by a noisy data.

This gives confidence to the fact that the increase in total enrollment at the 500 threshold point is a casual effect.

6 Conclusion

This paper presents a variety of instrumental variable estimates of the effect of better roads on students' enrollment in schools in rural India. Instrumental variable estimates constructed by using population rules as instruments for new all-season roads show a positive association between new all-season roads and students' enrollment. The most conservative estimates show that an improved access to school by better roads increased school enrollment by 22 percentage in 2009. These effects vary across different age cohorts. The effect is largest for students of younger cohorts. Further, I find that the enrollment of students is heterogeneous across different social background. The results indicate that the restriction in social mobility that determines their future earnings is an important factor for their participation in schooling. This paper also extends the human capital investment model and introduces cost of travel time to motivate an empirical analysis. The model provides some hypotheses, for virtually all of which I find empirical support.

The effects are larger than those that are reported by Duflo (2001) Jalan and Glinskaya (2003). However, for poorer sections of the economy the authors found a large impact on students' enrollment in schools with a better access. The findings reported here are important because they show that a large government intervention has been effective in increasing education in India. In India where the low caste resides in the vicinity of the villages the physical distance imposes a major hurdle to participation in school. Thus, these sections would especially benefit from an improved access to schools. An improved access is especially beneficial for girl children for whom travelling a long distance becomes additional barrier to participation in schools. In many countries there are social taboos for unmarried young girls for travelling far from home, but no such taboos exists for boys (Holmes, 2003). Travelling a long distance by foot may pose a personal safety issue for young girls but not for boys. Distance increases the opportunity cost of schooling for girls and will lead to an increase in the enrollment gap between boys and girls.

It is worth considering that the results for India are likely to be relevant for other developing countries as well. The schools in India are located at comparable distances compared with those in Bangladesh and Philippines (see Filmer, 2007). Culturally, India and Bangladesh are more similar than that of Phillipines. So, the results presented here may be showing evidence of an effect of an improved access to schools on students' enrollment for most developing countries.

There is usually a tradeoff between quantity and quality. This analysis have concentrated towards the quantity of education and left aside the quality aspect. Although, it cannot be tested here there is likelihood that this would also increase the quality of education. In developing countries the quality of education is highly sensitive to the time teachers allot to task (Epstein and Karwait, 1983). With an improved connectivity the perceived threat of monitoring increases subsequently reducing the absence rate (Kremer et al., 2004). A reduction in the teacher's absence rate will enhance the quality of education.

References

[1] Angrist, Joshua, Lavy, Victor, 1999. Using Maimonides'Rule To Estimate The Effect of Class Size on Scholastic Achievement. Quarterly Journal of Economics, 114 (2) , 533-575.

[2] Assorted Reports of Pradhan Mantri Gram Sadak Yojana, India.

[3] Banerjee, Abhijit, Somanathan, Rohini, 2007. The Political Economy of Public Goods : Some Evidence from India. Journal of Development Economics, Vol. 82 (2), 287-314.

[4] Banerjee, Abhijit, Iyer, Lakshmi, Somanathan, Rohini, 2008. Public Action for Public Goods. Handbook of Development Economics, Vol. 4,Chapter 49, 3119-3154.

[5] Basu, Kaushik, 1999. Child Labour: Cause, Consequence, and Cure, with Remarks on International Labor Standards. Journal of Economic Literature, Vol. 37, 1083-1119.

[6] Baum, Christopher F., Schaffer, Mark E., Stillman, Steven, 2007. Enhanced Routines for Instrumental Variables\Generalized Method of Moments Estimation and Testing. The Stata Journal , Vol. 7 (4), 465-506.

[7] Bell, Clive, 2010. Goods, Education and Health: A Combined Model for Evaluating PMGSY. University of Heidelberg, Germany.

[8] Carrell, Scott E., Hoekstra, Mark, West, James E., 2010. Does Drinking Impair College Performance? Evidence from a Regression Discontinuity Approach. NBER Working Paper No. 16330.

[9] Card, David, Chetty, Raj, Weber, Andrea., 2007. Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market. The Quarterly Journal of Economics, Vol. 122 (4), 1511-1560.

[10] Census of India Village Directory,2001.

[11] Duflo, Esther, 2001. Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. The American Economic Review, Vol 91 (4), 795-813.

[12] Dre'ze, Jean, Kingdon, Geeta Gandhi, 1998. Biases in education statistics. The Hindu, March 6.

[13] Epstein, J, Kavweit, N., 1983. Friends in School: Patterns of Selection and Influence in Secondary Schools. Academic Press, New York.

[14] Filmer, Deon, 2007. If you build it , will they come ? School availability and school enrollment in 21 poor countries. Journal of Development Studies,Vol. 43 (5),.901-928.

[15] Foster, A.D., Rosenzweig, M. R., 1996. Technical Change and Human Capital Returns and Investments: Evidence from the Green Revolution. American Economic Review, Vol. 86 (4), 931-953.

[16] Handa, S., 2002. Raising Primary School Enrollment in Developing Countries: The Relative Importance of Supply and Demand. Journal of Development Economics, Vol. 69 (1), 103–128.

[17] Holmes,J., 2003. Measuring the determinants of school competition in Pakistan: Analysis of censoring and selection bias. Economics of Education Review, 22, 249-264.

[18] http://pmgsy.nic.in/

[19] http://pmgsy.nic.in/downloads/esmf.pdf

[20] Imbens, Guido, Lemieux, Thomas, 2008. Regression Discontinuity Designs: A Guide to Practice. Journal of Econometrics, Vol. 142 (2), 615-635.

[21] Jalan, J., Glinskaya, E., 2003. Improving Primary School Education in India: An Impact Assessment of DPEP-Phase I. Unpublished Document, Indian Statistical Institute (New Delhi) and The World Bank.

[22] Jacoby, H. G., 2000. Access to Markets and the Benefits of Rural Roads. Economic Journal, 110 (465), 713–737.

[23] Jacoby, H. G., Minten, Bart, 2008. On Measuring the Benefits of Lower Transport Costs. World Bank Working Paper, 4484.

[24] McCrary, Justin, 2008. Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. Journal of Econometrics. Vol. 142 (2), 698-714.

[25] Mukul, Akshaya, 2009. Dalit kids cannot use school loo but have to clean them. The Times of India, June 25

[26] Lee, David S., Lemieux, Thomas, 2009. Regression Discontinuity Designs in Economics. Journal of Economic Literature, American Economic Association, Vol. 48 (2), 281-355.

[27] UNICEF 2006. Social exclusion of Scheduled Caste Children from Primary Education in India, UNICEF, India

[28] Ward, Michael (2007). Rural Education. India Infrastructure Report.

[29] Wills, Robert J, 1986. Wage Determinants: A Survey and Reintterpretation of Human Capital Earnings Functions. Handbook of Labor Economics, Vol. 1, Chapter 10, 525-602.

A Tables and Figures

Figure 1-McCrary(2008) Test for Manipulation at 500 cutoff

Figure 2-New Roads Construction between Eligible vs Ineligible Villages

Figure 3-Total Enrollment (All) between Eligible vs Ineligible Villages.

Figure 4-Trends in Total Enrollment across years for Eligible vs Ineligible Villages.

Figure 5-Placebo Test (Top Panel-Left Median),(Bottom Panel-Right Median)

Table 1-Descreptive Statistics (Full Sample)

Variables				Quantiles				
	N	Mean	S.D.	0.10	0.25	0.50	0.75	0.90
Enrollment	3326	97.44	64.48	40	55	82	119	171
Completed New Roads	3326	0.17	0.38	0	0	0	0	1
Year of Completion	585	2007	0.78	2005	2007	2007	2007	2007
Population	3326	489	86.39	375	412	486	562	613
Distance to town	3326	27.20	18.72	9	14	22	35	50
Number of Primary Schools	3326	1.0	0.25	1	1	1	1	1
Number of Middle Schools	3326	0.05	0.23	0	0	0	0	0

Variable definitions are as follows: Enrollment=Total enrollment in schools from September 2009-September 2010, Distance to town=Distance of the village from the nearest town in kms, Population=Total Village Population as of 2001 Census.

Table 2-Tests of the Identifying Assumption of the RD Analysis (Full Sample)

Variables	Base Year	Primary	Middle	Distance	Bank	Electricity
New Road	-1.06	-0.02	-0.05	2.03	-0.03	0.07
	(16.06)	(0.05)	(0.08)	(5.17)	(0.02)	(0.09)
Observations	3326	3326	3326	3326	3326	3326

The unit of observation is the village. Each cell represents results for separate regression where a key independent variable is an indicator for a new road. Standard errors are clustered both at district and block level in parentheses. The base year is the Total Enrollment in academic year 2002. Primary (Middle) is the number of schools in villages in 2001. Distance is the distance to the nearest town. Bank and Electricity are indicators for those facilities in the villages in 2001.

Table 3-First Stage Results for Discontinuity Samples

Outcome	New Road				
Discontinuity Sample-\+	(150)	(100)	(80)	(60)	(50)
Discontinuity of Population at 500	0.26**	0.28**	0.28**	0.27**	0.25**
-	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
K-P rk Wald F Statistics	18.45	14.60	12.83	15.97	10.28
Baseline Covariates	Yes	Yes	Yes	Yes	Yes
Base Year	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes
Optimal order of Polynomial	1	1	1	1	1
Population Polynomial	Linear	Linear	Linear	Linear	Linear
Observations	3326	2236	1734	1314	1116

**(5% significance level)*(10 % significance level). The unit of observation is the village. Each cell represents results for separate regression where the dependent variable is a new road and a key independent variable is an indicator for population at 500. Standard errors are clustered both at district and block level in parentheses.

Staiger and Stock(1997), states that the F statistic should be at least 10 for weak identification not to be considered a problem.

Table 4-2SLS Estimates for All grades in 2009 (Full Sample)

Outcome	Total Enrollment			Total Enrollment			Total Enrollment		
	All†	Boys†	Girls†	All†	Boys†	Girls†	All†	Boys†	Girls†
Specification	(1)	(2)	(3)	(4)	(5	(6)	(7)	(8)	(9)
New Road	26.95**	12.97	13.25**	30.31**	15.01*	14.61**	29.95**	14.89*	14.49**
-	(13.38)	(8.12)	(6.22)	(13.51)	(8.09)	(6.38)	(12.25)	(8.00)	(6.39)
Baseline Covariates	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Base Year	No	No	No	No	No	No	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3326	3326	3326	3326	3326	3326	3326	3326	3326
Population Polynomial	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
Discontinuity sample					-/+150 Population				

†The enrollment has been winzorized at 1 percent to take care of extreme values.**(5% significance level)*(10 % significance level). The unit of observation is the village. Each cell represents results for separate regression where the dependent variable is the total enrollment and a key independent variable is an indicator for a new road. Standard errors are clustered both at district and block level in parentheses. The base year is the academic year 2002. Baseline covariates include distance to the town, number of primary, middle, secondary and higher secondary schools in the village. Other covariates are indicators of bank facilities, electricity and newspaper in the village as of year 2001.

Table 5-Regression Discontinuity for Discontinuity Samples

Outcome	Total Enrollment (All)				
Discontinuity Sample-\+	(150)	(100)	(80)	(60)	(50)
New Road	29.74**	23.41*	21.83	25.36	21.12
-	(12.35)	(13.57)	(16.33)	(20.08)	(23.45)
	[0.00]	[0.00]	[0.00]	[0.12]	[0.36]
Partial Elasticity	+0.*30*	+0.*24*	+0.*22*	+0.*26*	+0.*22*
Baseline Covariates	Yes	Yes	Yes	Yes	Yes
Base Year	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes
Optimal order of Polynomial	1	1	1	1	1
Population Polynomial	Linear	Linear	Linear	Linear	Linear
Observations	3326	2236	1734	1314	1116

**(5% significance level)*(10 % significance level). The unit of observation is the village. Each cell represents results for separate regression where the dependent variable is the total enrollment and a key independent variable is an indicator for a new road. Standard errors are clustered both at district and block level in parentheses. Base year is the academic year 2002. P-values from the goodness-of-fit test in square brackets. The goodness-of-fit test is obtained by jointly testing the significance of a set of bin dummies included as additional regressors in the model. The bin width used to construct the bin dummies is 5. The optimal order of the polynomial is chosen using Akaike's criterion.

Table 6-Age -wise Effect of New Road on Total Enrollment (Full Sample)

Outcome	Total Enrollment (All)			
Age-Group(in years)	(less than 5)	(5-9)	(10-14)	(15 above)
New Road	0.84	19.68**	6.35	-0.57
-	(0.79)	(7.25)	(6.32)	(1.09)
Mean of Enrollment	2.07	63.53	33.29	1.05
Partial Elasticity	+0.*42*	+0.*30*	+0.*19*	-0.*53*
Baseline Covariates	Yes	Yes	Yes	Yes
Base Year	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes
Population Polynomial	Linear	Linear	Linear	Linear
Observations	3326	3326	3326	3326

**(5% significance level)*(10 % significance level). The unit of observation is the village. Each cell represents results for separate regression where the dependent variable is the total enrollment and a key independent variable is an indicator for new roads.
Standard errors are clustered both at district and block level in parentheses. The base year is the academic year 2002.
Baseline covariates include distance to the town, number of primary, middle, secondary and higher secondary schools in the village.
Other covariates are indicators of bank facilities, electricity and newspaper in the village as of year 2001.

Table 7-Effect of New Road on Total Enrollment by Caste (Full Sample)

Outcome	Total Enrollment (All)	
Type of Castes	(General)	(Backward)
New Road	9.76*	18.92**
-	(5.44)	(9.38)
Mean of Enrollment	9.75	86.40
Partial Elasticity	+1.*00*	+0.*21*
Baseline Covariates	Yes	Yes
Base Year	Yes	Yes
District Fixed Effects	Yes	Yes
Population Polynomial	Linear	Linear
Observations	3114†	3114†

†Observations are lost due to absence of caste information in the village school.**(5% significance level)*(10 % significance level).

The unit of observation is the village. Each cell represents results for separate regression where the dependent variable is the total enrollment and a key independent variable is an indicator for new roads. Standard errors are clustered both at district and block level in parentheses.

The backward caste comprise of schedule caste, schedule tribe and other backward class. The base year is the academic year 2002.

Baseline covariates include distance to the town,number of primary, middle,secondary and higher secondary schools in the village.

Other covariates are indicators of bank facilities,electricity and newspaper in the village as of year 2001.

B Theory Appendix

The representative individual's problem is the following

$$
\begin{aligned}
Max\ V(s,r) &= \frac{\{y(s,\tau)-\}\,e^{-rs}\left(1-e^{-rn}\right)}{r} - .C(s,\tau)w.r.t.\text{"}s\text{"},\\
\frac{dV(s,r)}{ds} &= \frac{\left(1-e^{-rn}\right)\left[e^{-rs}y_s - re^{-rs}y\right]}{r} - .C_s = 0,\\
F &= \frac{\alpha\left[e^{-rs}y_s - re^{-rs}y\right]}{r} - .C_s = 0,\\
Now, \frac{ds}{d\tau} &= -\frac{\left[\frac{\delta F}{\delta \tau}\right]}{\frac{\delta F}{\delta s}}\,\mathrm{Im}\,plicit\ Theorem\\
\frac{\delta F}{\delta \tau} &= \frac{e^{-rs}\left(1-e^{-rn}\right)\left[y_{s\tau} - ry_\tau\right]}{r} - C_{s\tau},\\
\frac{\delta F}{\delta s} &= \frac{e^{-rs}\left(1-e^{-rn}\right)\left[y_{ss} - 2ry_s + ry^2\right]}{r} - C_{ss},\\
\frac{ds}{d\tau} &= -\frac{e^{-rs}\left(1-e^{-rn}\right)\left[y_{s\tau} - ry_\tau\right] - rC_{s\tau}}{e^{-rs}\left(1-e^{-rn}\right)\left[y_{ss} - 2ry_s + ry^2\right] - rC_{ss}},\\
\left[y_{ss} - 2ry_s + ry^2\right] &< 0(by\ construction),\\
\frac{ds}{d\tau} &\lesseqqgtr 0 \text{ if } \left[y_{s\tau} - ry_\tau\right] \lesseqqgtr 0.
\end{aligned}
$$

C Data Appendix

Glossary of Pradhan Mantri Grammen Sadak Yojna (PMGSY)

All-weather road: A road which is negotiable during all weathers, except at major river crossings. This implies that the road-bed is drained effectively by adequate cross-drainage structures such as culverts, minor bridges and causeways. Interruptions to traffic as per permitted frequency and duration are, however, allowed. The pavement should be negotiable during all-weather, but this does not necessarily imply that it should be paved or surfaced or black-topped.

Black-Topped Road (BTR): A road provided with bituminous surfacing.

Core Network: Network that is essential to provide basic access to each habitation. It can also be defined as the network of all the Rural Roads that provide Basic access to all the Habitations. Basic access is defined as the single all-weather road connectivity to each Habitation. As already indicated, the effort under the PMGSY is to provide single all-weather road connectivity to each Habitation by way of connecting it to another Habitation having all-weather connectivity or to an all-weather road, in such a way that there is access to, inter alia, Market Centers.

Gravel Road (GR): A road constructed using well compacted crushed rock or gravel material, which is fairly resilient and does not become slippery when wet.

Habitation: A cluster of population, living in an area, the location of which does not change over time. Desam, Dhanis, Tolas, Majras, hamlets etc. are commonly used terminology to describe the Habitations. A Revenue village/ Gram Panchayat may comprise of several Habitations.

New Connectivity: The construction of roads on the existing alignments from earth-work stage.

Paved Road (PR): A road provided with a hard pavement course, which should be at least a water-bound-macadam layer. A paved road need not necessarily be surfaced or black-topped.

Unconnected Habitation: A habitation with a population of more than 500 persons and located at a distance of at least 500 meters (0.5 km) (from an all-weather road or a connected village/habitation).

Unpaved Road (UPR):A road not having a hard pavement course (which should be at least a water-bound-macadam layer). Thus, earthen road and gravel road will be unpaved roads.

Village Roads (VR): These are roads connecting villages/Habitation or groups of Habitation with each other and to the nearest road of a higher category.

Water Bound Macadam (WBM): This is the road layer made of crushed or broken aggregate mechanically interlocked by rolling and the voids filled with screening and binding material with the assistance of water.

Primary Data Construction Process from PMGSY, Concording to Census 2001 Village Directory and Concording to School Report Cards

Step1: The Core Network (CN) provides information on all eligible habitations with a population over 100 which satisfy the distance criteria. Each habitation in CN has been assigned a unique 10 digit code with first two -state, next two-district, next two-block and last four habitations by PMGSY. This provides the population information (according to 2001 Census) ,connectivity status (unconnected for new connectivity) of these habitations and 2001 Census villages from which they were drawn.

Step 2: The Completed Road and Proposed Roads status is provided in the District Profile, which has a unique project or road code The Completed Road information does not provide any habitation information but the proposed road information does. So merging the Completed Roads with Proposed roads would provide the districts, blocks and habitations name where the roads have been completed.

Step 3: The PMGSY Block and District code has to be concord with that from the census file to obtain their respective 2001 Census codes and names.

Step 4: The proposed road whose subset has been completed should be merged with the CN for a 1:1 mapping of the habitations and their completed roads status (or new roads status). Further, this allows forming the unique 16 digit village 2001 Census codes to which this habitation belongs. This allows the use of information from the census village directory.

Step 5: The new roads dummy variable has been constructed which assigns 1 if the road construction of a habitation is completed and zero if either under construction or not proposed.

Step 6: To rectify for manipulation of the habitation population information the respective census 2001 village total population is used as upper bound.

Step 7: The School Report Cards provides the unique 9-digit code for the village of whose 4 digit are the 2001 Census state and district codes. They also come with the 2001 Census block name and village name. This has to be manually matched with the 2001 Census using the state, district, block and village names. Thereafter the 16 digit 2001 Census codes have been formed for the respective 9-digit village code of the School report card. Also, every school has a unique 11 digit code of whose first 9-digits signifies the village code, thus these helps to concord the schools to their respective 2001 Census villages.

Step 8: Finally, the two data sets have to be merged on the basis of 16 digit 2001 Census code to get a master data whose unit of observation is the village with information on population, the new roads completion status, school enrollment and other village demographics.

Appendix1-Log Discontinuity Estimates

Outcome	Density Estimate of Population of Villages
Discontinuity Sample-\+	(150)
Discontinuity of Population at 500	-0.01
-	(0.13)
Observations	3326

Standard errors in parentheses.

Appendix 2-Mid-Day Meal Scheme Across Years.

Appendix3-Comparison of Child Labour for Sample States with National Average

	Child Workforce Participation Rates		
States	(5-9)	(10-14)	(5-14)
Andhra Pradesh	0.56	12.38	6.61
Kerala	0.00	0.39	0.20
Madhya Pradesh	0.14	5.74	2.82
Rajasthan	0.41	9.42	4.86
India	**0.26**	**6.38**	**3.33**

Source:Derived from Unit Level Records of NSS, 2004-05.