Beggar-thy-neighbourhood: Spatial Externalities in Access to Private Schools - Evidence from India

Aditi Acharya*

Abstract

Access to good quality schools in a neighbourhood is a strong correlate of long run income dynamics (Chetty et al 2014). But do more endowed neighbourhoods lower the probability of poorer neighbourhoods nearby having access to these good schools, thereby affecting their human capital accumulation? I use the context of a rural state of India where villages (that proxy rural neighbourhoods) in a "schooling market" compete for private middle schools (which are considered by people to be of higher quality than public schools) to locate them within the village locality. The first contribution of this paper is to construct a geographically demarcated "schooling market" -since private schools may not cater to one village. Using granular spatial location of all villages in the state and machine learning techniques, the paper constructs schooling markets: villages are spatially close to each other within each market but far from villages in other markets. The average geographical size of the constructed markets is consistent with the reported distance travelled by private upper primary (middle) school students, as recorded in education surveys. As a second exercise, this paper estimates the impact of the average incomes of villages in a schooling market on the location of private middle schools. Using decadal occurrence of exogenous weather shocks as instruments for village incomes, I find that a village with a higher income is more likely to have a private middle school but if a village is also surrounded by rich villages, it reduces the likelihood of having a private school. Thus, to have a shot at landing a private middle school, not only does a village have to be rich, it has to be rich relative to its neighbourhood. The paper also provides some suggestive evidence that when schooling markets have such unequally distributed neighbourhoods, the average human capital of school-going-age children living in such markets suffers.

*I would like to thank my thesis advisor, Prof. Abhiroop Mukhopadhyay, for his guidance, encouragement and all the long and insightful discussions. I am also grateful to Prof. Tridip Ray, Prof. Debasis Mishra and Dr. David Evans for their comments and helpful suggestions.

1 Introduction

The relationship between human capital, development and growth has been extensively researched in economics (Barro 1991) and (Mankiw et al 1992). More recent literature focuses not only on the aggregate human capital but also on the geography of its production and its implication for income and intergenerational mobility (Chetty et al 2014). This literature places the neighbourhood as a focal geographical unit giving schools a central role; neighbourhoods with better schools tend to have better prospects for children who exhibit higher upward mobility. In the context of the US, neighbourhoods are also racially segregated: thus ethnicity is tied to neighbourhoods and explains why some communities are stuck in low human capital-low income traps. An aspect that has received lesser attention in the literature, however, is whether certain endowed neighbourhoods crowd out access to human capital resources- schools, teachers and other supply side inputs- in poorer neighbourhoods nearby and affect their human capital with its ensuing effects on long run outcomes. I investigate this issue in the context of a state in rural India where villages form historically determined neighbourhoods and exhibit variation in ethnicity and income. In particular, I ask if greater income of a village increases the chance that a private (middle) school-often perceived to be of higher quality than the existing public school- locates within its locality but lowers the probability that such schools open up in relatively poorer villages nearby and whether this leads to a lowering of human capital accumulation, overall.

While the question this paper asks is of more general relevance to understand spatial externalities that some neighbourhoods pose on others, the specific context of private schooling and its access is especially important to developing countries-especially in South Asia, where the public schooling model has met with poor success. The demand for private schools have been driven by the failure of public schools (Muralidharan and Kremer 2006). Therefore, nearly 50% of all students in India are enrolled today in half a million privately managed schools across the country. Private schools have experienced massive growth in the last two decades and today serve almost 90 million students (U-DISE 2019). Rural areas have seen an analogous boom in private schools- from 4% of total schools in 1993 to 27% in 2018. Thus, private schooling is the new focal point of the national education policies[1] and diktats from international think-tanks call for a more private sector driven approach to meet schooling needs.[2]

[1]The New Education Policy, 2021 emphasises a big role for private schools to improve learning outcomes in India.

[2]Central Square Foundation: "State of the Sector Report on Private Schools in India" (2020)

The concerns that low-income households have poor access to quality schools and that poor neighbourhoods have worse quality of schools are well established in the literature and are equally true for both developed and developing countries (McCoy et al 2016). However, there has been less discussion on the externalities that some neighbourhoods impose on others. This paper contributes to the literature by focussing on spatial externaliies- supply of schooling inputs can be crowded out by geographically contiguous neighbourhoods that are relatively richer. In the context of the US, such a result would have an immediate implication on human capital-schools typically mimic the ethnic and economic composition of neighbourhoods-thus pointing out to limited movement of students across neighbourhoods.[3] However, in the context of rural private schools that the paper analyses, this implication is not obvious. Private upper primary (middle) schools may not open up for one village-hence one needs to think about what is the geographical size of the market. Further, it is not obvious that if a village does not have a private school, children living in such a village cannot go to the neighbouring village where the private school has opened up. In this paper, I address these issues head-on.[4]

To test the hypothesis about spatial externality, the paper proceeds in two steps. First, it uses data on 21,346 villages in the state of Rajasthan (a western arid state of India with the largest area) for whom geocodes are available and there is no missing information on schools. Using a machine learning technique: K-means clustering, it partitions villages into school markets, such that within each schooling market, the villages are geographically close to each other, but are far away from the the next schooling market. Such methods do not artificially impose distance cut-offs to demarcate markets, instead the total number of clusters are chosen and the algorithm partitions villages into distinct sets in an iterative process so as to minimise the within cluster distance between villages and maximises the distance between centroids of different clusters. Since the total number of clusters is arbitrary, the choice was made to generate schooling markets covering an area equal to a circular neighbourhood with a 10 km radius -roughly consistent with assumptions based on a secondary survey[5] that measured how far children enrolled in private middle schools travelled. This gives us 600 schooling markets.

Second, given the school markets and the villages contained in them, I consider new

[3] https://www.gao.gov/products/gao-22-104737.

[4] In the context of Indian villages however, movement of households in and out of villages is not common as there is no rural housing market and migration of households into villages for education is rare-in this respect it differs from the US context and this makes the neighbourhood composition less endogenous to schooling supply.

[5] National Sample Survey Organisation (NSSO) 2017-18

private school openings in a cross-section of villages: the main outcome variable of interest is a dichotomous variable representing whether any new school (offering middle school classes) opened between 2012-17. I look at the impact of a village's average income (proxied by mean per capita consumption expenditure) on whether a private middle school opens in its locality. Further, controlling for own income, I look at the impact of the surrounding villages in the schooling market. Income is measured in 2011-hence I take care of reverse causality. I also control for a rich host of other variables using the census and SHRUG (more in the data section). However there may still be concerns about endogeneity. These are addressed using decadal (2002-2011) weather shocks-precipitation and temperature shocks (cold spells) -the identification assumption being that these affect the income of the village and the surrounding villages but have no incremental direct effect on school openings in the future.

The results from a 2SLS regression show that a Rs. 1000 increase in the income of a village (with a mean income of Rs. 16,000) leads to 5% increase in private school provisioning while the same increase in income for only the neighbouring villages leads to a fall in probability of provisioning by the same magnitude. Thus, to have a shot at landing a private school, not only does a village have to be rich, it has to be rich relative to its neighbourhood. On the flip side, a poor village may even be worse off when surrounded by richer villages instead of poorer ones.

As pointed out above, the implication for human capital needs additional investigation when children can move to adjoining neighbourhoods to avail of the private school. Hence, I ask what happens to the overall human capital of the those residing in the schooling markets when incomes are high but inter-village inequality is also high. The results from an analogous 2SLS estimation suggests that in high mean income markets, greater inequality is associated with lower enrollment (controlling for population). Hence the spatial externality effect dominates the own high income effect to lead to fall in over-all human capital outcome. This evidence is supported by some suggestive evidence on learning levels: the ability to do at least subtraction. Having a private middle school within the village does affect math learning levels positively. Hence, its absence is likely to affect the relatively poor village adversely.

This paper contributes to the literature of neighbourhoods, schooling quality and human capital outcomes that is the recurrent theme in literature on intergenerational mobility (Chetty et al 2014). While the context of this work is different, it alludes to a potentially crowding out effect between neighbourhoods. Such concerns are less of a concern though when it comes to public schooling but similar effects are expected when private schools are involved. Neighbourhood effects interacting with income inequality have also been studied theoretically. Assuming a single private service provider and a uniform distribution of income over space, the impact of inequality was found to be non-monotonic owing to a trade-off: higher valuation of the rich attracts the supplier to enter into the neighbourhood but the service provider charges a higher price greater is the income or larger is the proportion of the rich in the neighbourhood (Gulati and Ray 2016). I empirically analyse the provisioning effect while departing from the uniform spatial distribution assumption with respect to income. Neighbouring villages have differing levels of incomes, which may not only influence provisioning within their own villages but may also impose spatial externalities on neighbouring villages. The empirical results show that spatial effects are of significant value as the market for schools extends well beyond the village it is located in.

The literature to which this paper contributes to more directly is that of private schooling in developing countries, particularly in India. Rising rural incomes, greater availability of both private and government services, better roads and better access to electricity in villages have all facilitated the growth of private schools in rural India. Population demographics is yet another factor driving location choices of private schools. Villages with private schools tend to be larger and have relatively more educated parents (Pal 2010). Presence of public schools and their performance also influences location of private schools, with some effects being temporal. Location patterns of private schools are, in part, a response to a shortage of teachers in rural areas. Government girls' secondary school in a village increases the probability of a private school by 300 percent—largely because yesterday's students in government schools are today's teachers in private schools (Andrabi et al 2007). Moreover, private schools are significantly more likely to exist in villages with high teacher absence in public schools (Muralidharan and Kremer 2006). Further, in line with my result, the income of the neighbourhood is a major determinant of private school placement. Such schools are located in intensely competitive schooling clusters: the average private primary school in rural Pakistan is located such that close to half of all other schools in the village are within a 5-minute walk, and less than a third are more than a 15-minute walk away. With 8 schools in every village, the average private school has close to 4 schools surrounding it.[6] Moreover, the clustering observed for private schools is around richer settlements - closer to banks, health centres and the main road (Andrabi et al 2007). This paper differs from all these papers are it focuses on an externality. I therefore add to this literature by bringing in the spatial aspect of what market a village lies in. This is increasingly more important as one

[6]Government schools tend to be less clustered, with just over a third of all schools in a village within a 5-minute walking distance of the average school.

looks at higher levels of schooling-where markets encompass many potential villages.

The structure of the paper is as follows. The next section describes the data. In section 3, I present the results of a machine learning exercise to demarcate schooling markets. Section 4 describes the empirical model. Section 5 discusses the potential endogeneity and the instrumental variables strategy. Section 6 presents the main 2 IV-2SLS results. Welfare implications are discussed in section 7. I conclude with a policy discussion in section 8.

2 Data

2.1 Sources

The analysis uses data for all villages in the state of Rajasthan and sources information on their geocodes, demographics and village level public goods (amenities) from the Census of India, 2011. However, information on new school openings between 2012 and 2017 require this data to be merged with data collected as a part of District Information System for Education (DISE), 2017-18. Given no synchronised village identifier, I use string matching algorithms to match names of the villages across the two data sets.[7] For cross validating the match so obtained, I compare the number of schools established before 2011 according to both sources. The validation exercise produces reliable results- a little over 93% of villages are such that the number of schools with middle (upper primary) schooling from the two sources are within 1 school deviation of each other.

Of the 44,900 villages in Rajasthan (Census 2011), 21487 (48%) villages get matched to the DISE database. These problems are well known- (Adukia et al 2020) use a similar matching for India and are able to match 65% schools. Moreover, it is important to note that there are two potential sources of non-matching. If a village is not matched across the two data sets, it is not clear whether the village has no school as of 2017- since DISE reports data for a village *only* if there is a school offering education in its locality[8]- or, whether the village indeed has a school but is not matched due to differences in village names- the names are in Hindi in official records and their conversion to English often introduces deviations that are hard to match. Therefore, I cannot conclude that unmatched villages have no schools. Instead, I am forced to limit the analysis to only matched villages.

[7] I use reclink in STATA

[8] These cover all schools in the locality: public schools managed by the state, those managed by local bodies or religious trusts and private schools-even those that may be unrecognised

The matched villages, constitute 70% of the total number of villages with a school (30,696) according to DISE. Some villages have improbable data on number of schools (more than 7 schools as on 2011) and are dropped from the sample. Such villages constitute 1% of the sample. Consequently, we are left with 21,356 villages which are used for analysis.

A crucial piece of information is the average income of the villages. I use village level mean consumption estimates for the village as a proxy for village income. These are sourced from SHRUG (Asher et al 2021) which reports small-area estimates following the methodology in (Elbers et al 2003).[9]

2.2 Descriptive Statistics

Schools differ in India based on who funds them. While state education departments often have their own schools, they also fund private schools-controlling fee and recruitment rules for hiring of teachers, making them de-facto public schools. In Rajasthan, private schools are those that are un-aided by the state and set their own fee and recruitment. I focus on such schools for this study.

To begin with, Table 1 provides a snapshot of the private schools in the state of Rajasthan. Almost 24% of villages had a middle school by 2011 and continue to grow with almost 8.5% of them showing growth of schools in the period 2012-2017. While the figures for primary school are comparable, I focus instead on middle schools as India has achieved almost universal enrolment in primary schools-hence the focus in India as well as all over the world (through Sustainable Development Goals) is to address issues of access at the upper primary level. Openings at higher levels- secondary and senior secondary- are rarer. Hence, I restrict attention to schools offering education at the upper primary level (I also refer to them as middle schools).

[9] This common used data source for India uses consumption and assets data from the 2011-12 India Human Development Survey (IHDS-II) to predict household-level consumption in the Socio-economic and caste census-2011 microdata- While the household level data are not available publicly, SHRUG reports the village level statistics for mean predicted consumption per capita.

Table 1: Private Schools

Level of Schooling	**% of Villages with at least one school by 2011**	**% of Villages with at least one new school in 2012-17**
Primary	25.17	12.00
Middle	23.88	8.46
Secondary	12.73	1.40
Senior Secondary	7.24	0.50

To give further context to this analysis, we describe the villages which form part of the study. As Table 2 shows, the annual average village income (consumption) is roughly Rs.16,000 per capita, with high variation across villages. These variations are high even within clusters. These villages vary substantially by population size with a mean size of 1200 people. Around 63% of the population is literate while 33% of the population (including women and children) is employed. Only 45% of the households living in the village in 2011 had access to electricity, with roughly similar proportions having access to drinking water, drainage and a bus stop. The terrain is somewhat sparse-most villages are far (on average, 60 kilometers away) from the district headquarters.

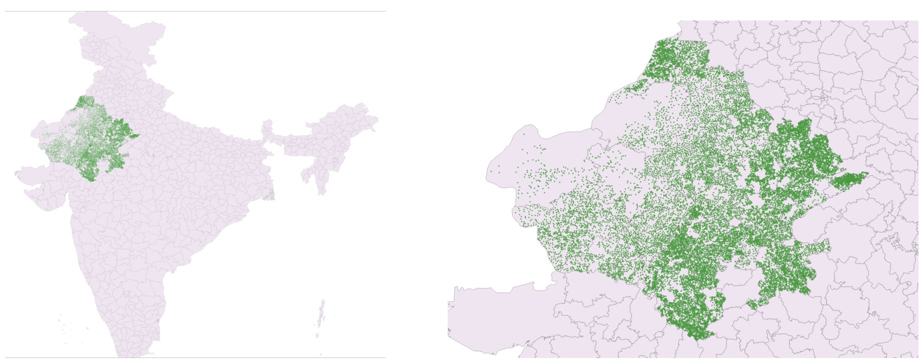
Table 2: Summary Statistics: Villages

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Village Per Capita Consumption (in 1000s)	16.507	4.021	7.714	39.836	20777
No. of Govt Middle Schools $<=$ 2011	0.807	0.666	0	6	20777
No. of Pvt Middle Schools $<=$ 2011	0.266	0.555	0	2	20777
No. of Govt Middle Schools $>$ 2011	0.007	0.082	0	2	20777
No. of Pvt Middle Schools $>$ 2011	0.077	0.302	0	4	20777
Village Population (in 100s)	12.236	10.095	0.01	103.75	20777
Literacy Rate	0.629	0.136	0	1	20770
Employment Rate	0.327	0.131	0	1	20777
Electricity	0.457	0.498	0	1	20777
Pucca Road	0.253	0.435	0	1	20777
Railways	0.02	0.14	0	1	20777
Drinking Water	0.395	0.489	0	1	20777
Drainage	0.393	0.488	0	1	20777
Bus Stop	0.457	0.498	0	1	20777
Post Office	0.011	0.103	0	1	20777
Distance to District Headquarters (kms)	60.807	35.83	0	260	20774
Elevation (meters)	309.484	130.033	9	1042	20777

3 Schooling Markets

Private middle schools are firms that locate strategically to garner the optimal market size, setting prices endogenously to maximise their profit. Typically, they cater to children in many villages simultaneously. This is consistent with student level data from the education survey conducted by the National Sample Survey Organisation (NSSO) in 2017-18 that shows that about 27% of students enrolled in private middle schools in Rajasthan, report travelling more than 5 kms to get to their school.[10] This motivates why, as a first exercise, I try to estimate the size of the geographical schooling market. Consider a school which locates within a village locality- I ask which are the other villages that would be part of its catchment area.

One way out is to assume that students travel upto 10 kms to get to their private school- given the data above, this is not an unreasonable upper bound. However, given that the notion of a 'market' in this context would revolve around distance between villages, should an area covered by a fixed distance radius be considered the market? Figure 1 shows us why in many contexts, and especially the context in Rajasthan, this may be an unreasonable assumption. Villages in Rajasthan showcase a combination of clustering and sparseness.



((a)) ((b))

Figure 1: Villages (Centroids) in Rajasthan

In such a scenario, a fixed radius area imposes an artificial boundary wherein villages just outside this area are unjustifiably excluded while adjacent villages are included. There is a natural clustering pattern to the villages which is likely to influence what 'markets' are- students in one cluster (set of villages close to each other) are likely to go a school situated

[10]The survey unfortunately top codes distance -hence while we know these students travel more than 5 kms, how much they actually travel needs further assumptions.

within that cluster rather than a village belonging to another cluster even if it is within some radius distance. Moreover, there is no clear radius that can be determined and in fact, this radius might even differ in accordance to the terrain and other geographic and economic factors. Therefore, to negate this ambiguity to some extent, I resort to a machine learning tool to identify villages 'close' to each other or in other words, to identify the nearest neighbours (villages).

I make use of the standard machine learning tool of K-means clustering to identify a 'market' building on the rough intuition about the distance typically travelled by students to access their schools. The *K-means* algorithm is an iterative algorithm that partitions the villages into K pre-defined distinct non-overlapping subgroups (clusters) where each village belongs to only one group. It tries to make the intra-cluster data points as same (close) as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the villages' centroid and the cluster's centroid (arithmetic mean of all the village coordinates that belong to that cluster) is minimized. The number of clusters has to be pre-defined- I observe that a 10km radius around a village, on average covers 35-40 other villages. To generate clusters which are nearly (but not exactly of 10km radius) as large, back of the envelope calculations result in setting the number of clusters to 600. Some villages, along with the centers of the clusters they belong to, are depicted in Figure 2 - villages belonging to different clusters are shaded with different colours and the cluster centers are depicted by the larger black dots.

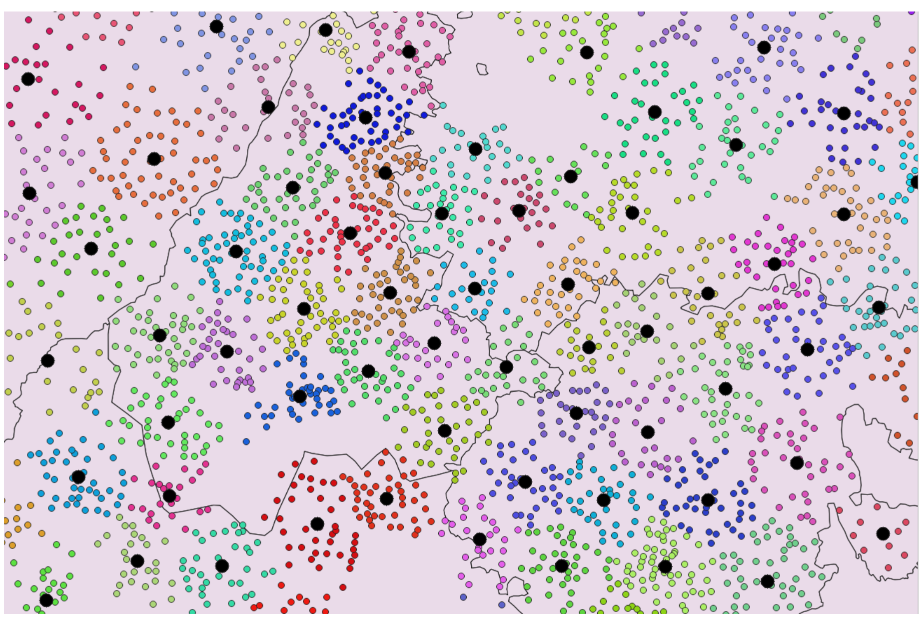


Figure 2: Clustered Villages in Rajasthan

Table 3 describes the clusters so formed. The clusters are large, encompassing 36 villages on average. They account for the mix of sparseness and clustering seen in the spatial distribution of the villages - some clusters are sparsely populated with close to 11 villages while some are dense with more than 80 villages. Consequently, the distribution of population across clusters is identical, with considerable variation ranging from less than 10,000 to more than 1.6 lacs. The total income for clusters is roughly Rs. 6 lacs with the poorest clusters being far behind the richest clusters.

Table 3: Summary Statistics: Clusters

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Cluster Size: No. of Villages	35.947	11.376	11	81	600
Cluster Population	49900.4	21200.15	8540	167385	600
Cluster Income	595702.18	219728.99	128821.97	1654445.6	600

4 Empirical Model

In the previous section, we have partitioned all villages into schooling markets. Given these schooling markets, each private school has a location choice- which schooling market to locate in and where within this market to locate its premises. Theory would suggest that there are two effects. If the profit maximizing fee is high enough to exclude most living in the poor neighbourhood, then the school should locate in the richer neighbourhood, exacting a higher price from well endowed consumers, who have easy access. On the other hand, if the profit maximizing price involves also catering to the poor, then locating in the poor neighbourhoods may be optimal since the rich may find it easier to bear the transport cost of longer travel. If one brings in cost considerations of acquiring land, it may well be optimal for a private school to locate in a poorer village contiguous to a richer one. This paper therefore treats this as an empirical issue.

The sample considered for this study is cross-sectional. Income (consumption expenditure) data, representative at the village level, is available only for 2011. The Census also reports the existing stock of private middle schools in 2011. However this correlation is spurious. For example, there can be serious issues of reverse causality: a higher stock of private schools in 2011 may lead to higher incomes in 2011, as such schools may have opened in the past, leading to better human capital outcomes and higher realised income, over time. To alleviate this problem, I bring in the DISE (at the cost of losing a substantial part of the sample). Given establishment year of private schools, a time series of the stock of private schools can be constructed using DISE. In particular, I construct a dependent variable s_{vs} which is a binary variable: which takes the value 1 if any new school (offering middle school classes) opens between 2012-17 in village v located in sub-district s.[11]

The main independent variables of interest are village income, denoted by Y_{vs} and average income of all other villages in the schooling cluster (Y^c_{vs}). The coefficient of Y_{vs} will measure the own income effect where as the coefficient of Y^c_{vs} would measure the impact of having a richer neighbourhood, holding the own income constant.

Private school openings are however determined by a host of other factors that can potentially confound the estimated marginal effects of income. The baseline existing stock of private as well government stocks in 2011 both play an important role in subsequent private school openings.[12]. Further, other village level characteristics like population, literacy and employment levels, population shares of ethnically backward communities - Scheduled Castes (SC) and Scheduled Tribes (ST), distance to district headquarters and elevation which are likely to affect physical accessibility and demand for schooling as well as income levels of villages have their own influence on where schools open up. Further, prior evidence points out to private schools opening up where infrastructure is better (Jagnani and Khanna 2020). Hence, I use principal components analysis (PCA) and create a village assets index- which captures access to electricity, all-weather roads, drinking water, drainage, bus stops, railways and post offices- factors likely to influence both income levels as well as access to schools, which rely on these public infrastructure facilities for profitability.

Given these determinants and the coefficients we are interested in, I estimate the following specification:-

$$s_{vs} = \alpha_s + \beta_1 Y_{vs} + \beta_2 Y^c_{vs} + \beta_3 no\_pvt_{vs} + \beta_4 no\_govt_{vs} + \gamma X_{vs} + \epsilon_{vs} \tag{1}$$

where v indexes villages and s indexes sub-districts. α_s accounts for sub-district level fixed effects, thus making comparing school markets over relatively small geographical areas. This also accounts for any large difference in topography, soil quality or any sub-state level macroeconomic shocks. Apart from the the variables defined above, X_{vs} is the vector of all demographic and infrastructure variables at the village level. I report robust standard errors

[11] A district (county) is divided geographically into sub-districts (also called blocks) for decentralization of administration.

[12] It can be contended that initial number of private schools is also endogenous, but in actual estimation, it has a negligible effect on the estimated coefficient

clustered at the level of a village council (a collection of contiguous villages are governed by a village council, which is also called a gram panchayat).

In this regression, the own income effect is captured by β_1 and the spatial externality from the neighbourhood is captured by β_2.

5 Identification Strategy

The income variables considered in the regression are likely to be endogeneous as incomes are correlated with several local factors like institutions, which is likely to influence school provisioning. Thus, I use an instrumental variable strategy to create exogeneous variation in these incomes, to study its impact on provisioning.

I consider two weather related factors: temperature and precipitation at the village level. Random shocks pertaining to these factors are likely to affect income in agrarian areas like rural Rajasthan. More importantly, identification relies on these factors having no independent impact (apart from those driven through income) on provisioning of schools. Areas more prone to such shocks might have adapted themselves to cope with these in ways that may influence provisioning of schools directly, thus challenging the exclusion restriction of the instruments. To avoid such variation, I look at precipitation and temperature anomalies. Precipitation (temperature) anomaly is the deviation of the decadal (2001-11) average of monthly precipitation (temperature) from the historical (1960-91) monthly average precipitation (temperature).

The identification strategy would be invalid if weather shocks impact provisioning of schools, independently of income. While research has established a persistent relationship between weather shocks and educational outcomes, particularly enrollment and learning outcomes, most of these effects are driven through income. Suffice it to say, concerns may still remain and may need a better identification strategy.

The existing empirical literature on the impact of weather shocks on education in developing countries focuses mainly on the extensive margin with an analysis of school enrollment and opportunity costs (Bjorkman-Nyqvist 2013) (Jensen 2000) (Shah and Steinberg 2017). It also discusses effects on the intensive margin (Zimmermann 2018). School enrollment effects have switched signs over the past 30 years, consistent with a decreased role for credit constraints and an increased importance of the opportunity costs of the child's time. Households are also increasingly re-optimizing educational expenditures and school type. School location decisions are likely to react to these demand-side considerations but these are driven via income. It is hard to imagine rainfall or temperature shocks directly influencing location choice of new private schools.

Since both of these instruments are spatially correlated, villages within a cluster are likely to witness very similar shocks. As a result, any one of these variables cannot be used simultaneously to generate exogeneous variation in income at the village (cluster) level while controlling for cluster (village) income. Therefore, I use rainfall anomaly at the level of the village and temperature anomaly at the level of the cluster as instruments for village and cluster incomes. Results from both stages of the 2SLS procedure are shown below. Alternatively, one could also use temperature anomaly at the level of villages and rainfall anomaly at the level of clusters- these results are consistent (the effect sizes are slightly larger but borderline significant with p values equalling 0.10 and 0.11 for own income and spatial externality effects respectively).

According to the Spatial Database for South Asia by the World Bank, only 20% of land is irrigated in Rajasthan. So, agriculture in these regions is fairly monsoon-dependent. In recent years, although variability in rainfall has been on the rise, compared to the historical average, 2001-11 saw positive rainfall shocks as depicted in the following figure.

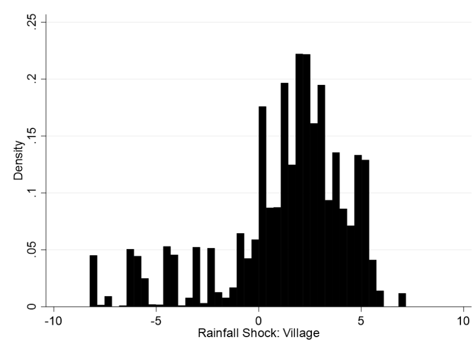
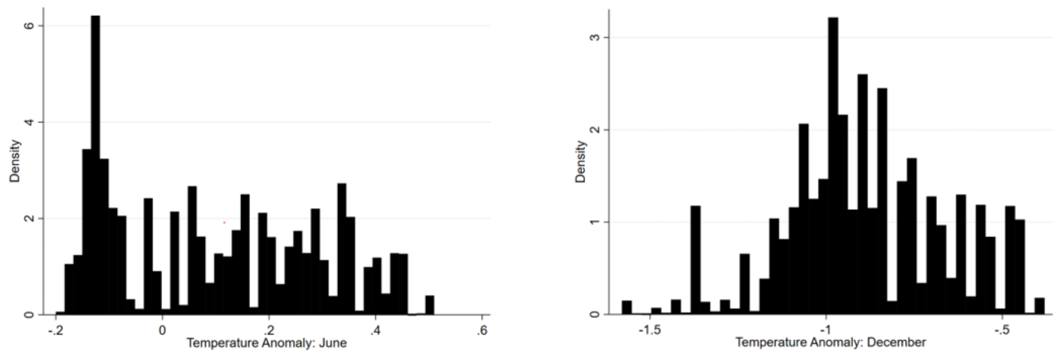


Figure 3: Rainfall Anomaly: Villages

Historical data analysis for extreme weather events indicates Rajasthan to be the second state after Jammu and Kashmir where maximum number of cold waves has occurred (TERI 2010 Draft Rajasthan State Action Plan on Climate Change New Delhi: The Energy and Resources Institute). The Spatial Database for South Asia by the World Bank is consistent with this finding: compared to the average for 1960-91, 2001-11 experienced colder conditions overall.



((a)) ((b))

Figure 4: Temperature Anomaly: Villages

Winters, as portrayed by temperature anomalies for December, have been cooler on average while in summers, there is more variation. Glancing at temperature anomalies for June, we see some villages experiencing cooler conditions while some others experienced warmer conditions. Overall, considering the average across all months, we see all villages experiencing colder spells, compared to their historical average. The intensity of the these spells is captured by the Cold Spell variable, plotted below at the cluster level (population weighted average of the village level cold spells to reflect the income metric of the cluster). It is defined as the absolute value of the temperature anomaly variable which takes a negative value (when anomalies are averaged across all months) for all villages.

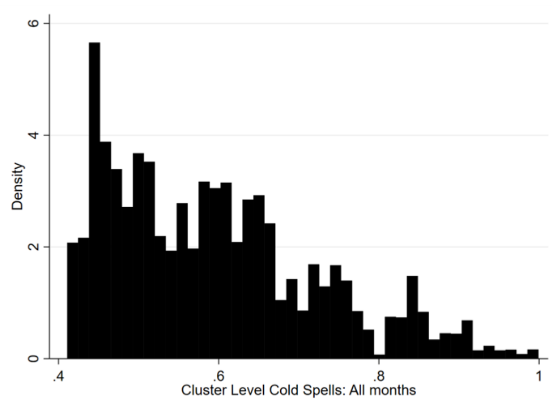


Figure 5: Cold Spells: Cluster

6 Results

Table 4 reports the OLS results for the own income effect, which reflect the impact of village's own income on provisioning within the village. I find evidence for all effects seen in literature- population size, literacy, SC-ST shares, infrastructure and of particular interest, income. Populous and literate villages are more likely to get a school, so do villages with greater access to public infrastructure while those with greater share of dis-empowered groups like SCs and STs are less likely to get a school. The own income effect (income however is endogeneous) is positive but small. A Rs.1000 increase in village per-capita income is associated with 0.17% increase in probability of middle school provisioning, an effect of 2% on the mean (8.46%).

Richer villages are more likely to get schools but upon that, does the neighbourhood matter? I now estimate our main specification (equation 1) which includes the income of the cluster. The OLS results for this are presented in Table 5. The own income effect is consistent and stable. Although borderline significant (p value= 0.135) in this specification, the spatial externality effect is negative and almost of the same magnitude as the own income effect: a Rs.1000 increase in (population weighted) per-capita income of the neighbourhood reduces the probability of provisioning by 0.26%, an effect of 3% on the mean.

One should, however, be cautious of interpreting the OLS results as both the village and cluster incomes are endogeneous variables. As discussed in the previous section, for identification of the effects of the spatial income distribution, I use random variation in climatic factors as instruments, to create exogeneous variation in both income variables. Instrumenting village income with rainfall anomalies of the previous decade and cluster income with corresponding temperature anomalies, the main specification (equation 1) is estimated and the main results from the first and second stage are presented below. Table 6 and 7 in the appendix have the complete results.

	(1)	(2)
Dep. Variable Per Capita Consumption (in 1000s):	Village	Cluster
Rainfall Shock: Village	0.1522***	0.0612***
	(0.0415)	(0.0207)
Cold Shock: Cluster	-3.8847***	-5.0428***
	(1.1098)	(0.5809)
Observations	21,346	21,346
R-squared	0.4478	0.8346

Robust standard errors, clustered at the GP level, in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The first stage is strong, with positive rainfall shocks affecting incomes positively and cold shocks having a negative association with income. While the former is more or less expected due to low irrigation penetration, the latter follows mainly from crop damage due to frost conditions as an excerpt from a 2008 article[13] showcases: "A good mustard crop insulates the country from the spikes in global vegetable oil prices. But there may be an over one-million-tonne drop in the mustard crop this year as frost damage and poor soil moisture pull down yields in parts of Rajasthan and Haryana. The maximum damage due to frost is in Hanumangarh, Ganganagar, Kota, Jhalawar and Patan areas of Rajasthan, and Rewari and Bhiwadi in Haryana." A report[14] by the Indian Council of Agricultural Research on impact of the cold wave of 2002-03 on agriculture underscores this phenomenon more generally: "Normally cold waves occur in western Rajasthan for 10-12 days during January and that too intermittently. However, during 2002-03, the cold wave prevailed continuously for three weeks. Previous records showed that the longest severe cold wave occurred in 1998 for a maximum period of two weeks. The peculiar feature of 2002-03 cold wave was that its occurrence was coupled with dense fog and less sunshine affecting crops through frost injury and reduced levels of photosynthesis. There was significant damage to crops, horticulture, forest trees, livestock, fisheries and other livelihood systems in these regions that led to significant economic losses to the farming community."

[13] https://economictimes.indiatimes.com/industry/cons-products/food/mustard-harvest-shortfall-turns-the-heat-on-cooking-oil/articleshow/2814540.cms?from=mdr

[14] "Cold Wave of 2002-03 – Impact on Agriculture" by Samra et al (2003)

VARIABLES	(1) Any New Pvt School > 2011
Village Per Capita Consumption (in 1000s)	0.0506*
	(0.0306)
Cluster Per Capita Consumption (in 1000s)	-0.0555*
	(0.0326)
Observations	21,346

Robust standard errors, clustered at the GP level, in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The own income effects carry over, with a 60% greater probability (on the mean) of provisioning. The spatial externality effect is negative, reducing the probability of provisioning equally. Overall, I find that provisioning of private schools requires a village to be relatively richer as compared to its neighbourhood. A rich or resourceful village is more likely to have a private school but also faces a spatial externality: it is less likely to get a school the richer is its neighbourhood. If other factors apart from pricing and perceived profits are driving these location choices, we might see similar patterns for Govt. schools as well. As a falsification test, I run the same IV regression for Govt. middle schools. Results reported in Table 8 of the appendix show that the while the cluster effect is significant, the own income effects are imprecisely estimated but if anything, show opposite patterns. Relatively poorer villages of a neighbourhood are more likely to have a Govt. school setup within the village, perhaps alluding to effective targetting of policies at the middle schooling level as well.

7 Welfare Implications

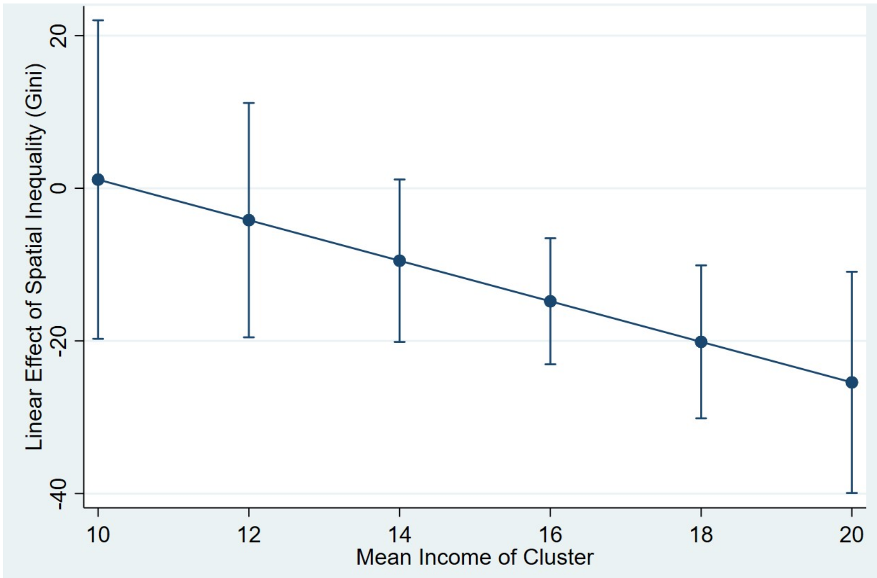
7.1 How far are schools?

We know that being a poor village in a rich neighbourhood leads to the worst outcomes when it comes to provisioning but we have looked at provisioning only in terms of whether or not the village has a school. What provides more perspective is to understand how far schools actually are. For each cluster, the poorest village is identified and distance to the nearest village with a private middle school is computed. I define relative poverty as the ratio of the income of the poorest village to the population weighted average of neighbouring villages of the cluster. I find a monotonically increasing relationship between relative poverty and distance to school, implying that being surrounded by richer villages (greater spatial inequality) negatively affects access to private schooling for the poorest in each market. A 10 percentage points increase in relative poverty increases distance to school by roughly half a kilometer.

7.2 Enrollment

We know that relatively rich villages of clusters tend to get schools. An important issue to consider next is the implications of accessibility for the poor villages within the clusters. To analyse this, I look at enrollment of schools, reported in DISE. Constraints of the data do not allow analysis of enrollment by residence of the student. We only have information about school-wise enrollments. Given our market structure, discussed in Section 3, we know a school serves children living in the same cluster as the school. These results show that the school is likely to be located in the relatively richer villages of the cluster. Are the children residing in the poorer villages of the cluster able to access these schools? We have seen that the distance from the poorest village to the nearest private school increases monotonically, the poorer the village is relative to its neighbourhood. These trends, however, would not be a concern if they are not detrimental to enrollment among children of the poorer villages.

It is natural to expect children from the relatively richer villages of the cluster to access schools, as they are located closer to them. Therefore, I assume differential enrollments across clusters to reflect difference in access for the poorer villages. I look at the interaction of the own income and spatial externality effects. The spatial externality effect implies that greater inequality within the cluster would aid private school provisioning, provided the own income does not dominate, as this reflects existence of villages which are rich, relative to their neighbourhood. I find this to be true only at higher levels of income. For low mean income markets, there is low provisioning at all levels of inequality. So, the own income effect appears to dominate the spatial externality effect at low income levels. In high mean income markets, however, the spatial externality effect kicks in and greater inequality is associated with lower enrollment. The marginal effects of inequality, as measured by the Gini of the cluster, is plotted for various levels of mean income in the figure below. So far as children from richer villages are more likely to be enrolled in these schools in all clusters, these results suggest that greater spatial inequality, or existence of rich villages in close proximity to poorer ones, reduces enrollment among the poorer villages. Therefore, I find suggestive evidence for poor villages in richer localities being worse off in terms of access, when compared to the poor villages in poorer localities.



7.3 Learning Loss

Imperative to this discussion on access to schools are its consequences for learning, one of the most important outcomes with respect to schooling. What is the association between better access to private schooling and learning outcomes? I use Annual Survey of Education (ASER) [15] data from 2017-18 to explore this relationship. Note that ASER covers only rural India, which is the domain of this analysis. I restrict the ASER sample to middle school (grades 6 to 8) students in Rajasthan.

The ASER scores for arithmetic ability varies from 1 to 5, with 1 implying child could not do any arithmetic, 2 implying child can recognise numbers 1-9, 3 implying child can recognise numbers 11-99, 4 implying child can do two-digit subtraction and 5 implying child can do division (3-by-1 form). The mean arithmetic ability of the sample is 3.8 - most students not being able to do subtraction. I consider that as the outcome: ability to do at least subtraction and find a positive relationship between this outcome and villages having a private school. This follows from multivariate OLS regression which controls for several possible confounders: parents going to school, household facilities like electricity, 2 wheelers,

[15] ASER, is an annual, citizen-led household survey that aims to understand whether children in rural India are enrolled in school and whether they are learning. Nation-wide ASER surveys provide representative estimates of the enrollment status of children aged 3-16 and the basic reading and arithmetic levels of children aged 5-16 at the national, state and district level

mobile phone access and presence of infrastructure in the village- pucca roads and Govt. middle schools (Table 9).

This aspect of the impact of spatial inequality needs further analysis but helps in understanding how varying incomes over space could translate into differential access to private schooling which could in turn, affect learning outcomes.

8 Conclusion

In this paper, I investigated whether richer neighbourhoods crowd out access in the relatively poor ones, which are contiguous to them. Further, the paper showed that these have implications on aggregate human capital accumulation. I find evidence for externalities imposed over space, wherein villages which are richer relative to their neighbourhood are more likely to gave access to private schools. This may call for public intervention, more so in developing countries, where private schools are being increasingly seen to provide quality education to the masses.

What sort of form could a public intervention take? It can be argued that richer students are more likely to have resources at their disposal which would aid them to travel a greater distance to be able to get to school. The socially optimal, access maximising location choice of schools, therefore, should be poorer villages. Private resources seem to be guiding choices in the opposite direction with relatively richer villages getting greater access.

This work in preliminary, but the issue is general, and if found to be robust, then certain policies that we see, even in urban areas, might worsen the problem: for example, preferential treatment for admissions based on proximity to school is likely to worsen location sub-optimality. Hence, this work calls for a deeper dive into the issue of spatial externality in the supply of schooling.

References

Adukia, Anjali, Sam Asher, and Paul Novosad. 2020. “Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction.” American Economic Journal: Applied Economics, 12 (1): 348-76.

Andrabi, Tahir, Jishnu Das, Asim Khwaja, Tara Vishwanath, Tristan Zajonc: “Learning and Educational Achievements in Punjab Schools (LEAPS): Insights to inform the education policy debate” (2007)

Asher, Sam, Tobias Lunt, Ryu Matsuura and Paul Novosad: “Development Research at High Geographic Resolution: An Analysis of Night Lights, Firms, and Poverty in India using the SHRUG Open Data Platform” (2021)

Barro, Robert J., Xavier Sala-i-Martin, “Convergence across States and Regions”. Brookings Papers on Economic Activity, No. 1, 1991, pp. 107-158

Björkman-Nyqvist,Martina (2013), Income shocks and gender gaps in education: Evidence from Uganda, Journal of Development Economics, 105, (C), 237-253
from Pakistan.” (2016) IZA Discussion Paper No. 9960

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez: “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States” (2014) The Quarterly Journal of Economics, Volume 129, Issue 4, November 2014, Pages 1553–1623

Elbers, Chris, Lanjouw, Jean and Lanjouw, Peter. (2003). “Micro–Level Estimation of Poverty and Inequality”. Econometrica. 71. 355 - 364.

Gulati, Namrata and Tridip Ray: “Inequality, Neighbourhoods and Welfare of the Poor” (2018) Journal of Development Economics, 122, (C), 214-228

Jagnani, Maulik and Khanna, Gaurav, 2020: “The effects of elite public colleges on primary and secondary schooling markets in India,” Journal of Development Economics, Elsevier, vol. 146(C)

Jensen, Robert 2000. ”Agricultural Volatility and Investments in Children.” American Economic Review, 90 (2): 399-404

Mankiw, N. Gregory, David Romer, David N. Weil, “A Contribution to the Empirics of Economic Growth”. The Quarterly Journal of Economics, Vol. 107, 1992, pp. 408-437

McCoy DC, Peet ED, Ezzati M, Danaei G, Black MM, et al. (2017) "Early Childhood Developmental Status in Low- and Middle-Income Countries: National, Regional, and Global Prevalence Estimates Using Predictive Modelling". PLOS Medicine 14(1)

Muralidharan, Karthik and Michael Kremer: "Private and Public Schools in Rural India" (2006)

Pal, Sarmishtha: "Public infrastructure, location of private schools and primary school attainment in an emerging economy" (2010) Economics of Education Review, Volume 29, Issue 5, October 2010, Pages 783-794

Shah, Manisha and Bryce Steinberg, (2017), Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital, Journal of Political Economy, 125, (2), 527 - 561

Zimmermann, Laura. "Remember When It Rained - Schooling Responses to Shocks in India" (2018) Working Paper

9 Appendix

Table 4: OLS: Own Income Effect

	(1)	(2)	(3)	(4)	(5)
	Any New Pvt School > 2011				
Village Per Capita Consumption (in 1000s)	0.0033***	0.0016***	0.0019***	0.0017***	0.0017***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
No. of Govt Middle Schools by 2011		0.0575***	0.0152***	0.0122***	0.0122***
		(0.0036)	(0.0040)	(0.0040)	(0.0040)
No. of Pvt Middle Schools by 2011		0.0759***	0.0235***	0.0216***	0.0215***
		(0.0040)	(0.0050)	(0.0050)	(0.0050)
Village Population (in 100s)			0.0066***	0.0054***	0.0054***
			(0.0004)	(0.0004)	(0.0004)
Literacy Rate			0.0553***	0.0432***	0.0407***
			(0.0144)	(0.0143)	(0.0145)
Employment Rate			-0.0059	-0.0058	-0.0057
			(0.0126)	(0.0126)	(0.0126)
SC, ST Share			-0.0073*	-0.0076**	-0.0076**
			(0.0039)	(0.0038)	(0.0038)
Village Assets Index				0.0142***	0.0141***
				(0.0023)	(0.0023)
Distance to District Headquarters					-0.0001
					(0.0001)
Elevation					-0.0000
					(0.0000)
Constant	0.0235***	-0.0240***	-0.0916***	-0.0612***	-0.0397**
	(0.0090)	(0.0087)	(0.0119)	(0.0128)	(0.0197)
Observations	21,357	21,357	21,350	21,350	21,347
R-squared	0.0401	0.1239	0.1541	0.1563	0.1564

Robust standard errors, clustered at the GP level, in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: OLS: Own Income and Spatial Externality Effects

	(1)	(2)	(3)	(4)	(5)
	Any New Pvt School > 2011				
Village Per Capita Consumption (in 1000s)	0.0034***	0.0017***	0.0020***	0.0018***	0.0018***
	(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Cluster Per Capita Consumption (in 1000s)	-0.0009	-0.0010	-0.0019	-0.0022	-0.0026
	(0.0018)	(0.0017)	(0.0017)	(0.0017)	(0.0017)
No. of Govt Middle Schools < 2011		0.0575***	0.0152***	0.0121***	0.0121***
		(0.0036)	(0.0040)	(0.0040)	(0.0040)
No. of Pvt Middle Schools < 2011		0.0760***	0.0234***	0.0215***	0.0215***
		(0.0040)	(0.0050)	(0.0050)	(0.0050)
Village Population (in 100s)			0.0066***	0.0054***	0.0054***
			(0.0004)	(0.0004)	(0.0004)
Literacy Rate			0.0571***	0.0451***	0.0425***
			(0.0146)	(0.0145)	(0.0146)
Employment Rate			-0.0059	-0.0057	-0.0056
			(0.0126)	(0.0126)	(0.0126)
SC,ST Share			-0.0073*	-0.0076**	-0.0076**
			(0.0038)	(0.0038)	(0.0038)
Village Assets Index				0.0142***	0.0142***
				(0.0023)	(0.0023)
Distance to District Headquarters					-0.0001
					(0.0001)
Elevation					-0.0001
					(0.0000)
Constant	0.0384	-0.0075	-0.0622**	-0.0271	0.0029
	(0.0296)	(0.0283)	(0.0281)	(0.0287)	(0.0342)
Observations	21,356	21,356	21,349	21,349	21,346
R-squared	0.0401	0.1239	0.1541	0.1564	0.1565

Robust standard errors, clustered at the GP level, in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: IV: First Stage

Dep. Variable Per Capita Consumption (in 1000s):	(1) Village	(2) Cluster
Rainfall Shock: Village	0.1522***	0.0612***
	(0.0415)	(0.0207)
Cold Shock: Cluster	-3.8847***	-5.0428***
	(1.1098)	(0.5809)
No. of Govt Middle Schools < 2011	0.0670*	-0.0157
	(0.0383)	(0.0128)
No. of Pvt Middle Schools < 2011	0.2663***	-0.0162
	(0.0359)	(0.0126)
Literacy Rate	9.3135***	1.1100***
	(0.2794)	(0.0891)
Employment Rate	1.3201***	0.1215*
	(0.2136)	(0.0726)
Village Population (in 100s)	-0.0348***	-0.0015
	(0.0031)	(0.0012)
SC,ST Share	-0.4597	-0.0106
	(0.3205)	(0.0127)
Village Assets Index	0.1753***	0.0252***
	(0.0222)	(0.0082)
Distance to District Headquarters	-0.0086***	-0.0109***
	(0.0015)	(0.0008)
Elevation	-0.0030***	-0.0013***
	(0.0007)	(0.0004)
Constant	14.2880***	19.9507***
	(0.7769)	(0.3662)
Observations	21,346	21,346
R-squared	0.4478	0.8346

Robust standard errors, clustered at the GP level, in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: IV: Second Stage

VARIABLES	(1) Any New Pvt School > 2011
Village Per Capita Consumption (in 1000s)	0.0506*
	(0.0306)
Cluster Per Capita Consumption (in 1000s)	-0.0555*
	(0.0326)
No. of Govt Middle Schools < 2011	0.0080
	(0.0051)
No. of Pvt Middle Schools < 2011	0.0076
	(0.0101)
Village Population (in 100s)	0.0071***
	(0.0011)
Literacy Rate	-0.3525
	(0.2568)
Employment Rate	-0.0633
	(0.0406)
SC,ST Share	0.0143
	(0.0194)
Village Assets Index	0.0069
	(0.0054)
Distance to District Headquarters	-0.0002
	(0.0002)
Elevation	0.0000
	(0.0001)
Observations	21,346

Robust standard errors, clustered at the GP level, in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: IV: Falsification Test with Govt. Schools

	(1)
VARIABLES	Any New Govt School after 2011
Village Per Capita Consumption (in 1000s)	-0.0221
	(0.0139)
Cluster Per Capita Consumption (in 1000s)	0.0240*
	(0.0134)
No. of Govt Middle Schools < 2011	-0.0091***
	(0.0022)
No. of Pvt Middle Schools < 2011	0.0076*
	(0.0042)
Village Population (in 100s)	0.0002
	(0.0005)
Literacy Rate	0.1930
	(0.1173)
Employment Rate	0.0190
	(0.0181)
SC,ST Share	-0.0113
	(0.0105)
Village Assets Index	0.0045*
	(0.0024)
Distance to District Headquarters	0.0000
	(0.0001)
Elevation	-0.0000
	(0.0000)
Observations	21,346
R-squared	-0.6111

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Presence of Private School and Learning

VARIABLES	(1) Student can do subtraction
Village has Pvt School	0.0825**
	(0.0359)
Village has Govt. Middle School	-0.0368
	(0.0351)
Village has pucca roads	0.0105
	(0.0574)
HH has electricity	0.0102
	(0.0426)
Mother went to school	0.0859***
	(0.0167)
Father went to school	0.0226
	(0.0240)
HH has a 2 wheeler	0.0457
	(0.0270)
HH has a mobile phone	0.0083
	(0.0435)
Constant	0.5786***
	(0.0909)
Observations	2,056
R-squared	0.0819

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1