UNIVERSITY OF MANNHEIM
Department of Economics
M.Sc. in Economics

MASTER THESIS

Does rural electrification affect primary school enrolments?

Evidence from India

Supervisors:

Prof. Ulrich Wagner, PhD
Dana Kassem, PhD
Dimitri Szerman, PhD

Marianna Magagnoli
1525997
mmagagno@mail.uni-mannheim.de

July 13, 2020

Abstract

Large amount of funds are still invested in giving rural areas access to electricity. The economic literature has studied its impact on a range of aspects. This thesis focuses on the effects which electrification had on education in Indian villages. It exploits the requirements entailed in the guidelines of the Rajiv Gandhi Grameen Vidyutikaran Yojana (RGGVY) electrification program, which determined eligibility for funds based on a population threshold. Through a RD identification strategy, it estimates that the program had no significant impact on consumption of electricity nor on education. However, results showcase a significant negative effect on school enrolments in proximity of towns. Compared to the baseline, in this subsample the intention-to-treat effect amounts to -20% in the medium-term and -22.7% in the long-term. These results are large and in contrast with most of the existing literature. They would represent an unexpected drawback of policies aimed at electrifying rural areas in order to allow economic development and improve welfare in the long term.

Declaration of Authorship

With this statement I, Marianna Magagnoli, declare that I have independently completed this master's thesis entitled "Does rural electrification affect primary school enrolments? Evidence from India" and have not used any other sources than those indicated in the thesis. All thoughts and illustrations that are taken directly or indirectly from external sources are properly marked as such. The thesis has not been submitted in this or a similar form to any other academic institution.

Mannheim, July 13, 2020

Marianna Magagnoli

Contents

1 Introduction

This thesis studies the impact which the Indian RGGVY [1] rural electrification program had on children's primary education, in the medium-term and in the long-term. As electrification allows local economic development, returns to education increase, incentivizing households to send their children to school. At the same time, in the context of a developing country such as India, where the poverty rate was still 21% in 2011 (World Bank, 2020d), households may care more about immediate economic opportunities than the long-run investment in children's education. This may cause an early exit from schooling, an unintended consequence of rural electrification policies. In other words, rural electrification creates a trade-off between increased returns to education and increased opportunity cost of schooling. Which of these effects prevails is an empirical question, to which the wide economic literature still finds contradicting results.

Overall, I find no significant effects of rural electrification on primary school enrolments. Nevertheless, this thesis suggests that the opportunity cost of schooling effect may be prevailing in villages closer to towns. Pointing to market integration and the economic drive exerted by towns, rural electrification seems to decrease primary school enrolments in those rural villages located in proximity to towns.

Very large amounts of money have been invested in rural electrification so far, which continues to be considered one of the primary means for poverty alleviation. The 2030 Agenda for Sustainable Development (United Nations, 2016) was adopted by all United Nations Member States in 2015. It recognizes the universal access to clean energy as one of the goals to be pursued hand-in-hand with those of ending poverty, ensure inclusive and equitable quality education and reducing inequalities. While electricity is pervasive in the industrialized world, energy poverty is still widely present in developing countries. The World Bank estimates that 12% of the world population still lacked access to electricity in 2017. In rural areas this figure raises to 22% of the population. People without electricity are mostly in developing Asia (51%) and Africa (44%) (Bonan et al., 2017). The Indian Government has invested in rural electrification since its independence in 1947. While only 43% of the rural population had access to electricity by 2001, in 2017 this covered the 89% (World Bank, 2020b).

The benefits of electrification have been studied on a range of socio-economic aspects. Replacement of kerosene lamps with electric lighting reduces indoor air pollution and provides better lightning at a cheaper price. This improves health conditions and potentially allows children to study longer. At the same time, the replacement of wood as source of fuel for cooking and the introduction of electric appliances inside the house free up time previously spent on wood collection and household chores. Electricity also opens opportunities for home-based businesses. Operating small appliances to provide market services becomes feasible; food preparation and storage become easier. Agriculture may also benefit from more efficient electric machinery. Telecommunications and mass media become accessible and offer a 'window to the world'. Village electrification may influence migration choices by making some places more desirable than others. As it will be discussed in Section 2, the relationship to educational outcomes ranges from positive to no effect.

Estimating causality of electrification is challenging. As it is the case for other kinds of infrastructure, the connection of a village to the grid is likely to be an endogenous factor. Which communities will be connected is often based on cost-effectiveness cri-

[1] Rajiv Gandhi Grameen Vidyutikaran Yojana

teria (distance to the existing grid, population size, average community income or productive potential), social allocation (based on poverty or other social indicators) or both (World Bank, 2008). Electricity has always played an important role in politics in India. As Baskaran et al. (2015) find, service provision may be influenced by election cycles. Thus, endogenous treatment allocation and confounding economic trends would bias a standard OLS estimation. The RGGVY program was introduced in 2005 and has allocated funds to Indian rural villages based on a minimum population threshold of 300 inhabitants. This implementation rule allows me to overcome the endogeneity and estimate the intention-to-treat effect of the program through a sharp regression discontinuity design.

This thesis follows Burlig and Preonas (2006). While Burlig and Preonas (2006) focus on many other aspects, I limit the analysis on the educational outcomes and extend it to a longer term. I also explore more of the heterogeneity in the effect, by gender as well as by proximity to town and cities. This thesis also contributes to the literature on causal effects of rural electrification through a relative new source of data first introduced in the economic literature by Min (2011) and Henderson et al. (2012). Remote sensing on satellite night lights images can be used to test for the actual increase in consumption of electricity, which can not be captured by binary variables. It has also been proven to provide a relatively good approximation to economic activity (Min et al., 2013), although it still performs poorly in detecting the one from the agricultural sector (Beyer et al., 2018). Furthermore, the data allow to perform the analysis at the smallest geographical level of the village, for which economic activity and energy consumption data are usually hard to obtain.

Using the Population Census of India of 2001[2] as running variable, I first test for the effective change in electrification usage caused by the program. I use village-level night lights data from the NOOAA satellites as a proxy for electricity consumption. While Burlig and Preonas (2006) find positive and significant effects, I find none. I provide some discussion about why this could be the case. I then turn to the educational outcomes. I use enrolment data of every Indian school contained in the DISE[3] dataset. For each school, it reports enrolment levels by class and by gender. The DISE also contains data on examination results (after Class 5 and Class 8) and many other school characteristics. I construct a panel data from 2005 until 2017 of all Indian schools and aggregate the data at the village level.

Overall, I find no significant effect of program eligibility on enrolment levels. The result is homogeneous across gender, level and time. I find a negative effect of around -1 in the number of children passing the Class 5 or Class 8 examination in 2017 with higher grades.

When taking into account the geographic location of the villages with respect to the nearest town, results reveal a different story. Eligibility to the electrification funds caused 2011 enrolment levels to drop on average by 17 students in villages which are in closer proximity of towns with at least 10k inhabitants. This corresponds to a 20% decrease in 2005 enrolments for a village at the mean of the subsample. I do not find any effect on enrolments in more isolated villages. These results are robust over time and by size of town (up to 100k inhabitants). The number of students passing the end-of-the-year examination with a grade higher than 60% also decreases only in villages closer to town, while remaining statistically unchanged in more isolated villages.

With the data available, I explore whether electrification has induced migration flows

[2]The last available population census before the introduction of the RGGVY program in 2001
[3]District Information System on Education

to or from eligible villages. Out-migration from rural areas around town towards the city itself may explain the drop in enrolment levels. In eligible villages, population increased in 2011 by 11% compared to 2001. I find no clear negative pattern in villages closer to a town. On the other hand, electrification seems to attract new inhabitants to rural, more isolated villages. While studying migration flows induced by investments in electrification is out of the scope of this master thesis, this evidence seems to exclude the possibility that the effect on school enrolments is driven by out-migration flows.

Lastly, I examine the extent to which program requirements were enforced. A regression discontinuity strategy identifies control and treatment group based on a hard threshold. Weak enforcement of program requirements and the presence of corruption in funds allocation may threaten the validity of the identification strategy.

The rest of this thesis is organized as follows: Section 2 describes the economic literature studying the diverse effects of electrification, Section 3 presents the theoretical framework underlying the potential impacts of rural electrification on education. Section 4 gives an overview of the history of electrification in India and of the current education system. Section 5 and Section 6 present the data used and summary statistics. The identification strategy and results are illustrated in Section 7. Section 8 and Section 9 wrap up this thesis with a discussion and conclusion.

2 Literature Review

The literature on the effects of rural electrification has focused on various socio-economic aspects. A large part of it studies the effect on local economic development, through employment, income and consumption data. Results have been so far ambiguous, ranging from positive to no effect. Rud (2012) finds that rural electrification in India increases manufacturing output by 14%, through an increase in production from smaller firms. Khandker et al. (2014, 2013) suggest electrification increases labor supply, particularly for women (+17% employment hours), by relieving time burdens spent in collecting fuel and making household chores more efficient. Household income and expenditure also benefit from the connection to electricity, but income effects seem to accrue mostly to the non-farm sector (Chakravorty et al., 2016). van de Walle et al. (2013) observe a relocation of casual wage work to regular work. On the other hand, the empirical literature also finds insignificant effects on income in Rwanda (Bensch et al., 2011) and no effect on employment, nor working hours in the Philippines (Chakravorty et al., 2016).

Electrification improves health conditions, mostly by lowering indoor air pollution from lightning (Bensch et al., 2011; Khandker et al., 2014). Unelectrified rural households rely on kerosene lamps. The lamp oil is known to generate only a low level of illumination and to produce harmful gases. Barron and Torero (2017) identify a reduction of 66% in the overnight PM2.5 concentration for households randomly connected to the electrical grid. This lowers the prevalence of acute respiratory infections such as tuberculosis (Pokhrel Amod K. et al., 2010).

Improved health and longer hours of light allow children to study longer and keep going to school, while adults may read more. The literature on the effects of electrification on education is relatively scarce compared to the one on economic outcomes. It has found predominantly positive effects on illiteracy rate, years of schooling, school enrolments and study time at home (Khandker et al., 2014; Lipscomb et al., 2013; van de Walle et al., 2013; Khandker et al., 2013). However, other studies find no effect on enrollments (Burlig and Preonas, 2006), study time (Bensch et al., 2011) or test scores (Lee et al., 2020).

This thesis contributes to this strand of literature, analyzing the effect of rural electrification on primary school enrolments and test scores. I initially follow Burlig and Preonas (2006) and I extend the analysis to explore longer-term effects, as well heterogeneity by gender and location.

Electrification seems to have gender heterogeneous effects. The shift from wood to electricity for cooking and lightning decreases the time women spend in household chores. This creates an endowment effect, which may allow them to increase their labor supply in the market. Dinkelman (2011) finds that electrification raises employment and working hours for women but not for men. van de Walle et al. (2013) and Khandker et al. (2013) show a disproportionate increase in school enrolments for girls compared to boys.

I study whether electrification has had spatially different effects on education. How urban proximity interacts with connection to the grid constitutes a gap in the literature. Empirical results from what is called 'new economic geography' show that urban demand exerts a distinct influence on the types of activities that take place in rural areas (Deichmann et al., 2009). Better infrastructure and connection to cities extend the size of the market for the closest rural villages. They may incentivize firms to relocate from urban to rural areas, where wages are cheaper or when cities are too congested and have higher costs of living. This results in a shift in the economy from farm towards non-farm sectors and more market-related activities in villages closer to urban areas (Asher and Novosad, 2016; Fafchamps and Wahba, 2006; Foster and Rosenzweig, 2004). Urban proximity seems to affect the rural local economy up to 3-4 hours of travel away from the city (Fafchamps and Shilpi, 2005). Fafchamps and Wahba (2006) explore how spatial location affects children's education and child labor. They point out that child labor is mostly a rural phenomenon in Africa, but an urban one in Asia and Latin America, where children are more likely to work in small-scale industrial enterprises or small trade and service businesses. The net effect of urbanization on child labor and education is a priori ambiguous. Wages are relatively higher in cities than in rural areas. Higher household incomes and higher returns to education increase the incentive for parents to send their children to school. At the same time, higher wages and more employment opportunities raise the opportunity cost of schooling. Using data from Nepal, Fafchamps and Wahba (2006) find that children residing in or closer to urban centers attend school more and work less in total. Fafchamps and Shilpi (2005) show that female labor supply is lower and more specialized in market-related activities or home-based chores when residing closer to town. While less time is spent on fetching water or firewood, households allocate more time to cooking, cleaning and shopping.

This thesis also contributes to the broad literature which uses quasi-experiment identification strategies to study the effect of infrastructures. As pointed out in Section 1, identifying the causal impacts of infrastructures such as electrification entails the threats of reverse causality and endogenous program placement. Researchers have less opportunities to manipulate or randomize large infrastructure developments. Few studies were able to implement a randomized control trial so far (Bernard and Torero, 2015; Lee et al., 2020). A large share of papers on rural electrification has relied on instrumental variables. Geographic variables such as groundwater availability or land gradient have been used by Rud (2012) and Dinkelman (2011), respectively. Chakravorty et al. (2016) and Lipscomb et al. (2013) estimate causality through an hypothetical grid extension uniquely based on least-cost principles. Another strand of studies such as Bensch et al. (2011) and Rao (2013) has used propensity score matching techniques.

This thesis follows Burlig and Preonas (2006) in using a regression discontinuity design which exploits the exogeneous variation introduced by the population requirements of the RGGVY program. RDD has been rarely used before to study electrification, but often applied in the recent literature on road infrastructure. Using the Indian PMGSY[4] road expansion program, Adukia et al.; Mukherjee (2012) find positive effects of road connection on school enrolments and test scores. Asher and Novosad (2016, 2019b) suggest the program has facilitated the movement of workers out of agriculture and an equivalent increase in wage work.

Many studies have pointed out that grid connection of a rural village does not translate into electrification of every households. Chakravorty et al. (2016) estimate that in the Philippines, village electrification leads to an increase in the household connection rate of 22.7 percentage points, while the share of households with electrification, conditional on a village being connected to the grid is only 43%. Khandker et al. (2014); Banerjee et al. (2014) report that affordability and poor quality of energy supply in India, resulting in frequent power outages, negatively affects households' incentive to connect to the grid and consume more electricity. Results from Kenya confirm that households' demand for connection to electricity is very elastic to price and lower than anticipated by policymakers. Among newly connected households, average electricity consumption is very low (Lee et al., 2020). Baskaran et al. (2015) mention that 35% of firms indicated access to reliable electricity as the number one obstacle facing their business in the World Bank Enterprise Survey of Indian businesses of 2006. In 2001, 75% of Indian citizens ranked electricity as an important problem in their lives.

As in Burlig and Preonas (2006), I use satellite images of night lights to measure whether the RGGVY program increased consumption of electricity. Since lightning is the most common usage of electricity, nighttime brightness may be a good proxy for electrification rate (Min, 2011). The use of remote sensing on night lights is a relatively new technique which has gained attention of economists since its introduction in the literature by Henderson et al. (2012). The reasons are many: the correlation between nightlight intensity and GDP is well established, also at the subnational level (Gennaioli et al., 2014; Prakash et al., 2019a; Bundervoet, 2015). The data capture informal activity, are available at very short time intervals and almost real-time. They are inexpensive to obtain compared to a household survey and allow to perform analysis at very low geographic specifications. However, Beyer et al. (2018) underlines that electricity may also be employed in agriculture for activities such as pumping water, which do not generate luminosity and are thus less well captured by night lights. Min et al. (2013) reveals that electrified villages are consistently brighter than unelectrified ones in Mali and Senegal, but that the correlation between nighttime brightness and household electricity use and access is low.

Several studies considering the effects of electrification on local employment have underlined the importance of considering migration flows induced by improved infrastructures. In my case, out-migration from villages surrounding towns towards the denser population centers may be driving the negative effects of electrification on school enrolments. Dinkelman (2011); Lipscomb et al. (2013) prove that, while electrification makes some places more desirable than others, the positive develop-

[4]Pradhan Mantri Gram Sadak Yojana

ments in the labor market can not be fully explained by in-migration. Dinkelman and Schulhofer-Wohl (2015) show why migration and related congestion externalities should be accounted for when estimating the benefits of rural electrification.
Finally, this thesis also considers the level of enforcement of the program requirements and sheds light on the potential role of corruption in the assignment of program funds. The literature on the political economy of electrification in India recognizes that even in a democratic state, there is room for politicians to strategically manipulate service provision (Besley et al., 2012; Baskaran et al., 2015). Ahlborg and Hammar (2014); Borang et al. (2016) study how poor institutional quality conditions progresses in rural electrification in Mozambique, Tanzania and India. Furthermore, Wilkinson (2006) documents cases of corruption allegations in large infrastructure investments in India.

3 Theoretical Framework

The theoretical framework which considers how rural electrification may affect education, centers on the trade-off between two major effects: returns to education and opportunity cost of schooling. Both stem from possible changes in the labor market.
First of all, household electrification allows a shift in sources of fuel for lightning and cooking. Kerosene lamps are mostly used for lightning in households with no access to electricity. The lamp produces weak light and is known to generate indoor air pollution. Electric bulbs are much cheaper, less dangerous for children's health and produce better light.
The shift to electric cookstoves and the possibility to use electric appliances both free up time previously spent on wood collection and household chores. This generates an endowment effect which may induce an increase in supply of labor to the market, especially for those whose time allocation is changed the most – women.
On the demand side, electrification opens opportunities for starting home-based activities. Conditional on other infrastructure levels, firms may locate production in more rural areas, where wages are lower.
Increased demand and supply of labor following rural electrification raise the opportunity cost of schooling for children and households. Poor households may have an incentive in letting children work for a wage and contribute to the household's income, rather than sending them to school. Alternatively, older children may be more able to substitute parents in household chores. We could expect the opportunity cost to increase with the labor market demand. At the same time, more possibilities in the labor market increase the returns to education. Better lightning and improved health could allow children to study longer at home and stay longer in school. Households' average income may benefit from a shift away from the agricultural sector, allowing children to attend school more. Which of these effects prevail, is ambiguous and thus subject of empirical research.
When considering the dynamics induced by rural electrification, it may be important to understand them in different contexts. In any economy, cities are an important driver for labor supply and demand, both within as outside the city borders. Work opportunities in and around cities are often more and better paid than those in rural areas. This may have implications on the household's schooling decision of their children, by raising its opportunity cost. On the other hand, a better labor market may incentivize households to send their children to school, in order to get a better education and better future perspectives. In other words, we could expect both the opportunity cost and the returns to education effects induced by rural electrification

to be of higher magnitude with urban proximity. Adding a spatial perspective to the analysis of the impact of electrification on education may serve well in better understanding its drivers.
Village electrification may also induce in- and out-migration, by making some places more desirable than others. The effect observed on school enrolments may thus mask what is actually a migration effect.

4 Background

4.1 Rural Electrification in India: from the independence to the RGGVY program

The Government of India has invested in electrification since its independence in 1947. However, the focus and pace of progress have changed throughout history. In 1947, only 1500 villages had access to electrification. Through the Electricity Supply Act in 1948, the creation of the electrical grid system was started, with which power was extended to semi-urban areas. The first Five Year Plan [5] (1951-56) set the focus on irrigation projects and village-level electrification. In the second part of the 50s, efforts were targeted to electrifying towns with population of at least 10.000. By the end of the 50s, 18.700 villages were connected to the grid. However, the rate achieved on the targeted towns remained very low.
Since the early 60s, electrification started to be seen as a crucial infrastructure for improved productive uses. Severe droughts and food shortages were a signal for the need to encourage the development of irrigation and commercial activities. The Green Revolution in India represented a period of drastic improvements in cultivation techniques starting in 1965. For the adoption of modern methods and technology such as high yielding variety seeds, irrigation facilities, pesticides and fertilizers, electrification of the agricultural sector was a prerequisite.
Households' grid connection was still considered as secondary. This started to change from the end of the 60s. In 1969, the Rural Electrification Corporation (REC) was established. The REC is a financial institution which aims at promoting investments in rural electrification, at that time primarily for agricultural production. In 1974, the access to electrification was included in the Minimum Needs Program. The program was "designed to assist in raising living standards and in reducing the regional disparities in development" (Planning Commission of India, 1973). This broadened the mandate of the REC to include rural households. Starting from the 80s, under the 6th and 7th Five-Year Plans programs were financed to address the low rates of rural electrification, with an emphasis on the poorest households. This period represented a steep improvement in village connections.
In the 90s, electricity started to be seen as core infrastructure for rural economic development. The Ministry of Power (MOP) and the Ministry of New and Renewable Energy (MNRE) were established in 1992 and 1994, respectively. The Electricity Act in 2003 mandates universal service obligation of the government to provide electricity supply to all areas, including villages and habitations (Bilolikar, 2004). The Act also aims at reforming the electricity sector. In 2004, the Government of India announced the ambitious goal of achieving universal electricity access by 2009 and fully meeting power demand by 2012.

[5] Five Year Plans are centralized and integrated national economic programs which set the objectives for the development of India in the following years.

The flagship electrification program Rajiv Gandhi Grameen Vidyutikaran Yojana (RGGVY) was launched in 2005 under the 10th Plan (2002 - 2007). It consolidated all the ongoing rural electrification programs under its umbrella. Its continuation in the 11th and 12th Plan (2007-2012, 2012-2017, respectively) was subsequently approved by the government, which subsumed it to the DDUGJY (Deendayal Upadhyaya Gram Jyoti Yojana) program in 2015.
The changing emphasis of electrification, first on agriculture, then shifting to productive activities and finally to households, is reflected in the different official definitions of 'electrified village' adopted throughout the years. Until 1997, a village was considered electrified whenever electricity was used within its boundaries for any purpose. This was partially modified to account for electricity used in 'inhabited localities' within their boundaries and was in place until 2004. Since then, in order to be declared 'electrified', a village must provide basic infrastructures such as distribution transformers and distribution lines. Public spaces like schools, panchayat offices[6] and health centers must be connected to the grid, as well as at least 10% of the households.

4.2 From the RGGVY program to the present

Under the 10th Plan, the RGGVY had the objectives of granting access to electricity to villages and habitations[7] with more than 300 inhabitants, installing small generators and distribution networks where grid extension was not cost-effective (off-grid) and providing free electricity connection to households below the poverty line (BPL). The program guidelines explicitly indicate irrigation pumpsets, small and medium industries, village industries, health care, education and IT among the sectors which would benefit from rural electrification expansion. Objectives of the program also include electrifying public places like schools, panchayat office, health centers etc. More broadly, overall rural development, employment and poverty alleviation are meant to be facilitated. The REC was given the responsibility to implement the program and disburse the funds for the projects to the concerned state governments (Ministry of Power, 2005). 90% of the project costs were covered by capital subsidies, the remaining 10% was financed through loans provided by the REC. Above poverty line households were given the possibility to purchase the connection to the grid and all connected households were required to pay for their energy consumption (Burlig and Preonas, 2006).
Under the 11th and 12th Plans, the population threshold was decreased to 100. Approval of projects by the government was based on the submitted Detailed Project Reports (DPRs). This implementation proposal had to be presented to the REC by the concerned state government. The DPRs grouped village-level proposals at the district level, collecting information such as number of households (and BPL households) to be connected, number of habitations, public places to be connected and size of electricity infrastructure to be installed. After reviewing the DPRs, the REC approved projects and disbursed funds to the states.
According to Burlig and Preonas (2006), funds for projects approved under the 10th Plan were disbursed between 2005 and 2010, the vast majority of which was released by 2008. 439.800 villages of the 593.732 listed in the 2001 census (74%) were connected to the grid by March 2005 (Planning Commission, GoI, 2011). As

[6]Panchayat raj are village councils.

[7]Administrative divisions in India are (in decreasing order): states, divisions, districts, subdistricts, blocks, villages, habitations.

of 1 September 2009, these were 85%. Funds under the 11th Plan were disbursed between 2008 and 2011.
Despite the overall significant progress, big disparities across states persist. In particular, village electrification rates in the states of Assam, Bihar, Jharkhand, Uttar Pradesh and Odisha were still well below the national average in 2009. These states also had the largest gap between village and household electrification rates in 2001. Even in states where village electrification already reached 100% in 2001 (Kerala, Andhra Pradesh, Nagaland), only 60% of the households had access to electricity. The national average was 50% (Nhalur et al., 2018). Overall, connection of BPL household remained largely incomplete (38.3% of the target in 2009) (Planning Commission, GoI, 2011; Nhalur et al., 2018), but progress has skyrocketed since then. As of March 2009, the RGGVY program had sanctioned 567 projects, for a total cost of 26.256,64 Rs crore (the equivalent of 3.1 billion euro as of today's conversion rate)[8]. The villages which were electrified were 54% of the ones initially targeted.
As of January 2019, village electrification has been completed, but the Prime Minister Narendra Modi's government announced it missed the self-imposed target of universal household electrification. Barriers to household adoption still exist and hinder the achievement of universal electrification. As the economic literature also finds, poor quality of energy supply (frequent power outages) negatively affects the household's decision to connect and pay for consumption (Khandker et al., 2014). Banerjee et al. (2014) also cite India's electricity pricing policies and household characteristics as decisive barriers halting the take-up of electricity. These factors appear to be more relevant in rural areas than in urban areas.

4.3 Education System in India

The current Indian education system is divided in Elementary School (grades 1 to 8), Secondary School (9 to 11) and Higher Education (12 to 18). The elementary education is divided into Lower Primary (1 to 5) and Upper Primary (6 to 8) school. At the end of each school year, students' learning progress is assessed through exams. The 2009 Right to Education Act stipulates that schooling is free and compulsory for all children aged 6 to 14 (Elementary School). However, ages of enrolment and completion vary (Ward, 2007).
Schools and other educational institutions in India are owned either by the government (central, state or local bodies) or by the private sector (individuals, trusts or societies). Private schools can be 'government-aided' and receive financial support from the government. Enrolment fees may be collected, but most of them are free to students. Those that are 'unaided' support themselves most commonly through student fees. At the Lower Primary level, 58% of schools were government schools in 2016, 69% at the Upper Primary level (British Council, 2019).
As it will be shown in Section 6, primary school enrolments in India have been increasing from 2005[9] to 2011. However, the DISE data showcase a decrease in the total enrolment levels from 2011. 2017 Lower Primary enrolments return to the 2005 value of 120 million students after having increased to 140 million in 2011. Upper Primary enrolments are still lower, but have been steadily increasing from 40 million in 2005 to 65 million in 2017. Enrolment levels have the drawback of not conveying the image of the ratio of children going to school compared to the school-age population. The aggregate data from the World Bank show an increase in gross school

[8] 1 lakh : 100.000 units; 1 crore : 10.000.000 units
[9] 2005 is the first available year in the DISE dataset

enrolment rates (World Bank, 2020c) but a decrease in the population aged 0-14 starting in 2012 (World Bank, 2020a). The share of population aged 0-14 is steadily decreasing since the mid-60s (from 40% to 27% in 2018). This indicates that the decrease in enrolment levels may be attributable to a decrease in the school-age population in the last years.

5 Data

I use multiple data sources combined at the village level. While Burlig and Preonas (2006) construct the panel data from scratch, I take advantage of the tremendous work done by Asher et al. (2019b) in constructing the Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG). This dataset is the result of a very time-intensive process of data merging at the village level across different sources. I link this data to village-level enrolments, habitation-level census and funds allocation data. This section describes the different sources, the relevant outcomes and illustrates the construction of the main dataset.

5.1 SHRUG

This version of the data became available in September 2019 and collects demographic, socioeconomic, firm and political outcomes at the village level for the period 1990-2018 (Prakash et al., 2019b; Jensenius and Verniers, 2017). It also includes remote-sensing data on nighttime brightness, forest cover and village ancillary (Asher et al., 2019a; Asher and Novosad, 2019a; Henderson et al., 2011a; Dimiceli et al., 2015). There are three main advantages in using this dataset. First, the SHRUG is a high geographic resolution census. Covering almost each one of India's 650.000 [10] towns and villages, it allows to take advantage of local variations and study socioeconomic developments at a very small geographic specification. National Sample Surveys are often only representative at the state or district level and do not constitute longitudinal data (Asher et al., 2019b). Second, the backbone of the SHRUG is a list of unique village identifiers (shrids) which allows to link villages' census codes to each other from 1990 to 2013. Because of the substantial number of geographical units which have merged and split across the different periods and the frequent inconsistencies in listed names, linking different waves of the Population and Economic Census is a very time-intensive process. In most cases, shrids match to single towns and villages in all census waves. However, when units merge in between any census years, they are merged in all periods in order to allow consistent analysis at the village level. Third, the SHRUG pools diverse data sources together, offering researchers the opportunity to work on a wide range of research questions. Furthermore, the SHRUG provides keys which allow to easily link the unique village identifiers to other data sources not yet included in the dataset. This represents a high potential for growth in the amount and kind of data assembled.

5.1.1 Population Census and Village Amenities

The SHRUG collects the most recent Indian Population Censuses of 1991, 2001 and 2011. They contain demographic data such as the number of households and population in various social groups and the share of literate population at the village

[10] Because of the aim of the SHRUG to consistently follow villages over time, splits and merges between geographical units modify the number of villages contained in the dataset.

level. Moreover, the Population Census also publishes village and town directories which contain data on local amenities. These include infrastructures (road connection, electrification status etc.), the distance from the nearest town, the number of medical facilities and schools located within the village boundaries. As pointed out by the authors, village electrification as reported in the population census in 1991 and 2001 has a poor correlation with the actual availability of electricity in the village. Electrification variables are missing for 60% of the villages. Starting from 2011, the data also contain the number of hours electricity was available per day by sector (agriculture, domestic, commercial and all uses). These values seem more reliable as proxies for electrification status of the village but are not available in previous years. Moreover, as reported in Section 4.1 the official definition of 'electrified village' changed in between census years, making the binary variables for power supply not comparable across periods. The SHRUG also includes some variables from the Socioeconomic Caste Census (SECC) undertaken in 2012. More precisely, the authors construct village-level variables for mean per capita consumption and share of households whose main source of income is agriculture.
The population census data provides me with the running variable for the RDD estimation strategy. In fact, the RGGVY guidelines mention the 2001 census as the source of data on which eligibility is based on. I use the 2011 population levels to test for migration across villages. Moreover, I use the distance to nearest town to explore heterogeneity in the intention-to-treat effect by proximity to urban centers. Share of literate population and belonging to low castes are used as control variables.
The original 2001 and 2011 population censuses included in the SHRUG contain around 585.000 and 591.000 unique shrids (consistent village identifiers), respectively.

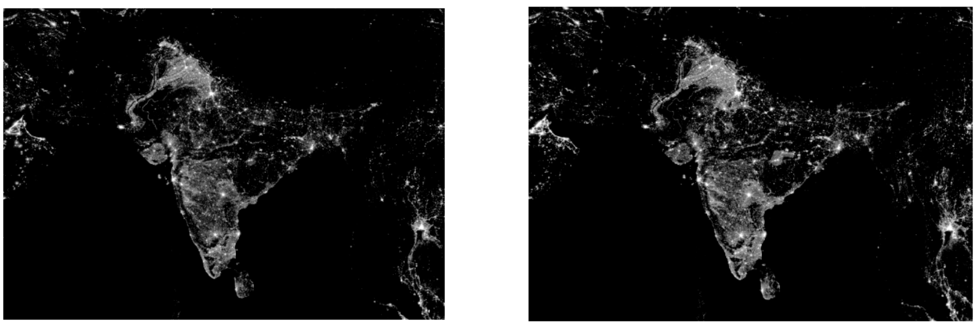
5.1.2 Night Lights

As mentioned in Section 2, the binary variables from the Population Census village amenities are not appropriate to detect actual change in households' electricity availability, access and consumption. There is no dataset available on village-level energy consumption in India. Since lighting is the most common employment of electricity for newly connected households (World Bank, 2008), using remote-sensing data on night lights represents a valid alternative proxy (Henderson et al., 2011b).
Under the Defense Meteorological Satellite Program – Operational Line Scan (DMSP-OLS), the American satellites photographed the earth daily during night hours from 1993 until 2014. A first elaboration from the NOAA (National Oceanic and Atmospheric Administration) translated each pixel's luminosity (approximately 1 square km at the equator) into a value (DN – Digital number) between 0 and 63, proportional to average observed luminosity.
Asher et al. (2019b) matched Indian villages and towns shapefiles to the NOAA night lights data and transformed them into a village-year-level dataset. While Henderson et al. (2012) point out that the use of the scale results in top-coding of the brightest pixels in urban centers, the issue of bottom-coding is for my research question more relevant. Since the analysis focuses on rural villages, where luminosity is much lower, there is a risk that the satellite sensors are not sensitive enough to catch the change in energy consumption. Bottom-coding would lead to underestimation of the effects of the RGGVY program on electricity consumption. However, these data have been used already in the economic literature and display a good correlation with rural

electrification. Min and Gaba (2014) compare DMSP-OLS night lights with rural electrification levels as reported in a village-level survey in Vietnam. They find that an increase in lightning from streetlights and homes is accurately detectable in the brightness from outer space. Min et al. (2013) and Min (2011) confirm these results in Senegal, Mali and India. Nevertheless, concerns remain that night lights fail to capture changes in electricity consumption in the agricultural sector (Beyer et al., 2018). Thus, images from satellites constitute a lower bound for usage of electrification in rural villages.
The NOAA releases three different elaborations of the night lights data. The visible night lights contain the yearly average luminosity level of each pixel. The calibrated night lights are a more processed version of the average visible lights. Based on Elvidge et al. (2014), calibrated lights do not include luminosity generated by fires or other sporadic events. Finally, the average visible × percent lights multiply the average visible DN value by the frequency with which it is observed. Because of the rural context in which the analysis takes place, my preferred measure of night lights is the maximum recorded brightness among the pixels within each village's boundaries. In this way, the analysis is targeted at the populated part of the villages, avoiding averaging brightness across unlit fields which typically surround the village (Burlig and Preonas, 2006).
Following Burlig and Preonas (2006) I also undertake an additional step to remove measurement error from the night lights. As described in more detail in Appendix A.2, I construct a projection of the night lights values for each year on the values of two years before and after. This aims at removing the random noise that cannot be explained by the brightness observed in adjacent years. Figure B.6 illustrates how projected maximum, average and average calibrated night lights compare to each other over time. I perform robustness checks on non-projected maximum brightness and other night lights measures. The night lights data included in the SHRUG contains originally 530.00 villages.

Figure 1: DMSP - OLS Nighlights for India, 2001 and 2011

(a) 2001 (b) 2011



Source: Burlig and Preonas (2006)

5.2 DISE

The District Information System on Education (DISE) collects a wide range of data on all Indian primary schools and is maintained by the National University of Educational Planning and Administration. Enrolment levels (by gender, by class and by social class), facility characteristics (for example number of classrooms, toilets, computers, availability of electricity and drinking water) and general data (school

type, language of instruction, school management etc.) are available starting from 2005 until 2017[11]. The Right to Education (RTE) data contain information such as whether the school is approachable by road, the status of the Mid-day Meal (MDM) program, Continuous and Comprehensive Evaluation (CEE) implementation status[12], number of enrolled children requiring special training and enrolled within the 25% quota established by the Right to Education Act[13]. Meaningful information on examinations at completion of Class 5 and Class 8 is available in 2017. The DISE also reports the number of children who pass the exam with a grade higher than 60%. These are on average about 65% of the students appearing at the C5 exam and 53% of those appearing at the C8 exam. Therefore, I use this information to proxy for the intensive margin of schooling.
Schools can be linked across years thanks to a unique numerical school code. I aggregate the data at the village level and use the first available year (2005) as baseline, 2011 and 2017 as medium- and long-run outcomes, respectively. My preferred outcomes are the total enrolment levels. I also perform the analysis separately by gender and grades. I aggregate values at the Lower Primary and Upper Primary school in order to explore whether electrification had different effects on younger or older children. In order to study the presence of a change in children's school efforts in the intensive margin, I use the number of students who passed the end-of-year exams with high grades. Although enrolment rates would be a more indicative measure of school attendance, information on the primary school-age population is not available at the village level. Originally, the DISE contains 1.1 million schools across all 36 states and union territories in 2005, 1.4 million in 2011 and 1.5 million in 2017. Of these, 753.400 schools form a balanced panel dataset.

5.3 National Rural Drinking Water Program

Under the RGGVY program, eligibility for funds was determined based on the 2001 habitation population (Burlig and Preonas, 2006). Habitations are sub-units of the villages and the smallest geographic specification existing in India. Villages with population larger than 300 may be mistakenly considered as eligible if they contain more than one habitation within their boundaries. If it is the case that none of the habitations exceeds the population cutoff, the village is not eligible despite counting more than 300 inhabitants overall. In order to overcome this issue, I use in the analysis only villages which are constituted by exactly one habitation (from now on *single-habitation villages*). In this subsample, the 2001 population census corresponds to the relevant population value which determined eligibility to the RGGVY program.
The only population census available at the habitation village is the one administered by the Ministry of Drinking Water and Sanitation, conducted for the National Rural Drinking Water Program (NRDWP) in 2003 and 2009. These had the purpose of assessing the availability of drinking water for all rural households in India and it is listed among the RGGVY implementation documents (Ministry of Power, 2011).

[11]The year refers to the one at the beginning of the school year (June/July).

[12]The Mid-day Meal scheme was launched in 1995 with the scope of improving nutrition for school-age children. It provides free lunch in Lower and Upper Primary schools. The Continuous and Comprehensive Evaluation was introduced in 2009 and aims at changing the way students are assessed in schools. Instead of single evaluations at the end of the year, students are evaluated throughout the whole academic year. Furthermore, children's ability includes not only school performance but also soft skills.

[13]reserved to children from lower castes

I am only able to obtain the 2009 habitation census from the Open Government Data Platform of India[14]. The dataset contains a list of all Indian villages and their habitations. The NRDWP habitation census originally consists of around 580.000 villages, which contain on average 3 habitations. 50% of these villages are single-habitation villages.

5.4 State-wise summaries and list of treated villages

As explained in Section 4.1, the RGGVY program was carried over from the 10th Plan to the 11th and 12th Plan. Funds were disbursed at the district level, which could apply for them more than once in different Five-Year Plans. Furthermore, different population cutoffs were applied over the years. While the 10th Plan required habitations to have population greater than 300, the cutoff was decreased to 100 under the 11th and 12th Plan. Failing to address this issue would result in biased estimates of the intention-to-treat effect. Consider the case where the estimation subsample includes villages located in districts which received funds under both the 10th and 11th Plan. Funds under the 10th Plan were disbursed from 2005 until 2010. Subsidies under the 11th Plan were paid out from 2008 until 2011. When analyzing the long-term effect (2017) of the program under the 10th Plan, villages with population between 100 and 300 fall into the control group. Nevertheless, these villages were eligible under the 11th Plan and could bias the result if they received the funds.

To overcome this threat, I restrict the analysis to the villages located in district which received funds exclusively under the 10th Plan. For this purpose, I use state-wise summaries published on the program website[15]. These are progress reports which document the districts which received funds from the RGGVY program and under which Plans. Figure B.8 shows the distribution of villages by Plans received by the relative district. Contrary to what reported in Burlig and Preonas (2006), these data show that the majority of the villages are located in districts which received funds under multiple Plans. Only about 10% of the villages are in districts which were treated uniquely under the 10th Plan.

From the same source I web-scrape the list of villages in which electrification works were completed under the 10th and 11th Plan[16]. This data would provide me with the most accurate treatment indicator. Nevertheless, because of the already discussed reverse causality and endogeneity of treatment allocation which characterize investments in infrastructures, relying on it would introduce omitted variable bias in a standard OLS estimation. A correlation between the error term and the treatment variable would be introduced by observable and unobservable characteristics of the village which influence both funds allocation and enrolments. This would violate the assumption of zero conditional mean of the error term underlying an OLS estimation.

These data contain the state, district, block and village names together with the 2001 census codes. The latter allows me to link the list to the SHRUG dataset. I further merge them to the habitation census and to the state-wise summaries on funds allocation.

I use this data to check the level of enforcement of program requirements. After restricting the list to districts which received funds under the 10th Plan only (for

[14] https://data.gov.in/, accessed in December 2019.
[15] http://www.ddugjy.gov.in/, accessed in January, 2020.
[16] http://www.ddugjy.gov.in/xxicompleteplan, accessed in January, 2020.

which the 300 population cutoff applies) and to single-habitation villages, I would expect to find no units with a 2001 population level lower than 300 if the program requirements were strictly enforced. Results are shown in Section 7.5.

5.5 Data preparation

The preparation of the data is a non-straightforward, time-intensive process which presents several threats due to the quality of the data. Both the NRDWP habitation census and the DISE contain no village identifiers. This means that I must follow a fuzzy match procedure on village names in order to merge the datasets. This procedure is further complicated by the presence of many different languages within India (Hindi, Bengali, Marathi to name the most spoken ones) and their ambiguous transcription to English. Furthermore, changing geographic boundaries and misspelling of village names across years increase the chances of introducing measurement error in the data due to false matching. At the expense of the sample size, I try to limit this possibility as much as possible, by implementing a rather conservative matching strategy. The whole process of data preparation and resulting datasets is described in more detail in the Appendix A.
For the analysis, I keep those villages which did not merge or split across census years. This is done by comparing how 2001 census codes merge to the SHRUG dataset. Whenever a shrid is matched to multiple 2001 census codes, it means the villages merged between 2001 and 2011. This implies that the 2001 population reported by the SHRUG is also aggregated at the resulting higher, merged level. This would not correctly represent the population level on which eligibility was based. These observations represent less than 2% of the villages contained in the 2001 Population Census and are thus dropped.
After merging all datasets together, I further drop observations which could bias the estimation. To avoid including villages which have been treated under other Plans into the control group for Plan 10, I keep only those units which are located in districts which received funds exclusively under the 10th Plan. As shown in Figure B.8, these represent around 10% of the observations. I use for the analysis only the single-habitation villages, for which the 2001 Population census represents the actual measure on which eligibility to the RGGVY program was based. These villages are little more than half of those possibly treated under the 10th Plan only. Finally, to reduce the risk of including false matches in the analysis, I keep only those villages for which one unique possible match was found by the fuzzy merge. In the subsample of single-habitation villages in districts receiving funds under the 10th Plan, these represent the 96%. The final dataset includes 13 of the 36 states and special territories, 56 districts and 16.516 villages. As Section 7.1 explains, each regression will be run on a different subsample, according to the optimal bandwidth chosen relying on Calonico et al. (2014a).

6 Descriptive Statistics

For sake of exposition, this section presents summary statistics using an average bandwidth size of 150. Table 1 allows to compare baseline (2001) village characteristics contained in the SHRUG by subsample and over time.
As shown also in Figure B.8, the number of villages located in districts which received funds under the 10th Plan only are around 10% of the full sample. Interestingly, villages located in 10th Plan funds-receiving districts are comparable to the full

dataset on all village variables, except for availability of electricity for the agricultural sector. While in the full sample the 48% of villages have access to electricity for agricultural purposes at the baseline, this is true for only 21% of the villages in funds-receiving districts. This signals that funds allocation may have been endogenous to village characteristics, stressing the need to apply econometric models able to identify causality. Single-habitation villages are also comparable to the full sample, in size as in characteristics. Although this might be surprising, single-habitation villages are not necessarily smaller villages, rather represent a different kind of administrative division. There is a high heterogeneity in habitations-to-village ratio across states. In the full dataset, a village contains on average 3 habitations. In states like Haryana and Mizoram, most of the villages are single habitations, while in Kerala a village contains on average 9 habitations. Smaller villages within an average bandwidth of 150 around the 300 cutoff have on average less schools and lower access to electricity, especially for cultivation. They are also less likely to be connected by an all-weather road. The share of workers relying on agriculture as main source of income and the data on distance to town are only available in 2011. It is therefore not possible to compare these dimensions at the baseline. All I can observe is that in 2011 – after the Plan 10 funds for the RGGVY program were completely disbursed – a higher share of workers (50%) is employed in the agricultural sector in eligible villages compared to those in the full sample. Units within the 150 bandwidth are also on average more isolated than average. In general, the villages in the estimation sample have at baseline worse than average levels of infrastructures. Literacy rate and number of primary schools per village both slightly increased in 2011 compared to 2001. When considering changes in village characteristics over time, we should keep in mind that the dummy variables for access to electricity are not comparable across years, since the official definition of 'electrified village' has changed in 2004 (see Section 4.1).

Village electrification was already very heterogeneous across states at baseline. This is confirmed by the great cross-state variation in nighttime brightness shown in Table 2. The estimation subsample does not contain all Indian states, since some did not receive any funds under the 10th Plan. From this measure of electrification, there does not seem to be a consistent pattern in funds allocation under the 10th Plan. The subsample of villages located in funds-receiving districts include states where night lights were already very high (for example Arunachal Pradesh), as well as very low (for example Jharkhand). Nevertheless, within states, most of the districts which received funds under the 10th Plan were darker at night than the state average. As I would expect, because of their size and infrastructure levels, smaller villages within the 150 bandwidth are on average less bright than the full dataset. For a full overview of the night light values by state and over time (1994-2013) see Table B.9. For a graphical comparison of the different measures of night lights at the aggregate level (whole India) see Figure B.6.

Summary statistics of village-level enrolments, exam results and school amenities are presented in Table 3. In this case, I use 2005 as the baseline year, as it is the first year in which the DISE dataset is available. Panel A presents village-level statistics on enrolments and examination results. Because of their smaller population size, villages within the 150 bandwidth count much less primary school students in 2005 than average. The ratio boys to girls remains constant (around 1.13) across subsamples. As represented graphically in Figure B.7, total enrolments in India are highest in 2011 and decreased since then. In the subsample within the bandwidth, total enrolments steadily decrease over time. While Lower Primary school enrolments are more than three times higher than Upper Primary school enrolments at baseline, they follow opposite trends over time: Lower Primary enrolments are steadily decreasing over time while Upper Primary enrolments are increasing.

As pointed out in Section 4.3, Indian demography is an important factor to keep in mind when analyzing school enrolments. As the World Bank shows (World Bank, 2020a), the school-age population is decreasing since 2012. Furthermore, another relevant aspect to consider is the school retention rate between Lower Primary and Upper Primary school. For a full overview of village-level enrolments and number of schools over time (2005-2017), see Table B.10.

Panel B of Table 3 describes village-level school amenities. Only 25% of the schools in the full dataset had access to electricity in 2005. The proportion is even lower among villages within the 150 bandwidth. Nevertheless, availability of electricity increases fast over time, reaching 76% in 2017 among 10th Plan, single-habitation villages with 2001 population within the bandwidth. It should be noted that access to electricity does not imply that power is available and used in the school. In these villages there were also less schools at baseline (1.3 per village). Only 74% of them disposed of drinking water within its premises. Surprisingly, the schools in these villages reported in 2005 a higher number of computers than average. Over time, the number of schools per village has increased, as well as access to electricity and drinking water. The decreasing share of schools approachable by road raises some concerns about the accuracy of these data.

Section 7.3 will explore whether there is heterogeneity in the effect of village electrification by geographic location. It is therefore important to understand whether baseline village characteristics are also different in proximity and further away from town. Table B.11 and Table B.12 present these statistics by proximity to town. Villages located closer to a town with at least 10k inhabitants have on average a lower share of population belonging to the lower castes and a slightly higher literacy rate. Electricity availability is also higher in proximity of towns than in more isolated villages, both at baseline as in 2011. Only 36% of the villages further away from town were approachable by tar road in 2001.

Despite being comparable in population size, villages closer to town had higher total enrolments at baseline. Coherent with the SHRUG data, the share of schools with access to electricity is higher in proximity of towns (22%) compared to more isolated villages (13%).

Over time, total enrolments decrease more in the more isolated villages than in proximity to town. It would be important in this case, to consider whether there was a rural-urban migration in this period, which could explain part of the time trend.

Table 1: Summary Statistics - SHRUG Dataset

	All Districts		10th Plan		10th Plan single-hab.		10th Plan single-hab. 150-450 pop.	
	2001	2011	2001	2011	2001	2011	2001	2011
total population	1468.59	1721.85	1414.70	1602.74	1335.59	1504.12	294.08	298.45
	(2045.83)	(2409.54)	(2082.97)	(2413.71)	(2060.93)	(2368.46)	(204.19)	(87.23)
number of households	270.48	346.28	280.32	345.79	261.65	320.58	59.03	66.63
	(401.11)	(511.85)	(436.38)	(552.91)	(419.85)	(531.25)	(35.75)	(22.27)
area (in ha)	509.88	509.06	587.63	592.44	564.23	562.02	219.93	223.42
	(994.40)	(898.45)	(798.86)	(1160.17)	(726.44)	(721.81)	(292.21)	(534.83)
share of pop. SC or ST	0.35	0.36	0.35	0.36	0.33	0.34	0.37	0.39
	(0.31)	(0.31)	(0.30)	(0.30)	(0.29)	(0.30)	(0.36)	(0.36)
literacy rate	0.47	0.57	0.50	0.60	0.51	0.61	0.52	0.62
	(0.16)	(0.14)	(0.15)	(0.12)	(0.14)	(0.12)	(0.15)	(0.13)
number of primary schools	1.28	1.61	1.42	1.77	1.22	1.55	0.87	1.01
	(1.22)	(2.09)	(1.42)	(1.73)	(1.02)	(1.40)	(0.47)	(0.52)
number of middle schools	0.39	0.76	0.47	0.76	0.44	0.71	0.14	0.22
	(0.65)	(0.99)	(0.74)	(1.04)	(0.68)	(1.00)	(0.36)	(0.49)
number of secondary schools	0.15	0.29	0.19	0.31	0.18	0.30	0.02	0.05
	(0.41)	(0.65)	(0.48)	(0.70)	(0.48)	(0.68)	(0.17)	(0.24)
number of higher secondary schools	0.04	0.12	0.05	0.12	0.05	0.11	0.01	0.03
	(0.22)	(0.42)	(0.24)	(0.42)	(0.23)	(0.41)	(0.09)	(0.18)
number of colleges	0.01	0.02	0.01	0.02	0.01	0.01	0.00	0.00
	(0.10)	(0.17)	(0.10)	(0.16)	(0.08)	(0.16)	(0.03)	(0.04)
electric power for domestic use (0/1)	0.83	0.89	0.82	0.93	0.81	0.98	0.82	0.96
	(0.37)	(0.31)	(0.39)	(0.25)	(0.39)	(0.14)	(0.38)	(0.19)
electric power for agriculture (0/1)	0.48	0.69	0.21	0.63	0.31	0.76	0.18	0.53
	(0.50)	(0.46)	(0.41)	(0.48)	(0.46)	(0.43)	(0.38)	(0.50)
electric power for all end uses (0/1)	0.70	0.60	0.72	0.70	0.80	0.81	0.66	0.66
	(0.46)	(0.49)	(0.45)	(0.46)	(0.40)	(0.39)	(0.47)	(0.47)
approachable by tar road	0.57	0.69	0.57	0.73	0.63	0.74	0.45	0.52
	(0.49)	(0.46)	(0.50)	(0.44)	(0.48)	(0.44)	(0.50)	(0.50)
agriculture main source of income (0/1)		0.39		0.43		0.42		0.50
		(0.28)		(0.32)		(0.31)		(0.35)
distance from nearest 10k town (in km)		12.52		14.19		12.77		16.17
		(9.20)		(9.85)		(8.77)		(10.50)
distance from nearest 50k town (in km)		30.12		31.03		28.22		33.62
		(18.99)		(18.63)		(16.89)		(19.45)
distance from nearest 100k town (in km)		43.93		44.94		39.06		46.59
		(29.93)		(25.91)		(23.21)		(25.82)
distance from nearest 500k town (in km)		96.56		102.35		98.78		100.06
		(58.08)		(47.19)		(48.46)		(43.63)
Observations	303263	303263	31164	31164	16516	16516	3140	3140

Note: 'All Districts' sample is the one obtained after merging all datasets and keeping only the best quality matches. The '10th Plan' subsample includes only villages which are located in districts which received funds from the RGGVY program uniquely under the 10th Plan. The '10th Plan, single-hab.' subsample includes only sngle-habitation villages in 10th Plan districts. The last two columns include only villages with 2001 population within an average bandwidth of 150. Distance to town and share of agricultural workers are not available in 2001. SC and ST are Scheduled Castes and Scheduled Tribes (low castes).

Table 2: Summary Statistics - 2001 Night Lights

	All Districts					10th Plan					10th Plan single-hab.					10th Plan single-hab. 150-450 pop.				
	mean	sd	count	min	max	mean	sd	count	min	max	mean	sd	count	min	max	mean	sd	count	min	max
Jammu Kashmir	3.37	(5.57)	1875	0	63	3.68	(5.32)	617	0	37	3.46	(4.83)	362	0	37	4.18	(4.74)	110	0	33
Himachal Pradesh	3.69	(4.32)	5782	0	63	1.76	(2.96)	574	0	17	1.59	(2.72)	49	0	13	2.07	(3.54)	16	0	13
Punjab	11.86	(7.05)	7869	0	63															
Uttarkhand	2.94	(5.87)	6219	0	63	2.94	(5.87)	6219	0	63	4.13	(6.90)	2460	0	63	2.97	(5.70)	894	0	63
Haryana	11.69	(8.08)	5121	0	63	14.29	(9.40)	806	4	63	14.27	(9.41)	802	4	63	13.19	(5.86)	33	5	29
Rajasthan	4.22	(4.59)	27636	0	62															
Uttar Pradesh	9.39	(18.92)	44131	0	63															
Bihar	3.29	(10.89)	22497	0	63															
Arunachal Pradesh	62.96	(1.57)	1607	0	63	62.64	(4.76)	175	0	63	62.59	(5.10)	152	0	63	63.00	(0.00)	41	63	63
Manipur	6.99	(11.66)	108	0	63															
Mizoram	62.74	(4.07)	478	0	63															
Meghalaya	63.00	(0.00)	3391	63	63	63.00	(0.00)	342	63	63	63.00	(0.00)	280	63	63	63.00	(0.00)	70	63	63
Assam	5.71	(14.36)	9601	0	63															
West Bengal	4.92	(5.15)	3080	0	62	4.23	(4.04)	1209	0	55	3.97	(2.88)	599	0	31	3.26	(2.96)	83	0	18
Jharkhand	1.75	(6.05)	15918	0	63	0.60	(3.46)	3173	0	63	1.11	(6.65)	188	0	63	1.26	(7.85)	65	0	63
Odisha	5.10	(13.66)	20784	0	63															
Chhattisgarh	2.62	(5.80)	12511	0	63															
Madhya Pradesh	3.29	(3.90)	35922	0	63	2.05	(3.10)	1925	0	40	2.04	(2.74)	1234	0	38	1.60	(2.16)	465	0	20
Gujarat	7.27	(7.30)	13829	0	63	7.16	(8.37)	1968	0	63	7.75	(9.46)	1354	0	63	4.75	(3.27)	295	0	30
Maharashtra	6.35	(6.43)	30609	0	63	5.54	(6.68)	2549	0	63	5.61	(6.55)	2005	0	63	4.98	(8.34)	267	0	63
Andhra Pradesh	6.03	(4.55)	7552	0	63	6.37	(4.60)	4481	0	63	6.31	(4.87)	2424	0	63	5.85	(8.73)	240	0	63
Karnataka	5.46	(3.88)	17344	0	63	5.25	(3.26)	7117	0	63	5.14	(3.23)	4607	0	55	4.54	(2.08)	961	0	34
Kerala	10.42	(5.11)	924	2	59	5.50	(2.16)	9	2	8										
Tamil Nadu	9.52	(6.79)	8475	0	63															
Total	6.69	(12.88)	303263	0	63	5.48	(9.37)	31164	0	63	7.02	(11.05)	16516	0	63	5.70	(11.50)	3540	0	63

Note: This table shows projected max night light values obtained as explained in Appendix A.2. Each village is assigned a DN (Digital Number) which represents the highest brightness value detected in the year. This corresponds to the value of the brightest pixel within the village boundaries. This table reports state 2001 averages. DN are on a scale between 0 and 63. 'All Districts' sample is the one obtained after merging all datasets and keeping best quality matches. The '10th Plan' subsample includes only villages which are located in districts which received funds from the RGGVY program uniquely under the 10th Plan. The '10th Plan, single-hab.' subsample includes only sngle-habitation villages in 10th Plan districts. The last columns include only villages with 2001 population within an average bandwidth of 150.

Table 3: Summary Statistics - DISE Dataset

	All Districts			10th Plan			10th Plan. single-hab.			10th Plan single-hab. 150-450 pop.		
	2005	2011	2017	2005	2011	2017	2005	2011	2017	2005	2011	2017
A. Students per village												
Lower Primary, Boys	114.32	110.04	91.41	95.79	93.48	81.74	88.30	88.56	80.45	32.07	29.19	24.42
	(194.80)	(207.80)	(183.21)	(222.34)	(209.42)	(205.06)	(225.17)	(214.36)	(220.93)	(75.58)	(69.47)	(64.82)
Lower Primary, Girls	104.11	103.97	86.05	90.87	87.84	75.10	84.12	82.50	73.20	31.06	27.73	23.20
	(179.28)	(194.43)	(165.29)	(215.45)	(193.49)	(172.67)	(214.05)	(196.95)	(181.69)	(70.22)	(63.85)	(58.38)
Upper Primary, Boys	39.36	48.06	47.83	38.95	45.43	45.27	36.76	43.28	44.24	9.17	10.87	10.34
	(100.97)	(109.03)	(109.40)	(130.27)	(127.13)	(132.49)	(130.79)	(126.38)	(137.58)	(48.50)	(42.18)	(41.10)
Upper Primary, Girls	32.03	45.92	45.82	34.59	43.26	42.64	33.03	40.61	41.03	8.15	10.78	9.89
	(87.26)	(105.35)	(106.22)	(121.78)	(123.96)	(122.22)	(122.00)	(121.54)	(123.33)	(39.13)	(40.76)	(39.00)
Total, Boys	153.68	158.10	139.24	134.74	138.91	127.01	125.06	131.84	124.68	41.24	40.06	34.76
	(281.71)	(304.83)	(283.47)	(345.62)	(330.07)	(332.65)	(349.84)	(335.13)	(354.17)	(117.83)	(107.22)	(103.17)
Total, Girls	136.14	149.89	131.87	125.46	131.11	117.74	117.14	123.11	114.23	39.21	38.51	33.09
	(254.75)	(288.43)	(262.85)	(330.85)	(311.60)	(290.28)	(329.93)	(313.63)	(301.08)	(104.26)	(99.69)	(94.69)
Total	289.82	307.99	271.11	260.20	270.02	244.76	242.20	254.95	238.91	80.45	78.57	67.85
	(533.34)	(590.43)	(541.98)	(674.37)	(638.95)	(620.47)	(678.20)	(646.18)	(652.92)	(221.20)	(205.40)	(196.90)
Appeared at the C5 test			30.96			23.92			22.58			7.94
			(85.43)			(68.13)			(71.15)			(26.60)
Appeared at the C8 test			24.38			20.91			20.27			5.95
			(69.60)			(74.74)			(76.67)			(33.65)
Passed the C5 test			29.98			23.57			22.30			7.81
			(72.71)			(67.15)			(70.27)			(25.84)
Passed the C8 test			23.85			20.65			20.10			5.94
			(69.14)			(74.04)			(76.16)			(33.59)
Passed the C5 test with more than 60%			19.21			16.80			16.59			5.21
			(54.30)			(52.12)			(55.72)			(17.43)
Passed the C8 test with more than 60%			14.39			12.65			13.05			3.14
			(47.86)			(51.40)			(53.61)			(16.97)
B. School amenities												
Schools per village	2.07	2.40	2.52	2.22	2.60	2.69	1.91	2.28	2.38	1.33	1.45	1.47
	(2.74)	(3.15)	(3.26)	(3.67)	(4.14)	(4.35)	(3.51)	(3.97)	(4.17)	(1.56)	(1.79)	(1.83)
Number of Computers	0.12	0.40	0.67	0.22	0.39	0.63	0.27	0.48	0.83	0.49	0.28	0.42
	(2.97)	(3.68)	(7.50)	(6.79)	(1.84)	(2.19)	(7.39)	(2.15)	(2.59)	(12.50)	(1.86)	(2.14)
Classrooms need major repair (0/1)	0.11	0.08	0.11	0.12	0.11	0.14	0.10	0.10	0.13	0.11	0.13	0.16
	(0.24)	(0.19)	(0.21)	(0.24)	(0.22)	(0.25)	(0.22)	(0.22)	(0.24)	(0.26)	(0.27)	(0.30)
Electricity access (0/1)	0.25	0.44	0.62	0.28	0.66	0.79	0.33	0.73	0.85	0.17	0.60	0.76
	(0.39)	(0.46)	(0.44)	(0.40)	(0.44)	(0.37)	(0.43)	(0.42)	(0.34)	(0.37)	(0.47)	(0.41)
Drinking water in the premises (0/1)	0.84	0.95	0.97	0.77	0.94	0.97	0.77	0.94	0.97	0.74	0.92	0.95
	(0.32)	(0.18)	(0.14)	(0.37)	(0.21)	(0.15)	(0.39)	(0.22)	(0.16)	(0.43)	(0.25)	(0.20)
Residential Schools (0/1)	0.02	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.02	0.02	0.01	0.01
	(0.11)	(0.08)	(0.12)	(0.09)	(0.07)	(0.10)	(0.10)	(0.07)	(0.11)	(0.12)	(0.08)	(0.10)
Private (0/1)	0.04	0.06	0.08	0.03	0.05	0.07	0.03	0.05	0.07	0.01	0.03	0.03
	(0.14)	(0.17)	(0.20)	(0.11)	(0.15)	(0.18)	(0.11)	(0.15)	(0.18)	(0.09)	(0.13)	(0.14)
Co-educational (0/1)	0.98	0.99	0.99	0.98	0.99	0.99	0.98	0.99	0.99	0.99	1.00	1.00
	(0.11)	(0.09)	(0.07)	(0.11)	(0.08)	(0.07)	(0.12)	(0.09)	(0.07)	(0.10)	(0.05)	(0.05)
Approachable by all-weather roads (0/1)		1.00	0.88		1.00	0.81		1.00	0.88		1.00	0.79
		(0.00)	(0.29)		(0.00)	(0.35)		(0.00)	(0.30)		(0.00)	(0.39)
Computer-Aided Learning (0/1)		0.07	0.10		0.09	0.10		0.09	0.12		0.04	0.05
		(0.21)	(0.24)		(0.22)	(0.24)		(0.23)	(0.27)		(0.18)	(0.20)
Observations	253067	270735	259863	27623	27942	27071	15056	15093	14536	3186	3201	2997

Note: Lower Primary school includes classes from 1 to 5, Upper Primary school from 6 to 8. Co-education schools are those where both girls and boys are educated. Computer-Aided Learning is defined as the integration of technology in the education process. 'All Districts' sample is the one obtained after merging all datasets and keeping best quality matches. The '10th Plan' subsample includes only villages which are located in districts which received funds from the RGGVY program uniquely under the 10th Plan. The '10th Plan, single-hab.' subsample includes only sngle-habitation villages in 10th Plan districts. The last columns include only villages with 2001 population within an average bandwidth of 150.

7 Empirical Analysis

7.1 Identification Strategy

As pointed out in Section 1, reverse causality and endogeneity of investments in infrastructures make assessing causality non-straightforward. The population requirements included in the guidelines of the RGGVY program offer a source of exogenous variation in the likelihood of treatment which can be exploited through a regression discontinuity design. My estimation strategy builds on that in Burlig and Preonas (2006) and makes some adjustments.

The core of a RDD estimation strategy is the existence of a continuous running variable (or forcing variable) and a cutoff on which treatment allocation is fully or partly based. Units whose value of the running variable is higher than the cutoff are offered the treatment (treatment group), while those whose score is below the cutoff (control group) are not. Using a potential outcome framework, we can express $Y_i(0)$ and $Y_i(1)$ as the outcomes that would be observed for unit i under the control or treatment condition. Ideally, we would like to study $Y_i(1) - Y_i(0)$. The so-called fundamental problem of causal inference is that we are not able to observe both outcomes for the same unit i. We only observe the actual outcome:

$$Y_i = Y_i(0) \cdot (1 - Z_i) + Y_i(1) \cdot Z_i = \begin{cases} Y_i(0) \text{ if } Z_i = 0 \\ Y_i(1) \text{ if } Z_i = 1 \end{cases}$$

where Z_i is the treatment status of unit i. Thus, we rely on comparison between subgroups. In a RDD estimation, a good counterfactual is found in the units just below the cutoff. The estimation is based on the units within an optimal bandwidth h which includes the cutoff. The necessary assumption of comparability between units with very similar values of the running variable but on opposite sides of the cutoff c has been formalized by Hahn et al. (2001) as

$$\lim_{\epsilon \to 0} E[Y_i(Z)|X = c + \epsilon] = \lim_{\epsilon \to 0} E[Y_i(Z)|X = c - \epsilon] \text{ for } Z = \{0, 1\}$$

where X is the running variable. This implies that the running variable is allowed to be correlated with the potential outcomes, but this relationship is assumed to be smooth at c (Imbens and Lemieux, 2008). If the continuity assumption is satisfied, any discontinuity of the observed outcome at the cutoff is to be attributed to the treatment. For this reason, the running variable constitutes a 'local IV'. While the continuity assumption is by nature not testable, pre-treatment outcomes and covariates correlated to both treatment and outcome will be used as tests.

Other assumptions needed for validity of a RDD estimation include no manipulation of the running variable and no other ongoing program which sets the same threshold for eligibility. If the cutoff is publicly known, individuals may have an incentive in misreporting the value of the forcing variable in order to receive the treatment. This would create a discontinuity in the density of the running variable at the cutoff and invalidate the estimation strategy. In my case, the forcing variable (2001 population size) is determined before the program guidelines are announced (2005), thus eliminating this concern. I will prove this through a McCrary test (McCrary, 2008). Manipulation of treatment allocation by the entities responsible for the program implementation still remains a possible issue which will be addressed in Section 7.5. I am not aware of any other program which uses the 300 population threshold in the period of analysis.

In a sharp RDD, Z_i is a deterministic function of X_i. In a fuzzy RDD, $Pr(Z_i = 1|X_i = c)$ needs not change from 0 to 1. As Burlig and Preonas (2006), I rely on a sharp RDD estimation where I exploit the discontinuity in eligibility status of villages with 2001 population just above 300[17]. In this case, the RD estimator identifies the intention-to-treat effect of electrification. If the above-mentioned assumptions are satisfied, the average intention to treat effect at the cutoff is identified as:

$$\tau_{SRD} = E[Y_i(1) - Y_i(0)|X_i = c] = \lim_{\epsilon \to 0} E[Y|X_i = c + \epsilon] - \lim_{\epsilon \to 0} E[Y|X_i = c - \epsilon]$$

In practice, I estimate the intention-to-treat effect for the subsample of single-habitation villages which are located in districts which received funds under the 10th Plan (see Section 5.5).
In a regression form, this is obtained by estimating the following:

$$Y_{vs}^{t} = \alpha_0 + \alpha_1 Z_{vs} + \alpha_2 X_{vs} + \alpha_3 Z_{vs} \cdot X_{vs} + \alpha_4 Y_{vs}^{pre} + \alpha_5 W_{vs} + \eta_d + \nu_{vs} \tag{1}$$

$$\text{where} \quad Z_{vs} = \mathbb{I}\{X_{vs} \geq 0\}, \text{ for} \quad h^- \leq X_{vs} \leq h^+$$

Y_{vs}^t is the outcome of interest in village v in state s in the relevant outcome year t (2011 and 2017). X_{vs} is the 2001 standardized village population obtained by subtracting the cutoff value 300 from the 2001 population size. Z_{vs} is equal to one when the 2001 village population is greater than the cutoff, so that α_1 is the RD coefficient of interest. Y_{vs}^{pre} is the pre-treatment value of the outcome variable. W_{vs} are village-level, pre-treatment control variables. η_d is a district fixed effect and ν_{vs} is the robust error term.
While the inclusion of covariates is not necessary in a RDD, Calonico et al. (2019) and Frölich and Sperlich (2019) advocate the use of controls with the purpose of reducing imprecision and sample imbalances. Frölich and Sperlich (2019) suggest including covariates which are good predictors of the outcome or when sample limitations force researchers to use observations further away from the threshold. The covariates included should be pre-determined (pre-intervention) and continuous at the threshold. In my case, covariates included are the pre-treatment outcome, the baseline share of literate population, share of ST or SC population [18] and distance to nearest town with at least 10k inhabitants. Although this last control variable is not pre-treatment, location of the village is not affected by electrification. In Section 7.2 I test for the robustness of the results to the inclusion of covariates.
Since DPRs and funds are prepared and allocated at the district level and I only include treated districts in the analysis, I choose not to cluster the standard errors (contrary to Burlig and Preonas (2006)) but to use district fixed effects and robust standard errors instead. In this way I hope to account for the unobservable characteristics which may induce a local district government to set higher priorities on electrification of its villages – and to engage more in order to obtain the funds. In fact, such characteristics may also influence the importance given to rural primary education at the district level. Furthermore, controlling for district fixed effects aims at accounting for the spatial correlation between nearby villages within districts.
For the choice of the optimal bandwidth h I rely on Calonico et al. (2014a) and the related Stata package `rdrobust` (Calonico et al., 2014b). Since this optimal

[17]The data I collected would allow me to apply a fuzzy RDD. However, not all assumptions needed are satisfied. I thus return to the settings proposed by Burlig and Preonas (2006).
[18]SC and ST are Scheduled Castes and Scheduled Tribes (low castes).

bandwidth choice is data-driven, it is different for each outcome Y^t. In Section 7.2 I perform a sensitivity analysis to assess whether results are robust to the choice of the bandwidth size. Within each bandwidth size, I drop the top first percentile of the data, since baseline enrolment levels seem likely inconsistent with the population size of the village.

Equation (1) estimates a local linear regression on both sides of the threshold, allowing for different slopes. In Section 7.2 I show that results are robust to a second-order polynomial estimation. I use a triangular kernel, which weights observations closer to the threshold more than those further away.

Before turning to the results, I test the above-mentioned assumptions underlying a RD strategy. I show continuity of potential outcomes and included covariates by estimating eq. (1) on the 2001 and 2005 value of the outcomes and controls. The results are presented in the first three panels of Table B.13 and graphically in Figure B.9. I am not able to test the continuity of the pre-treatment values of the exam results, since this information is not available before 2017. None of the RD coefficients is statistically significant, indicating that the continuity assumption underlying the RD strategy holds. I also show that there is no discontinuity in the distribution of the running variable at the threshold. The McCray test (McCrary, 2008) performed in Figure B.10 estimates a statistically insignificant jump of -0.11 with a standard error of 0.12.

7.2 Main Results

In this section I present the main results of the analysis on nighttime brightness, enrolments and exam results at the village level. I perform several robustness checks and a falsification test. I then turn to the analysis of the presence of heterogeneity in the intention-to-treat effect of the RGGVY program.

7.2.1 Electrification

Before turning to school enrolment outcomes, I explore whether eligibility for the RGGVY program had a significant effect on consumption of electricity. As discussed earlier, connection of the village to the grid does not imply an increased household electrification and electricity consumption. Contrary to a binary variable, satellite nighttime images have the potential to capture the actual usage of electricity. I estimate eq. (1) on the 2011 level of projected maximum nighttime brightness.

Results are shown in Table 4 and graphically in Figure 2.

Surprisingly, and contrary to Burlig and Preonas (2006), I find a very small and negative, not statistically significant intention-to-treat effect on night lights. Compared to the 2001 nighttime brightness of a village at the mean of the bandwidth subsample, eligibility for the RGGVY program decreases the brightness of 2011 night lights by 1.79%.

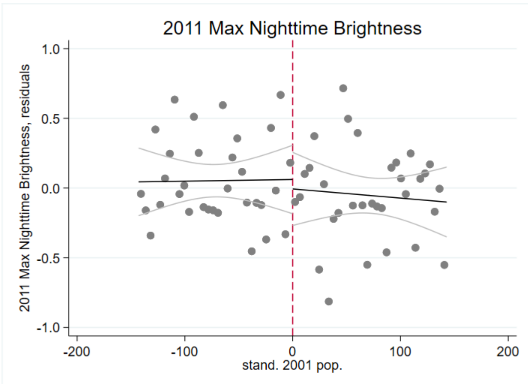
I test for the robustness of the result to the inclusion of covariates. Table B.15 shows that controlling for pre-treatment outcome level and village characteristics does not change the direction nor the statistical significance of the point estimate. Including the pre-treatment outcome value as control significantly decreases the magnitude of the estimated effect and its imprecision. This suggests that the 2001 brightness level is an important predictor of the 2011 level, thus justifying its inclusion in the specification according to Frölich and Sperlich (2019). Not taking it into account would overestimate the effect of eligibility for electrification. Controlling for village characteristics and district fixed effects further decreases standard errors.

Table 4: RD - 2011 Nighttime Brightness

Variables	Max Nighttime Brightness 2011
1[2001 pop. >= 300]	-0.102
	(0.197)
stand. 2001 pop.	-0.000
	(0.002)
1[2001 pop. >= 300]*stand. 2001 pop.	0.001
	(0.003)
Observations	3,361
R-squared	0.957
2001 Control	Yes
District FE	Yes
Baseline covariates	Yes
Bandwidth	143
Mean of dependent variable	7.436

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

Figure 2: RD graphical results - Nighttime Brightness



This figure shows the numerical results obtained in Table 4. The dots represent the conditional average residuals obtained from regressing the outcome variable on all covariates but the running variable. Each dot contains on average 50 observations, averaged in 5-person population bins. Lines are estimated separately on each side of the 300-person cutoff. The sample contains only unique-habitation villages located in districts receiving funds under the 10th Plan and with 2001 population within the optimal bandwidth chosen for estimation based on Calonico et al. (2014a).

The optimal bandwidth size of 143 is chosen following Calonico et al. (2014a). I check whether the result is sensitive to the choice of the bandwidth. I estimate eq. (1) using a fixed set of bandwidths from 50 to 300. Figure B.11 plots the obtained point estimates and their 95% confidence intervals. The effect remains negative and not statistically significant, regardless of which bandwidth is chosen for estimation. Following Gelman and Imbens (2019) and what is the standard practice for the implementation of a RD design, my estimation uses a local linear function of the running variable. I control whether the results are sensitive to the use of a second-order polynomial. In this specification I add a quadratic form of the running variable, also interacted with the eligibility dummy. The first column of Table B.16 displays the numerical results, while Figure B.12 shows them graphically. The RD coefficient does not significantly change in magnitude nor in statistical significance. Standard

errors slightly increase with the inclusion of a quadratic form of population.
I use the other night light measures from the NOAA satellites available in the SHRUG, together with the share of schools with electricity connection, to check whether the effect is sensitive to the kind of measure used. Table B.17 presents the results. They are consistently very close to zero and not statistically significant.
These results spark some doubts about the extent to which the RGGVY program brought an increase in electricity consumption of rural households. As previously discussed, satellite images are a powerful resource to measure electricity usage for lightning purposes. However, in rural areas the agricultural sector can also significantly benefit from village electrification. This kind of use of electricity would not be captured by nightlights. Thus, night lights constitute a lower bound for consumption of electricity. Furthermore, households' connection to the grid and consumption of electricity strictly depend on the quality of power supply. High disparity in power supply across states of India still exists nowadays. In Section 7.3 I explore whether there is heterogeneity in the effect of eligibility to the RGGVY program between states with above- or below-average supply deficit.

7.2.2 Education

I then turn to the analysis of the intention-to-treat effect on education. I first estimate eq. (1) on total village-level enrolments, both in the medium-term (2011) as in the long-term (2017). Results are shown in the first two columns of Table 5 and graphically in the first row of Figure 3.

Table 5: RD - Total Enrolments

	Total Enrolments		Lower Primary		Upper Primary	
Variables	2011	2017	2011	2017	2011	2017
1[2001 pop. >= 300]	0.480	-4.098	-0.395	-2.815	1.174	0.125
	(5.441)	(6.837)	(3.464)	(4.406)	(2.602)	(2.608)
stand. 2001 pop.	-0.059	0.013	-0.004	0.037	-0.086	-0.079
	(0.099)	(0.113)	(0.040)	(0.043)	(0.053)	(0.064)
1[2001 pop. >= 300]*stand. 2001 pop.	-0.012	-0.010	0.006	-0.005	0.085	0.077
	(0.132)	(0.146)	(0.054)	(0.064)	(0.077)	(0.080)
Observations	1,832	1,835	2,524	2,399	1,660	1,507
R-squared	0.538	0.385	0.429	0.285	0.474	0.388
2005 Control	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline covariates	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	95	103	131	136	86	85
Mean of dependent variable	62.41	52.17	46.91	37.82	15.37	13.31

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

Eligibility to the program slightly increases total enrolments in 2011 but decrease them in the longer term. On average, eligibility for the RGGVY program decreases total enrolments by 4 students in 2017. Consistent with Burlig and Preonas (2006), none of the effects is statistically significant.
Depending on the child's age, electricity may change both the opportunity cost of schooling as well as the returns to education in a different way. Older children are physically stronger and would be better able to substitute the parents in household

chores or work for a wage (Islam and Sivasankaran, 2014). They are also those who could take most advantage of improved learning and employment conditions. To take this into consideration, I estimate the effect separately on Lower Primary and Upper Primary enrolments. Although age at enrolment varies greatly even within schools, children enrolled in Lower Primary school are on average between 6 and 11 years old. The Upper Primary school is supposed to be concluded at the age of 14 (Ward, 2007). Columns 3 to 6 of Table 5 illustrate the results. A graphical representation is given in the second and third rows of Figure 3. The effects are slightly negative for younger children but slightly positive for the older ones. However, none of these coefficients is significantly different from zero. These results are also consistent with those of Burlig and Preonas (2006).
Besides school enrolments, eligibility for electrification may improve educational outcomes as well. Better and longer light hours at home, together with the time potentially freed up from helping with household chores may allow children to focus more on studying and thus obtain better results. In order to study this possibility, I make use of the variables contained in the DISE dataset which refer to exam results. This information is collected in the Indian states which adopt end-of-the-year examinations at the completion of the Lower Primary school (grade 5) and of the Upper Primary school (grade 8). The DISE collects information on the number of students who attend the exam, who pass the exam and who pass it with more than 60% of the grade. I use this last variable to study effects on the intensive margin of schooling[19]. Unfortunately, due to data limitations I am only able to study the 2017 examination results[20]. This implies that I am not able to test for the RDD continuity assumption on pre-treatment outcomes and that I can not control for the pre-treatment outcome level in the specification. For the latter, I use instead the number of children in the respective grade in 2005. Table 6 illustrates the numerical results. Graphical results are shown in the last row of Figure 3

Table 6: RD - Examination Results

Variables	Students passed the C5 test with high grades	Students passed the C8 test with high grades
1[2001 pop. >= 300]	-0.884*	-0.914**
	(0.468)	(0.390)
stand. 2001 pop.	-0.009	-0.006
	(0.008)	(0.005)
1[2001 pop. >= 300]*stand. 2001 pop.	0.044***	0.023***
	(0.012)	(0.007)
Observations	1,898	2,380
R-squared	0.189	0.182
2005 Control	Yes	Yes
District FE	Yes	Yes
Baseline covariates	Yes	Yes
Bandwidth	106	135
Mean of dependent variable	4.072	1.991

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

[19]It would be preferable to use the share of students passing with high grades instead of the absolute levels. The transformation does not result in plausible values. This in turn raises questions about the overall quality of the data.

[20]This information is collected in the DISE since 2009, but the quality of the data remain very low until 2017.

Eligibility for electrification seems to have decreased the number of children passing the examinations with high grades, both at the end of Class 5 as of Class 8. The effects are statistically significant and amount to little less than one student on average.

I perform on these educational outcomes the same robustness checks that I carried out on the night lights outcome. Table B.18 to Table B.25 display the robustness of all outcome variables to the inclusion of covariates. Once again, controlling for the pre-treatment outcome level greatly reduces the magnitude and the imprecision of the estimate. Including covariates changes the direction of the estimates for some enrolment outcomes, but not their significance. This does not hold for exam results. Not including controls seems in this case to underestimate the effect of eligibility to the program. As the RD coefficients become bigger and standard errors remain stable after the covariates' inclusion, the intention-to-treat effect becomes significant at the 10% and 5% level.

I also test for the sensitivity of these results to the choice of the bandwidth size. Figure B.13 illustrates how point estimates and confidence intervals change with different bandwidths. All enrolment outcomes remain insignificant, regardless of the subsample used for estimation. On the other hand, the effect on exam results is sensitive to the bandwidth choice. However, both results in Class 5 and Class 8 seem to be robust to two bandwidth sizes higher or lower than the optimal one [21].

Columns 2 to 9 of Table B.16 illustrate the robustness of the results to a second-order polynomial. The same are plotted in Figure B.14. The inclusion of the running variable in quadratic form increases the imprecision and changes the magnitude of some RD coefficients. The effect on enrolment outcomes remains insignificant, while that on exam results become insignificantly different from zero.

Since the distribution of these variables is left-skewed, I estimate eq. (1) on their log transformation. As shown in Table B.26, this results in intention-to-treat effects which are not statistically significant on any of the educational outcome variables.

I perform a falsification test in which I vary the subsample of estimation and on which I expect to find no effect of the 300 cutoff for the RGGVY program eligibility. Since eligibility was determined on the population size at the habitation level, the 2001 census does not represent the relevant population level when a village contains more than one habitation. Similarly, since the threshold was decreased from 300 to 100 under the 11th Plan, estimating eq. (1) on the subsample of villages located in districts which received funds under the 11th Plan should not result in a significant RD coefficient. Table B.27 compares the results between those on the 10th Plan, single-habitation villages and those obtained from the falsification tests. As expected, there is no significant discontinuity at the 300 cutoff for all outcome variables analyzed. This provides some evidence that there is no spurious correlation in the data around the threshold and that there is no other program which induces discontinuity at the 300 cutoff.

While the impact on exam results seems relatively sensitive to bandwidth size, polynomial order and log transformation, the results on enrolment levels are robust and insignificant.

These conclusions are in sharp contrast with the strand of literature which finds positive results on education (Khandker et al., 2014; Lipscomb et al., 2013; Khandker et al., 2013) and which has so far seemed to shape the common belief that electrification improves schooling (Kohlin et al., 2011). However, they are consistent with

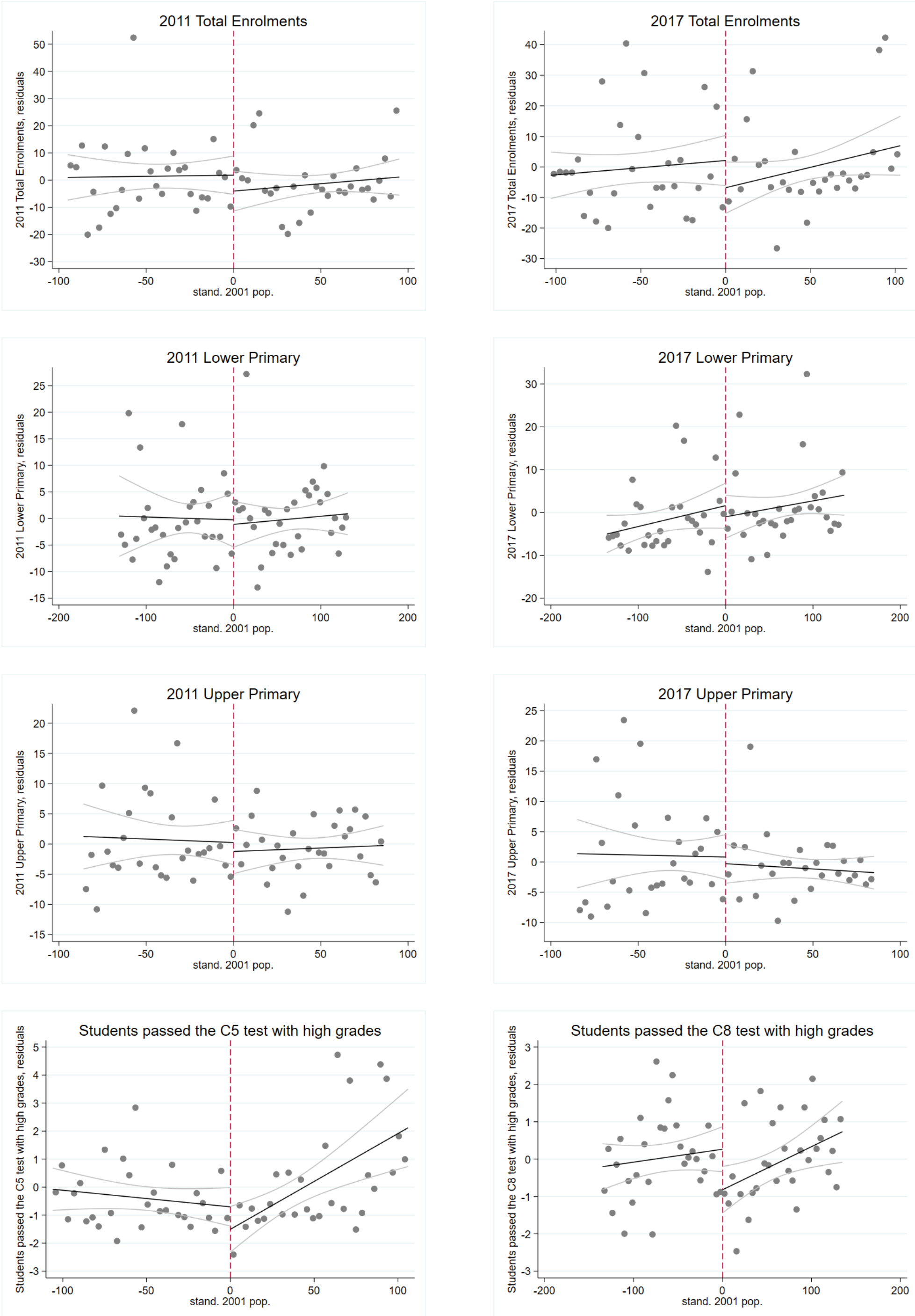
[21] Since the fixed set of bandwidth sizes increases in steps of 10, this means that results are robust within the range of bandwidth sizes $\pm$ 20 than the optimal one.

the literature which is more cautious on the impact electrification has on education. Bensch et al. (2011) find no effect of electrification on time spent studying, Lee et al. (2020) point to no improvement in children' test scores, while Burlig and Preonas (2006) do not encounter any significant effect on school enrolments.
The existing literature has also focused on the role which gender plays in determining the impact of electrification on schooling. In Section 7.3 I address this aspect and explore further the presence of heterogeneous effects by village characteristics.

7.3 Heterogeneity of the effect

As Burlig and Preonas (2006) mention, electricity consumption and households take-up of the service depend on the quality of power supply. As pointed out in Section 4.1, Indian states still largely differ in the quality of the service supply. The ESMI (Electricity Supply Monitoring Initiative) aims at providing transparency about the quality of electricity supply in India. The September 2019 report snapshots an average of 124 interruptions of supply in ESMI's rural consumer locations in Uttar Pradesh. 20% of these lasted more than 3 hours (Prayas (Energy Group), 2019). Even if households consumed more electricity in states with better power supply, the positive effect could be obscured if averaged with that of other states where power supply is low.
In order to address this concern, I split the sample by states with above- and below-average power deficit. According to Central Electricity Authority (2011), among the states present in my sample those with lower than average supply deficit are Karnataka, West Bengal, Gujarat and Haryana. Panel A of Table B.28 shows the results by subsample. Surprisingly, the intention-to-treat effect is negative in states with lower power deficit. However, none of the coefficients is statistically different from zero. Note also that the average 2011 maximum brightness is lower in states with higher power availability. Although this may be unexpected, it is still plausible. The maximum nighttime brightness is not an average measure. It records the brightest pixel within village boundaries, which may not be related to the quality of electricity supply. Average brightness, on the other hand, reflects the levels of power availability. The average 2011 DN value is 7.65 in states with better power supply and 5.5 in states with higher power deficit. These results seem to confirm that eligibility for the RGGVY program under the 10th Plan did not substantially increase consumption of electricity - at least not for lightning purposes.
I then turn to the heterogeneity of the effect on enrolment outcomes. A substantial part of the literature on electrification considers the distinction by gender. Women are those who mostly take care of household chores and spend time collecting fuels for cooking (Dinkelman, 2011). They are thus those whose time allocation is affected the most by the availability of electricity. If girls are more likely than boys to help with household chores, their time allocation also may change. With that, the incentive to enrol in school or work for a wage.
In order to account for this possibility, I run eq. (1) separately on girls and boys enrolments. In Panel B of Table B.28 I show the numerical results. The RD coefficient tends to be more positive for girls than for boys. Compared to the 2005 total enrolments of a village at the mean of the estimation sample, the effects on the medium-term represent a 0.63% increase for boys but a 3.42% increase for girls.

Figure 3: RD graphical results - Education variables



These figures show the numerical results obtained in Table 5 and Table 6. The dots represent the conditional average residuals obtained from regressing the outcome variable on all covariates but the running variable. Each dot contains on average 40 observations, averaged in 5-person population bins. Lines are estimated separately on each side of the 300-person cutoff. The sample contains only unique-habitation villages located in districts receiving funds under the 10th Plan and with 2001 population within the optimal bandwidth chosen for estimation based on Calonico et al. (2014a).

In the long term, total enrolments seem to decrease with eligibility to the RGGVY program. However, also in this case none of the coefficients is statistically significant.

These findings contradict the literature which predicts higher benefits from electrification for girls than for boys. Khandker et al. (2014) use a 2005 Indian household survey and finds an increase in school enrolments of the magnitude of 6% for boys and 7.4% for girls. My results indicate no significant effect of eligibility for the RGGVY program on school enrolments, neither for boys nor for girls.

Proximity to town is an important driver of labor demand. The availability of electricity may allow small family enterprises to produce in the village and sell in town. Commuting to the urban center from surrounding villages is also much easier than from more isolated villages. Conditional on the quality of the infrastructure, production of tradeable goods may be moved outside the urban center, where wages and land are cheaper (Deichmann et al., 2009; Foster and Rosenzweig, 2004). As a result, proximity to town exerts an important influence on the nature and amount of job opportunities available in the surrounding rural villages. Economic intuition would predict both returns to education and opportunity cost effect of electrification to increase with labor demand - in villages closer to town. Which of the effects prevails is still an empirical question. In Table B.29 I estimate the intention-to-treat effect of the RGGVY program separately by distance from the nearest town. The subsamples are defined by being above or below the median distance in the sample of single-habitation villages in 10th Plan districts. The SHRUG contains three measures for different sizes of town. The median distance from the nearest town with at least 10k, 100k or 500k inhabitants is 10.7km , 36km and 94.6 km, respectively. Results considering 10k towns are presented graphically in Figure B.16 and Figure B.17. The first column contains the estimates obtained in the subsample further away from town, while the second column shows results obtained on the subsample of villages closer than the median.

While I find no significant effects on enrolments and exam results in villages further away from town, I find robust and significant negative effects in proximity to town. In this subsample, eligibility for the RGGVY program decreases total enrolments by an average of 18 students in 2011 and 20 students in 2017. Compared to the 2005 enrolments of a village at the subsample mean, these represent a drop of 20% and 22.7%, respectively. Both estimates are statistically significant at the 5% level, but standard errors are quite high. While Lower Primary enrolments seem not to be significantly affected by electrification, the opposite is true for Upper Primary enrolments. Compared to 2005, Upper Primary enrolments decrease by 52% in 2011 and by 43% in 2017. The number of students passing the Class 5 and Class 8 examinations with high grades in 2017 is also significantly affected by eligibility for electrification. On average, it decreases by 1.7 and 2.4 students, respectively.

When considering bigger cities with at least 100k inhabitants (which are on average further away from the villages in the sample), significant negative effects on enrolments are still present. Total enrolments decrease on average by 12 students in 2011 (13.4% compared to the 2005 levels). The RGGVY program seems not to have had long lasting effects on total enrolments in this case. The RD estimate is of -10 students, but not statistically significant. However, Upper Primary school enrolments are still affected by the program. Compared to 2005, they decrease by 38.4% in 2011 and by 29% in 2017. Eligibility also negatively affects the number of students passing the Class 8 exam with high grades in 2017. Importantly, eligibility to the RGGVY program has mostly insignificant effects on enrolments when the average distance to town increases. The rural villages in the sample are in fact on average much further to towns with at least 500k inhabitants. As an indication, a village in the subsample closer to town is on average 7km far away from a 10k town, 22km from a 100k town and 65km from a 500k town. On the other hand, more isolated villages are on average 20.6km, 60km and 128km far away from the nearest 10k, 100k or 500k town, respectively.

In Table B.14 I perform the same tests for validity of the RD continuity assumption as I did on the aggregate sample. There does not seem to be any significant discontinuity at the 300 threshold before the program.

Fafchamps and Wahba (2006) show that children residing in and closer to urban centers in Nepal are more likely to attend school and work less in total. However, they are also more likely to be involved in wage work or in small family businesses. Urban proximity is generally associated with an improvement in children's welfare in terms of child labor and education. Table B.29 confirms that average enrolments are higher in the subsample of villages closer to a 10k town than in those further away.

Here I consider the causal effect of (eligibility to) electrification. The literature is very scarce on how electrification and geographic location interact with each other. My results strongly point to an important negative effect on school enrolments, conditional on the geographic location of the village with respect to the nearest urban center. The findings seem to confirm the spillover effects of urbanization indicated in Fafchamps and Shilpi (2005), which affect the labor market of surrounding villages up to 3 to 4 hours of travel time away. This would explain why the effect of electrification seems to fade away when considering bigger towns and larger distances. Moreover, my results seem to confirm the hypothesis that electrification changes the time allocation of older children more than for younger children. Stronger physical and cognitive capabilities make them more likely to be able to help with household chores, work in the child labor market or in a family business.

My results seem to indicate that in villages closer to a urban center, electrification increases the opportunity cost of schooling more than the returns to education. Because of data limitation, I am not able to prove that the decrease in enrolments is mirrored by an increase in child labor supply. Nevertheless, my findings indicate that there might be unintended effects of electrification on rural development. If households prefer to use the new opportunities created by electrification for short-term economic gains rather than invest in the long-term through the education of their children, policies targeted to rural economic development may fail in their long-term objectives.

The extent to which rural villages integrate in the market with surrounding towns or other villages, greatly depends on the level of local infrastructure. Roads and transportation means provide the necessary physical infrastructure needed for market integration. According to the Indian Population Census, 45% of the villages in the sample of single-habitation, 10th Plan districts was connected by all-weather roads in 2001. The share increased to 55% in 2011. In Table B.30 I explore whether eligibility for electrification has impacted enrolment levels in a different way, depending on whether the village was connected by tar road in 2001. Although results are negative and of higher magnitude in villages connected by road in 2001, none of the RD coefficients is statistically significant, with the exception of exam results. However, the evidence does not indicate a clear pattern between the two subsamples. Road connection alone does not seem to play an important role in shaping how electrification affects enrolment decisions.

7.4 Migration

As Dinkelman (2011) notes, a channel through which electrification may affect the labor market is migration. The same holds for school enrolments. I may be mistaking the effect of electrification on school enrolments for what is actually an effect on migration between villages or towards towns. For example, the decrease in enrolments in villages in proximity of towns may be caused by households out-migrating towards towns or other villages further away. In the attempt to analyse this possibility, I run eq. (1) on 2011 village population.

I first test whether the RDD assumption of continuity of potential outcome is satisfied. I use the 1991 Population Census contained in the SHRUG as pre-treatment outcome. Results of the validity test are presented in Panel D of Table B.13 and in Figure B.9. The assumption seems to be satisfied, as there is no discontinuity at the 300 threshold when considering pre-treatment population levels. I thus turn to the 2011 population levels. Results are shown in Table 7 and illustrated graphically in Figure 4. Population seems to significantly increase by around 35 inhabitants in villages eligible for the RGGVY program. Compared to the 2001 population of a village at the mean of the estimation sample, this represents a 11% increase. Figure B.15 shows that the coefficient looses significance with larger bandwidth sizes but remains positive.

I perform the same analysis taking into account the geographic location of the village. Table B.31 illustrates the numerical results. The same are presented graphically in Figure B.18. Irrespective of the size of town considered, electrification seems to induce an increase in population in the relatively more isolated villages. Compared to 2001, the effect on population ranges from 8% (10k towns) to 15% (100k towns). In villages relatively closer to the urban center, electrification does not seem to have a clear impact on population. While the effect is positive and significant in villages close to a 10k town, it becomes insignificant and even negative for 100k towns.

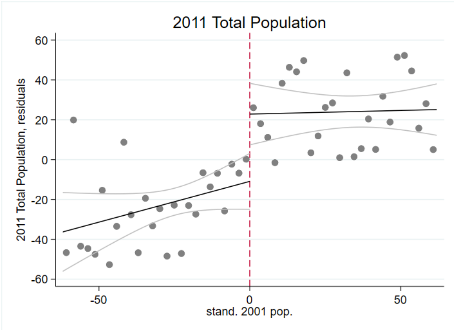
Studying migration flows induced by investment in rural infrastructure is out of the scope of this master thesis. However, this evidence allows me to exclude the possibility that the drop in enrolments in proximity of towns is driven by an out-migration flow. At the same time, these results are consistent with the idea that infrastructures make rural villages more attractive.

Table 7: RD - 2011 Total Population

Variables	2011 Total Population
1[2001 pop. >= 300]	37.135***
	(9.918)
stand. 2001 pop.	0.717***
	(0.248)
1[2001 pop. >= 300]*stand. 2001 pop.	-0.682**
	(0.346)
Observations	1,373
R-squared	0.317
1991 Control	Yes
District FE	Yes
Baseline covariates	Yes
Bandwidth	62
Mean of dependent variable	340.2

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This specification includes the 1991 population level as pre-treatment control, in order to avoid multicollinearity between the 2001 control level and the running variable.

Figure 4: RD graphical results - 2011 Population



These figures show the numerical results obtained in Table 7. The dots represent the conditional average residuals obtained from regressing the outcome variable on all covariates but the running variable. Each dot contains on average 30 observations, averaged in 5-person population bins. Lines are estimated separately on each side of the 300-person cutoff. The sample contains only unique-habitation villages located in districts receiving funds under the 10th Plan and with 2001 population within the optimal bandwidth chosen for estimation based on Calonico et al. (2014a).

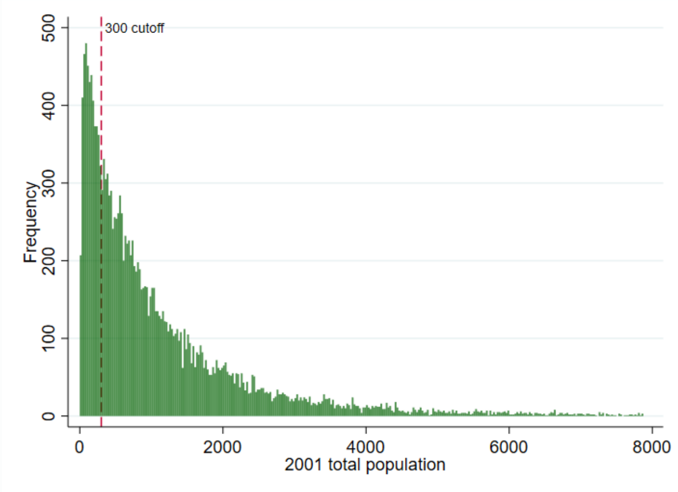
7.5 Enforcement of program requirements

As explained in Section 5.4, I web-scrape the list of villages in which electrification works were completed under the RGGVY program. I use these data to study whether the population cutoff contained in the program requirements was strictly enforced. After keeping only single-habitation villages located in districts which received funds only under the 10th Plan, I plot the distribution of treated villages by 2001 population size. Figure 5 shows how the list of treated villages as reported on the official program website contains many villages which had population size lower than 300 in 2001. 18% of the listed villages should not have been treated under the 10th Plan. This evidence sparks some doubts about the level of enforcement of the program guidelines. Furthermore, it raises some concerns about the potential role of corruption in funds allocation.

There is a strand of the literature on electrification which focuses on the role of politics in determining allocation of funds for large-scale infrastructure investments. Chhibber et al. (2004); Baskaran et al. (2015) report how electrification in India is a central topic of political campaigns. Besley et al. (2012) study how politics shapes the public provision of various infrastructures, while Ahlborg and Hammar (2014) and Borang et al. (2016) show how poor institutional quality in terms of corruption affects progresses in rural electrification in Mozambique, Tanzania and India. Wilkinson (2006) refers to the corruption allegations which regarded the Indian Golden Quadrilateral project managed by the National Highway Authority. State-level audits revealed that "investments were often directed to people and places that appear not to fit the declared criteria for the program". Surveys suggest that up to 56% of beneficiaries of small grants (including infrastructure grants) were supposed to be ineligible for the respective programs according to their guidelines.

Whether because of corruption or misreporting of the progresses of the program, the non-enforcement of the program guidelines would invalidate a RD strategy.

Figure 5: Distribution of treated villages



This histogram plots the distribution of the villages contained in the web-scraped list of treated units over 2001 population. The ratio of villages falling below the 300 cutoff is 18%.

8 Discussion

The results of this master thesis strongly question the effectiveness of the RGGVY electrification program in increasing households' access to electricity and improving the welfare of the rural population. While the Government of India claims that village electrification is complete, I stress the importance of relying on actual electricity consumption in order to assess the progresses made. In fact, to the extent that night lights constitute a good proxy for it, no significant increase is found in the medium-term. This ultimately remains a data limitation problem. There is no dataset available for energy consumption at the village level. While night lights have been proved to be a powerful instrument for researchers, concerns remain that they only represent a lower bound for the actual consumption of electricity in rural areas. Furthermore, this thesis only studies changes in nighttime brightness in the medium-term. It would be important to consider the long-term effects of the electrification program. In fact, households in rural areas may be too income constrained to pay for the connection to the grid and for the electricity consumed. If income levels benefit from the new economic opportunities coming with electricity, households may be more willing and able to pay in the long-term. For this purpose, night lights from the Visible Infrared Imaging Radiometer Suite (VIIRS) [22] could be used. While the DMS-OLS data I use are collected until 2013, the VIIRS was launched in 2011 and photographs the Earth still today. Related studies also document how the quality of power supply and affordability are crucial in the decision of households to take-up the service. Connecting the village to the grid may thus be not sufficient in order to introduce electricity in all households' everyday life. Policies such as subsidization of energy consumption for households, also above the poverty line, may play an important role in increasing electrification rates.

These results also point to important unintended effects of investments in rural electrification. Large amount of money has been invested in policies which identify electrification as a crucial instrument for local development. A decrease in enrolment levels risks undermining the achievement of the long-term objectives set by

[22] https://maps.ngdc.noaa.gov/viewers/VIIRS_DNB_nighttime_imagery/index.html

these policies. Capital constrained households may prefer to capitalize on the new economic opportunities brought by electrification instead of investing in the long-term through the education of their children. Policymakers should consider the integration of electrification programs with other policies which aim at limiting the unintended effects as much as possible. For example, conditional cash transfer programs such as Progresa, have been proved to be effective in improving schooling in the long term (Parker and Vogl, 2018).

Moreover, this thesis briefly touches on the issue of weak enforcement of program guidelines. If funds allocation has been manipulated under the RGGVY program, the validity of the RD strategy is severely threatened. Weak enforcement and corruption have been shown to hinder the achievement of program's goals, thus conditioning the social benefits which accrue to the most disadvantaged population originally targeted by the program.

This thesis has also other limitations. The fuzzy match undertaken to link datasets which have no village identifiers is by nature imperfect. Although I try to exclude the imprecise links, false matches may still be present. Furthermore, I interpret the negative effect on school enrolments as a consequence of new working opportunities created by electrification, including an increase in child labor. Positive employment and income effects have been found in the literature on rural electrification (Khandker et al., 2014, 2013; Rud, 2012). Child labor of older children seems to be positively correlated to working hours of the adults in the household (Islam and Sivasankaran, 2014). Nevertheless, this is a speculation I am not able to prove empirically. Data from the Indian economic census could be linked to the SHRUG in order to understand whether changes in the labor market can explain those in enrolments. However, data on child labor are likely only obtained through households' surveys. I use enrolment levels as outcome variables. Enrolment rates would be a preferable measure because they allow to take into account the school-age population at the village level.

From this thesis, ideas for further research can be drawn. The effect which electrification has on enrolments showcases a strong heterogeneity by geographical location of the village relative to the nearest urban centers. This calls for an important role of market integration in the way electrification impacts local economies. It would be interesting to explore the heterogeneity in more detail through quantile regressions and by discerning whether the spillovers on surrounding villages are driven by the size of the town or by its physical distance. Moreover, market integration is conditional on the quality of physical infrastructures such as roads. The interaction between geographical location and road connection of the village might be important in shaping the structural changes caused by electrification. Ideally, detailed data on migration flows would be extremely useful in studying in which way electrification changes the composition of the village population. It is important to understand who migrates as well as the origin and destination places in order to take this into account in the policy-making process. As Dinkelman and Schulhofer-Wohl (2015) point out, congestion of public services due to migration flows induced by better infrastructure could hinder the realization of the full potential benefits. Lastly, heterogeneity by social background and spillover effects between neighboring villages would also constitute a relevant evidence.

9 Conclusion

This thesis uses Indian population census, night lights and school enrolments to study the impact of the RGGVY rural electrification program on education at the village level. It finds that eligibility for the program did not substantially increase consumption of electricity, as proxied by satellite images. At the aggregate level, the intention-to-treat effect of electrification on school enrolments does not seem to be significantly different from zero. However, it seems to negatively affect the intensive margin of education. When distance to the nearest town is taken into account, enrolments decrease with eligibility in villages relatively closer to a population center. On the other hand, no effect is found in more isolated villages. In economic terms, electrification seems to generate an opportunity cost effect which outweighs returns to education when the local economy is geographically closer to a urban market. Based on the existing literature, I speculate that the negative effect on education may be attributable to changes in the labor market caused by electrification. It would thus represent a non-intended effect of policies aimed at electrifying rural areas. Further research is needed in order to better understand what the real drivers are.

Bibliography

A. Adukia, S. Asher, and P. Novosad. Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction.

H. Ahlborg and L. Hammar. Drivers and barriers to rural electrification in Tanzania and Mozambique – Grid-extension, off-grid, and renewable energy technologies. *Renewable Energy*, 61, Jan. 2014.

S. Asher and P. Novosad. Evidence from Rural Roads in India. 2016.

S. Asher and P. Novosad. Rural Roads and Local Economic Development. *American Economic Review (forthcoming)*, 2019a.

S. Asher and P. Novosad. Rural Roads and Local Economic Development. 2019b.

S. Asher, T. Lunt, R. Matsuura, and P. Novosad. The Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG). Working paper, 2019a.

S. Asher, T. Lunt, R. Matsuura, and P. Novosad. The Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG). 2019b.

S. G. Banerjee, D. Barnes, B. Singh, K. Mayer, and H. Samad. *Power for All: Electricity Access Challenge in India*. The World Bank, Nov. 2014.

M. Barron and M. Torero. Household electrification and indoor air pollution. *Journal of Environmental Economics and Management*, Nov. 2017.

T. Baskaran, B. Min, and Y. Uppal. Election cycles and electricity provision: Evidence from a quasi-experiment with Indian special elections. *Journal of Public Economics*, 126, June 2015.

G. Bensch, J. Kluve, and J. Peters. Impacts of rural electrification in Rwanda. Dec. 2011.

T. Bernard and M. Torero. Social Interaction Effects and Connection to Electricity: Experimental Evidence from Rural Ethiopia. *Economic Development and Cultural Change*, 63, Apr. 2015.

T. Besley, R. Pande, and V. Rao. Just Rewards? Local Politics and Public Resource Allocation in South India. *The World Bank Economic Review*, Jan. 2012. Publisher: Oxford Academic.

R. C. M. Beyer, E. Chhabra, V. Galdo, and M. Rama. *Measuring Districts' Monthly Economic Activity from Outer Space*. Policy Research Working Papers. The World Bank, July 2018.

R. Bilolikar. Rural Electrification in India – an overview. 2004.

M. Blasnik. RECLINK: Stata module to probabilistically match records, Jan. 2010. URL https://ideas.repec.org/c/boc/bocode/s456876.html.

J. Bonan, S. Pareglio, and M. Tavoni. Access to modern energy: a review of barriers, drivers and impacts. *Environment and Development Economics*, Oct. 2017.

F. Borang, S. Jagers, and M. Povitkina. How corruption shapes the relationship between democracy and electrification. 2016.

British Council. The school education system in india - an overview, 2019.

T. Bundervoet. *Bright Lights, Big Cities: Measuring National and Subnational Economic Growth in Africa from Outer Space, with an Application to Kenya and Rwanda*. Policy Research Working Papers. The World Bank, Oct. 2015.

F. Burlig and L. Preonas. Out of the Darkness and Into the Light? Development Effects of Rural Electrification. 2006.

S. Calonico, M. D. Cattaneo, and R. Titiunik. Robust Data-Driven Inference in the Regression-Discontinuity Design. *The Stata Journal: Promoting communications on statistics and Stata*, Dec. 2014a.

S. Calonico, M. D. Cattaneo, and R. Titiunik. Robust Data-Driven Inference in the Regression-Discontinuity Design. *The Stata Journal: Promoting communications on statistics and Stata*, Dec. 2014b.

S. Calonico, M. D. Cattaneo, M. H. Farrell, and R. Titiunik. Regression Discontinuity Designs Using Covariates. *The Review of Economics and Statistics*, July 2019.

Central Electricity Authority. Load generation balance report, 2010-2011. 2011. http://cea.nic.in/annualreports.html.

U. Chakravorty, K. Emerick, and M.-L. Ravago. Lighting Up the Last Mile: The Benefits and Costs of Extending Electricity to the Rural Poor. *SSRN Electronic Journal*, 2016.

P. Chhibber, S. Shastri, and R. Sisson. Federal arrangements and the provision of public goods in india. *Asian Survey*, June 2004.

U. Deichmann, F. Shilpi, and R. Vakis. Urban Proximity, Agricultural Potential and Rural Non-farm Employment: Evidence from Bangladesh. *World Development*, Mar. 2009.

C. Dimiceli, M. Carroll, R. Sohlberg, D. Kim, M. Kelly, and J. Townshend. Mod44b modis/terra vegetation continuous fields yearly l3 global 250 m sin grid v006 [data set]. *NASA EOSDIS Land Process*, 2015.

T. Dinkelman. The Effects of Rural Electrification on Employment: New Evidence from South Africa. *American Economic Review*, Dec. 2011.

T. Dinkelman and S. Schulhofer-Wohl. Migration, congestion externalities, and the evaluation of spatial investments. *Journal of Development Economics*, May 2015. doi: 10.1016/j.jdeveco.2014.12.009.

C. Elvidge, F.-C. Hsu, K. Baugh, and T. Ghosh. National Trends in Satellite-Observed Lighting: 1992–2012. In Q. Weng, editor, *Global Urban Monitoring and Assessment through Earth Observation*. CRC Press, May 2014. doi: 10.1201/b17012-9.

M. Fafchamps and F. Shilpi. Cities and Specialisation: Evidence from South Asia. *The Economic Journal*, 115, Apr. 2005.

M. Fafchamps and J. Wahba. Child labor, urban proximity, and household composition. *Journal of Development Economics*, Apr. 2006.

A. Foster and M. Rosenzweig. Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970–2000. *Economic Development and Cultural Change*, Apr. 2004.

M. Frölich and S. Sperlich. *Impact evaluation.* Cambridge University Press, 2019.

A. Gelman and G. Imbens. Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*, 37, July 2019.

N. Gennaioli, R. La Porta, F. Lopez De Silanes, and A. Shleifer. Growth in regions. *Journal of Economic Growth*, 19, Sept. 2014.

J. Hahn, P. Todd, and W. Van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, 2001.

J. V. Henderson, A. Storeygard, and D. N. Weil. A Bright Idea for Mesuring Economic Growth. *American Economic Review*, 2011a.

J. V. Henderson, A. Storeygard, and D. N. Weil. Measuring Economic Growth from Outer Space. *American Economic Review*, Apr. 2012.

V. Henderson, A. Storeygard, and D. N. Weil. A Bright Idea for Measuring Economic Growth. *American Economic Review*, 101, May 2011b.

G. W. Imbens and T. Lemieux. Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2):615–635, 2008.

M. Islam and A. Sivasankaran. How does Child Labor respond to changes in Adult Work Opportunities? Evidence from NREGA. 2014.

F. R. Jensenius and G. Verniers. Studying indian politics with large-scale data: Indian election data 1961–today. *Studies in Indian Politics*, 2017.

S. R. Khandker, D. F. Barnes, and H. A. Samad. Welfare Impacts of Rural Electrification: A Panel Data Analysis from Vietnam. *Economic Development and Cultural Change*, Apr. 2013.

S. R. Khandker, H. A. Samad, R. Ali, and D. F. Barnes. Who Benefits Most from Rural Electrification? Evidence in India. *The Energy Journal*, Apr. 2014.

G. Kohlin, E. O. Sills, S. K. Pattanayak, and C. Wilfong. *Energy, Gender and Development: What are the Linkages? Where is the Evidence?* Policy Research Working Papers. The World Bank, Sept. 2011.

K. Lee, E. Miguel, and C. Wolfram. Experimental Evidence on the Economics of Rural Electrification. *Journal of Political Economy*, Apr. 2020.

M. Lipscomb, A. M. Mobarak, and T. Barham. Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil. *American Economic Journal: Applied Economics*, Jan. 2013.

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

B. Min. Electrifying the poor: distributing power in India. *Ann Arbor*, 2011.

B. Min and K. M. Gaba. Tracking Electrification in Vietnam Using Nighttime Lights. *Remote Sensing*, Oct. 2014.

B. Min, K. M. Gaba, O. F. Sarr, and A. Agalassou. Detection of rural electrification in Africa using DMSP-OLS night lights imagery. *International Journal of Remote Sensing*, Nov. 2013.

Ministry of Power. Rggvy scheme of rural electricity infrastructure and household electrification - memorandum, 2005.

Ministry of Power. Guidelines for preparation of DPRs under XII Plan of RGGVY, 2011.

M. Mukherjee. Do Better Roads Increase School Enrollment? Evidence from a Unique Road Policy in India. *SSRN Electronic Journal*, 2012.

S. Nhalur, A. Josey, and M. Mandal. Rural Electri[U+FB01]cation in India. (45), 2018.

S. Parker and T. Vogl. Do Conditional Cash Transfers Improve Economic Outcomes in the Next Generation? Evidence from Mexico. Technical report, National Bureau of Economic Research, Cambridge, MA, Feb. 2018.

Planning Commission, GoI. Mid-Term Appraisal - Eleventh Five Year Plan 2007-2012, 2011.

Planning Commission of India. Minimum Needs Programme, 1973.

Pokhrel Amod K., Bates Michael N., Verma Sharat C., Joshi Hari S., Sreeramareddy Chandrashekhar T., and Smith Kirk R. Tuberculosis and Indoor Biomass and Kerosene Use in Nepal: A Case–Control Study. *Environmental Health Perspectives*, Apr. 2010.

A. Prakash, A. K. Shukla, C. Bhowmick, and R. C. M. Beyer. Night-time Luminosity: Does it Brighten Understanding of Economic Activity in India? 2019a.

N. Prakash, M. Rockmore, and Y. Uppal. Do criminally accused politicians affect economic outcomes? evidence from india. *Journal of Development Economics*, 2019b.

Prayas (Energy Group). Electricity supply monitoring initiative (esmi) - summary analysis, september 2019. 2019. https://www.watchyourpower.org/uploaded_reports.php.

N. D. Rao. Does (better) electricity supply increase household enterprise income in India? *Energy Policy*, June 2013.

J. P. Rud. Electricity provision and industrial development: Evidence from India. *Journal of Development Economics*, Mar. 2012.

United Nations. *Sustainable Development Goals Report*. 2016. URL https://www.un-ilibrary.org/content/publication/3405d09f-en.

D. van de Walle, M. Ravallion, V. Mendiratta, and G. Koolwal. *Long-Term Impacts of Household Electrification in Rural India*. Policy Research Working Papers. The World Bank, June 2013.

M. Ward. Rural education in india. *India Infrastructure Report*, 2007.

S. I. Wilkinson. *The Politics of Infrastructure spending in India*. 2006.

World Bank. *The Welfare Impact of Rural Electrification: A Reassessment of the Costs and Benefits*. The World Bank, Mar. 2008.

World Bank. Total population between the ages 0 to 14. - india, 2020a. Data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SP.POP.0014.TO?locations=IN.

World Bank. Access to electricity, rural (% of rural population), 2020b. Data retrieved from World Development Indicators, https://data.worldbank.org/indicator/EG.ELC.ACCS.RU.ZS?locations=IN.

World Bank. School enrollment, primary (% gross) - india, 2020c. Data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SE.PRM.ENRR?locations=IN.

World Bank. Poverty headcount ratio at $1.90 a day (2011 ppp) (% of population) - india, 2020d. Data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SI.POV.DDAY?locations=IN.

A Data Preparation

A.1 Fuzzy merge

The NRDWP habitations census and the DISE datasets have no numerical village identifier. Therefore, I must rely on state, district and village names in order to merge these data to the SHRUG. This process is relatively time-intensive and by nature imperfect. I try to limit the prevalence of imperfections as much as possible throughout the procedure. This implies a greater loss of observations. Because of the many languages spoken in India and their ambiguous translation to English, district and village names are very often spelled differently across years and data sources. I start by cleaning district and village names, in order to make them as comparable as possible among the SHRUG, the NRDWP and the DISE datasets. In order to perform a fuzzy merge on the names, I make use of the `reclink2` program in Stata (Blasnik, 2010). `reclink2` is an extension of `reclink`, which uses record linkage methods to match observations between two datasets where no perfect key fields exist. It allows to specify a list of variables on which the fuzzy merge is to be performed, their associated match and nonmatch weights and the minimum matching score for which two observations are considered a match. `reclink2` introduces the possibility to perform a many-to-one linking procedure, in which multiple observations of the master dataset can be matched to the same observations from the using dataset. I use this option in order to increase the chances of finding correct matches. I base the fuzzy merge on state, district and village names. Unfortunately, the combination of these three specifications still does not uniquely identify a village. There are two levels of geographic specification (subdistricts and blocks) which I can not use in the matching procedure. While subdistrict names are included in the SHRUG but not in the NRDWP nor in the DISE, the opposite is true for block names. This implies that even if two observations result in a perfect match, they may still refer to different villages. In fact, a district may contain villages which are called the same (or have very similar names) but located in different subdistricts or blocks. While these two are non-comparable administrative divisions, I used them to clean the resulting matches from duplicates generated by the many-to-one option.

In the fuzzy merge process, I take the following steps. I separately link the NRDWP and the DISE datasets to the SHRUG. I set the minimum matching score to 0.85 (on a scale from 0 to 1, default is 0.60) and perform a many-to-one merge. For each village contained in the SHRUG (identified by a unique shrid), I keep the match with the highest matching ratio. This still results in duplicate shrids, since multiple observations can match to the same shrid with the same (highest) matching ratio. At this point of the process, before eliminating the remaining duplicates, I construct a variable for the quality of the match. The number of different villages with which each shrid matches is an indicator for the uncertainty of the link. The higher the number of possible links, the higher the probability that the final match will be wrong. On the other hand, shrids which match to one unique village name from the NRDWP or DISE dataset are likely to represent a good quality match. From the final dataset, best quality matches are the only one which are used in the analysis. I take three other steps in order to clear the resulting dataset from duplicates. Among the duplicates of each shrid, I identify whether there is one for which subdistrict and block names are the same. While these divisions are not comparable, their pairing in names is likely to be an indicator for a correct match among villages. I then use the village population levels reported by the 2009 NRDWP habitation census to compare them with those of the 2011 population census. Among the duplicates of each shrid, I keep the one with the lowest population disparity. This step is omitted when matching the SHRUG to the DISE dataset, where no population values are included. Lastly, the remaining duplicates are randomly dropped. For the DISE dataset, I take a further step before matching. Because of misspelling of village names or due to merge and split of units, the same school identification code may be located into different villages across years. In order to correct for this, I create a list of all the schools present in the dataset in the years 2005, 2011 and 2017 and I assign them the village names as reported in 2005. This is also a way to take a 'snapshot' of where the school was located at the time of the RGGVY funds assignment. Those schools which were open after 2005, which find no merge with the 2005 names, are assigned the 2011 or 2017 village names, depending on when the school code is first found in the data. From the NRDWP, 87% of the villages finds a match in the SHRUG. Of these, 98% belong to the best quality match category. From the DISE, 67% of the villages which have a school in either 2005, 2011 or 2017 are matched to the SHRUG. Of these, 93% are best quality matches. The large difference in performance of the fuzzy merge can be explained by the relatively large share of villages which do not have a primary school within their boundaries. According to the population census, these represented the 21% in 2001 and the 17% in 2011. These villages will not be listed in the DISE dataset. Once the NRDWP and the DISE datasets contain the shrids, I merge all the data sources together. Table A.8 illustrates step by step the number of observations which successfully merge among datasets. The table also breaks down the observations by subsamples. For the sake of illustration of the results, an average bandwidth of 150 is used.

Table A.8: Count of Villages by merged dataset

Merge	Total	10th Plan	10th Plan single-hab.	10th Plan single-hab. 150-450
2001 Census Codes	628.706			
+ 2001-2011 SHRUG Census	578.601	58.807		
+ NRDWP habitations	500.261	49.196	27.268	6.600
+ SHRUG night lights	459.735	45.205	24.774	5.761
+ DISE schools	308.748	32.509	17.190	3.629

Note: This table reports the number of observations which successfully merge between datasets (including all match quality types). The DISE dataset includes only Indian schools existing in 2005, 2011 and 2017. 37% of these constitute a balanced dataset. In the interest of presentation clarity, an average bandwidth of 150 around the 300 population cutoff is taken.

A.2 Night lights projection

Following Burlig and Preonas (2006), I use the projection of maximum night lights as preferred outcome variable for assessing changes in electrification. With the aim of reducing the random noise in yearly night lights, I project each year on two years before and after. The obtained values represent a measure of maximum night lights which is more consistent to that detected in the same village two years before and after. In practice, I estimate the following regression:

$$NL_v^t = \alpha_0 + \alpha_1 NL_v^{t-1} + \alpha_2 NL_v^{t-2} + \alpha_3 NL_v^{t+1} + \alpha_4 NL_v^{t+2} + \epsilon_v \tag{2}$$

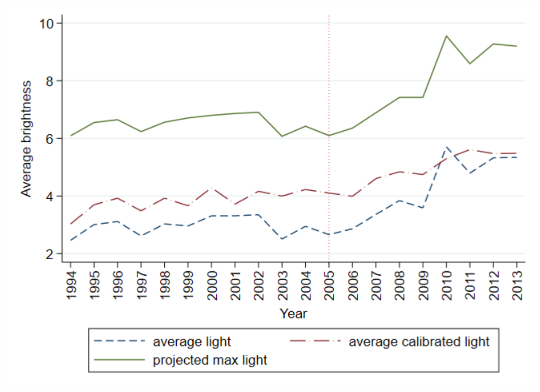
where NL_v^t is the value of maximum night lights detected in village v in year t. Using the estimated $\hat{\alpha_i}$ for $i = \{0, 1, 2, 3, 4\}$, I calculate $\hat{NL}_v^t$ which I use in eq. (1).

The transformation slightly increases the average brightness and its standard deviation. In Table B.17 I use non-projected night lights as robustness check. Compared to projected max night lights, the RD coefficient is less negative but still not significant.

B Other Tables and Figures

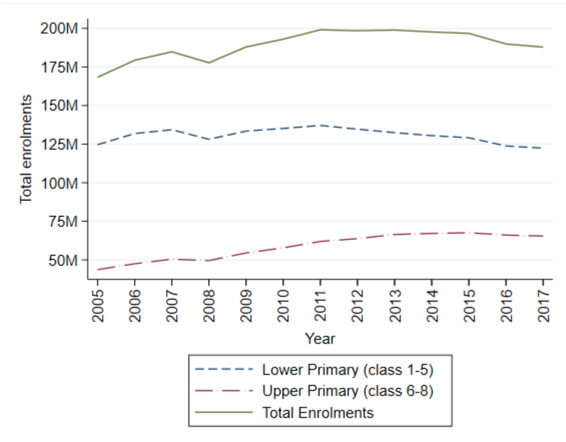
B.1 Summary statistics

Figure B.6: Average night lights over time



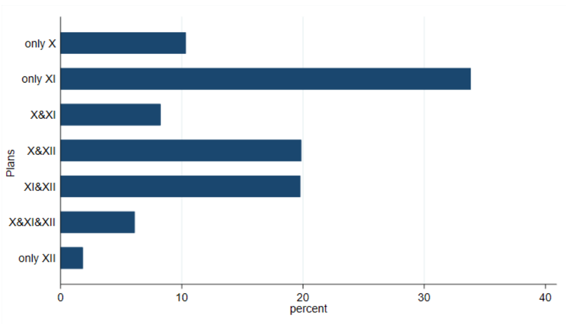
Note: The sample is the full one obtained after merging all datasets.

Figure B.7: Enrolments by level over time



Note: The sample is the full one obtained after merging all datasets.

Figure B.8: Villages in funds-receiving districts, by Plan



Note: This histogram plots the percentage of villages by Plans received by the district they are located in. The sample is the full one obtained after merging all datasets.

Table B.9: Summary Statistics - Night Lights by State over time

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Jammu Kashmir	2.5	3.1	3.2	2.9	3.0	3.3	3.3	3.4	3.6	2.9	3.4	3.2	3.7	4.0	4.9	4.8	7.4	6.4	7.0	7.7
	(4.7)	(5.3)	(5.5)	(5.1)	(5.1)	(5.6)	(5.4)	(5.6)	(5.6)	(4.9)	(5.2)	(4.8)	(5.2)	(5.6)	(6.4)	(6.7)	(8.6)	(8.1)	(8.6)	(9.0)
Himachal Pradesh	1.6	2.0	2.1	2.1	2.8	3.1	3.4	3.7	3.9	3.1	3.7	3.7	4.2	4.8	5.5	5.3	7.5	6.5	7.2	7.3
	(3.4)	(3.6)	(3.7)	(3.6)	(3.9)	(4.2)	(4.2)	(4.3)	(4.3)	(3.8)	(4.1)	(4.0)	(4.4)	(4.9)	(5.4)	(5.6)	(6.9)	(6.5)	(6.9)	(6.9)
Punjab	8.9	10.2	10.7	9.9	11.0	11.4	11.6	11.9	11.9	9.9	10.8	10.3	11.2	12.5	14.2	13.7	18.3	16.1	16.3	17.1
	(5.7)	(6.1)	(6.5)	(6.1)	(6.5)	(6.9)	(6.7)	(7.1)	(7.1)	(6.1)	(6.5)	(6.1)	(6.6)	(7.0)	(7.6)	(7.5)	(9.2)	(8.6)	(9.3)	(9.2)
Uttarakhand	1.8	2.0	2.1	1.9	2.3	2.4	2.7	2.9	3.3	2.6	3.0	3.1	3.2	4.1	4.5	4.9	7.1	6.1	6.9	7.1
	(4.7)	(4.9)	(5.1)	(4.8)	(5.2)	(5.5)	(5.5)	(5.9)	(6.0)	(5.3)	(5.9)	(5.7)	(6.3)	(6.9)	(7.8)	(7.8)	(10.0)	(9.3)	(9.8)	(9.8)
Haryana	9.3	10.8	10.9	10.0	10.5	11.2	11.2	11.7	11.9	10.1	10.9	10.1	10.8	12.2	14.1	14.4	20.8	18.9	19.9	20.2
	(6.4)	(7.1)	(7.4)	(6.9)	(7.3)	(7.9)	(7.7)	(8.1)	(8.3)	(7.1)	(7.6)	(7.1)	(7.6)	(8.2)	(9.1)	(9.1)	(11.7)	(11.0)	(11.8)	(11.4)
Rajastan	3.4	3.9	4.0	3.5	4.0	4.0	4.1	4.2	4.4	3.4	3.9	3.6	3.9	4.6	5.4	5.5	8.5	7.5	8.1	8.3
	(3.9)	(4.3)	(4.5)	(4.3)	(4.4)	(4.6)	(4.5)	(4.6)	(4.6)	(3.9)	(4.1)	(4.0)	(4.3)	(4.8)	(5.4)	(5.8)	(7.7)	(7.4)	(8.0)	(8.1)
Uttar Pradesh	9.6	10.1	9.8	9.4	9.2	9.6	9.5	9.4	9.5	8.8	8.9	8.7	8.9	9.4	9.9	9.8	11.3	10.5	11.1	10.9
	(18.8)	(18.7)	(18.8)	(18.8)	(18.9)	(18.8)	(18.8)	(18.9)	(18.9)	(19.0)	(19.0)	(19.0)	(19.0)	(19.0)	(18.9)	(19.0)	(19.0)	(19.2)	(19.3)	(19.4)
Bihar	2.8	3.0	3.2	2.9	3.1	3.2	3.2	3.3	3.3	2.9	3.2	3.0	3.0	3.5	3.6	3.6	4.7	4.3	5.1	4.2
	(10.9)	(10.9)	(10.9)	(10.9)	(10.9)	(10.9)	(10.9)	(10.9)	(10.9)	(10.8)	(10.8)	(10.8)	(10.9)	(10.9)	(11.0)	(11.0)	(11.3)	(11.3)	(11.4)	(11.5)
Arunachal Pradesh	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0
	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.5)	(1.6)	(1.6)	(1.6)
Manipur	5.9	6.4	6.8	6.2	6.8	7.0	6.9	7.0	6.8	5.7	6.0	5.3	5.3	5.9	6.0	6.2	7.9	7.2	8.1	7.4
	(11.7)	(11.7)	(11.8)	(11.7)	(11.7)	(11.6)	(11.6)	(11.7)	(11.6)	(11.6)	(11.6)	(11.7)	(11.6)	(11.6)	(11.7)	(11.6)	(11.6)	(11.7)	(11.6)	(11.7)
Mizoram	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7	62.7
	(4.1)	(4.1)	(4.1)	(4.1)	(4.1)	(4.1)	(4.1)	(4.1)	(4.1)	(4.1)	(4.1)	(4.1)	(4.1)	(4.1)	(4.1)	(4.1)	(4.0)	(4.1)	(4.1)	(4.1)
Meghalaya	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Assam	5.2	5.7	5.8	5.5	5.8	5.7	5.7	5.7	5.7	5.1	5.4	5.2	5.4	5.9	6.2	6.3	7.9	7.0	7.4	7.6
	(14.3)	(14.3)	(14.4)	(14.3)	(14.3)	(14.4)	(14.3)	(14.4)	(14.4)	(14.2)	(14.2)	(14.2)	(14.2)	(14.1)	(14.2)	(14.2)	(14.3)	(14.4)	(14.5)	(14.3)
West Bengal	3.8	4.2	4.4	3.9	4.4	4.6	4.8	4.9	5.0	4.1	4.6	4.2	4.5	5.2	5.8	5.8	8.6	7.6	8.7	8.5
	(4.7)	(5.0)	(5.2)	(4.6)	(4.9)	(5.0)	(4.9)	(5.2)	(5.0)	(4.3)	(4.5)	(4.0)	(4.4)	(4.7)	(5.5)	(5.8)	(7.4)	(7.2)	(7.8)	(7.4)
Jharkhand	1.4	1.7	1.7	1.4	1.6	1.5	1.7	1.8	1.8	1.5	1.7	1.6	1.7	2.1	2.3	2.2	3.7	2.8	3.2	3.1
	(6.1)	(6.2)	(6.2)	(5.9)	(5.9)	(5.9)	(5.9)	(6.0)	(6.0)	(5.8)	(5.9)	(5.7)	(5.9)	(6.1)	(6.4)	(6.7)	(7.7)	(7.5)	(8.0)	(7.8)
Odisha	4.5	4.7	4.9	4.7	4.9	5.1	5.0	5.1	5.2	4.7	4.9	4.7	4.9	5.2	5.5	5.3	7.0	6.1	6.6	6.5
	(13.7)	(13.7)	(13.6)	(13.7)	(13.6)	(13.7)	(13.6)	(13.7)	(13.7)	(13.6)	(13.6)	(13.6)	(13.7)	(13.7)	(13.7)	(13.9)	(13.9)	(14.0)	(14.2)	(14.1)
Chhattisgarh	2.4	2.5	2.7	2.3	2.3	2.4	2.5	2.6	2.8	2.4	2.9	2.7	3.1	3.6	4.1	4.4	6.8	6.0	7.0	6.7
	(5.5)	(5.7)	(5.7)	(5.5)	(5.6)	(5.7)	(5.6)	(5.8)	(5.9)	(5.6)	(5.8)	(5.8)	(6.0)	(6.3)	(6.8)	(7.1)	(8.3)	(8.2)	(9.1)	(8.7)
Madhya Pradesh	3.3	3.7	3.7	3.3	3.4	3.3	3.3	3.3	3.0	2.2	2.4	2.0	2.2	2.4	2.6	2.5	3.9	3.1	4.0	3.6
	(3.6)	(3.9)	(4.0)	(3.8)	(3.9)	(4.0)	(3.8)	(3.9)	(3.8)	(3.2)	(3.2)	(3.1)	(3.3)	(3.6)	(3.9)	(4.2)	(5.3)	(5.2)	(6.0)	(6.0)
Gujarat	5.7	6.3	6.7	6.0	6.9	7.0	7.1	7.3	6.9	5.5	5.9	5.1	5.4	6.0	6.8	6.7	9.4	8.2	9.0	8.9
	(6.5)	(7.0)	(7.3)	(7.0)	(7.3)	(7.5)	(7.2)	(7.3)	(7.0)	(6.3)	(6.4)	(6.0)	(6.3)	(6.6)	(7.1)	(7.3)	(8.7)	(8.5)	(9.1)	(9.1)
Maharashtra	4.9	5.6	5.8	5.3	6.0	6.2	6.3	6.4	6.3	4.9	5.2	4.5	4.7	5.2	5.9	6.0	8.7	7.5	8.4	8.6
	(5.6)	(6.0)	(6.1)	(6.0)	(6.2)	(6.4)	(6.3)	(6.4)	(6.4)	(5.8)	(5.9)	(5.6)	(5.7)	(5.8)	(6.2)	(6.6)	(7.9)	(7.7)	(8.4)	(8.1)
Andhra Pradesh	4.8	5.4	5.6	5.1	5.6	5.7	5.9	6.0	6.0	4.9	5.4	5.0	5.3	6.2	7.0	7.2	10.2	9.0	9.6	9.8
	(3.9)	(4.1)	(4.2)	(4.0)	(4.2)	(4.4)	(4.4)	(4.5)	(4.6)	(4.1)	(4.3)	(4.1)	(4.4)	(4.7)	(5.1)	(5.5)	(6.6)	(6.4)	(7.1)	(6.9)
Karnataka	3.8	4.4	4.7	4.1	4.8	5.1	5.3	5.5	5.6	4.5	5.1	4.6	5.1	5.8	6.5	6.6	9.4	8.1	8.8	9.0
	(2.9)	(3.4)	(3.4)	(3.3)	(3.5)	(3.8)	(3.7)	(3.9)	(3.9)	(3.4)	(3.7)	(3.5)	(3.8)	(4.2)	(4.7)	(5.1)	(6.3)	(6.0)	(6.9)	(6.7)
Kerala	8.9	10.1	10.0	8.8	9.5	9.9	10.0	10.4	10.3	8.5	9.3	8.6	9.2	10.2	11.4	11.7	15.2	13.8	14.8	14.7
	(4.6)	(5.1)	(5.0)	(4.4)	(4.8)	(4.9)	(4.8)	(5.1)	(4.9)	(4.3)	(4.5)	(4.2)	(4.7)	(5.0)	(5.5)	(6.0)	(7.0)	(6.7)	(7.4)	(7.2)
Tamil Nadu	6.9	7.5	7.9	7.3	8.3	8.8	9.2	9.5	9.7	7.9	8.6	7.8	8.4	9.4	10.8	11.2	14.5	12.8	13.5	13.0
	(5.1)	(5.5)	(5.7)	(5.5)	(5.9)	(6.5)	(6.6)	(6.8)	(6.9)	(6.0)	(6.3)	(5.9)	(6.3)	(6.8)	(7.6)	(7.9)	(9.2)	(8.7)	(9.3)	(8.9)
Total	5.9	6.4	6.5	6.1	6.4	6.5	6.6	6.7	6.7	5.8	6.2	5.8	6.1	6.6	7.2	7.2	9.4	8.4	9.1	9.0
	(12.7)	(12.8)	(12.8)	(12.8)	(12.8)	(12.9)	(12.8)	(12.9)	(12.9)	(12.7)	(12.8)	(12.7)	(12.8)	(12.9)	(13.0)	(13.1)	(13.6)	(13.6)	(13.8)	(13.8)

Note: This table shows projected maximum night light values, obtained as explained in Appendix A.2, over the whole time period the NOAA satellite images are available. Each village is assigned a DN (Digital Number) which represents the highest brightness value detected in the year. This corresponds to the value of the brightest pixel within the village boundaries. This table reports state 2001 averages. DN are on a scale between 0 and 63. The sample is the full one obtained after merging all datasets.

Table B.10: Summary Statistics - DISE Schools Dataset

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
A. Students enrolled per village													
Lower Primary, Boys	118.9	123.2	123.1	122.3	119.9	120.5	121.7	118.8	117.0	115.1	113.8	109.4	108.3
	(299.5)	(286.3)	(293.0)	(283.0)	(282.9)	(283.7)	(306.9)	(307.0)	(305.8)	(304.3)	(300.3)	(296.3)	(303.2)
Lower Primary, Girls	108.8	114.1	114.5	114.5	112.7	113.1	114.0	111.3	108.9	107.1	105.9	101.5	100.2
	(276.8)	(260.9)	(266.9)	(257.1)	(257.2)	(254.7)	(276.4)	(276.6)	(273.0)	(272.2)	(268.1)	(262.1)	(267.8)
Upper Primary, Boys	43.3	45.7	47.3	48.1	49.3	51.6	54.7	55.8	58.2	58.8	59.1	57.9	57.5
	(163.7)	(159.8)	(168.9)	(163.1)	(165.3)	(166.4)	(177.1)	(187.7)	(194.5)	(192.3)	(191.3)	(194.1)	(187.5)
Upper Primary, Girls	36.5	39.8	41.9	43.5	45.7	48.4	51.8	53.2	55.2	55.6	55.9	54.6	54.0
	(149.8)	(147.4)	(154.2)	(149.8)	(152.6)	(153.4)	(163.3)	(171.0)	(176.2)	(174.2)	(174.8)	(176.4)	(170.6)
Total, Boys	162.2	168.9	170.4	170.3	169.2	172.1	176.5	174.6	175.2	173.9	172.8	167.3	165.8
	(441.0)	(415.3)	(432.0)	(419.9)	(423.6)	(427.1)	(461.4)	(467.2)	(473.6)	(471.2)	(469.9)	(469.3)	(470.7)
Total, Girls	145.4	153.9	156.4	158.0	158.4	161.5	165.8	164.4	164.1	162.7	161.8	156.1	154.2
	(405.8)	(377.2)	(391.1)	(380.0)	(384.2)	(384.0)	(415.7)	(419.6)	(422.3)	(420.1)	(420.1)	(416.3)	(417.6)
Total	307.5	322.8	326.8	328.4	327.6	333.6	342.2	339.0	339.3	336.7	334.7	323.4	320.0
	(841.1)	(786.8)	(817.3)	(794.1)	(801.7)	(804.8)	(871.1)	(880.8)	(889.9)	(885.2)	(883.9)	(879.8)	(882.3)
B. Observations per year													
Schools per village	2.1	2.2	2.2	2.3	2.3	2.4	2.4	2.4	2.6	2.6	2.6	2.6	2.7
	(3.7)	(3.2)	(3.3)	(3.3)	(3.2)	(3.3)	(3.7)	(3.5)	(3.9)	(3.8)	(3.8)	(3.8)	(4.4)
Number of villages	547.238	555.601	565.476	541.188	573.656	578.442	581.678	585.355	586.212	587.147	587.792	587.115	586.978
Number of schools	1.124.033	1.194.912	1.258.055	1.228.675	1.303.856	1.361.630	1.412.178	1.431.703	1.518.160	1.516.892	1.522.346	1.535.610	1.558.943

Note: this table shows enrolments and schools per village over the whole time period the DISE is available. The sample is the full one obtained after merging all datasets.

Table B.11: Summary Statistics - SHRUG Dataset by proximity to town

	2001		2011	
	close 10k	far 10k	close 10k	far 10k
total population	293.83	294.34	304.04	292.87
	(144.55)	(250.76)	(86.32)	(87.79)
number of households	59.67	58.39	68.61	64.65
	(27.90)	(42.26)	(22.58)	(21.77)
area (in ha)	216.86	223.01	211.15	235.69
	(267.06)	(315.41)	(222.05)	(722.95)
share of pop. SC or ST	0.32	0.43	0.33	0.44
	(0.33)	(0.38)	(0.33)	(0.38)
literacy rate	0.54	0.50	0.63	0.60
	(0.14)	(0.16)	(0.12)	(0.14)
number of primary schools	0.86	0.88	0.99	1.02
	(0.48)	(0.47)	(0.54)	(0.50)
number of middle schools	0.13	0.15	0.20	0.24
	(0.34)	(0.37)	(0.49)	(0.49)
number of secondary schools	0.02	0.03	0.05	0.06
	(0.18)	(0.17)	(0.23)	(0.25)
number of higher secondary schools	0.01	0.01	0.02	0.03
	(0.08)	(0.10)	(0.17)	(0.18)
number of colleges	0.00	0.00	0.00	0.00
	(0.03)	(0.04)	(0.04)	(0.03)
electric power for domestic use (0/1)	0.84	0.81	0.99	0.94
	(0.37)	(0.39)	(0.10)	(0.24)
electric power for agriculture (0/1)	0.25	0.14	0.68	0.41
	(0.43)	(0.34)	(0.47)	(0.49)
electric power for all end uses (0/1)	0.75	0.55	0.76	0.55
	(0.43)	(0.50)	(0.43)	(0.50)
approachable by tar road	0.54	0.36	0.58	0.46
	(0.50)	(0.48)	(0.49)	(0.50)
agriculture main source of income (0/1)			0.48	0.53
			(0.35)	(0.35)
distance from nearest 10k town (in km)			8.55	23.79
			(3.20)	(9.70)
distance from nearest 50k town (in km)			26.08	41.16
			(15.61)	(19.99)
distance from nearest 100k town (in km)			37.15	56.03
			(21.40)	(26.41)
distance from nearest 500k town (in km)			95.46	104.67
			(42.18)	(44.57)
Observations	1570	1570	1570	1570

Note: This table allows to compare the two subsamples between each other and over time. Far 10k villages are those further from town than the median village. The median distance to 10k town is 10.7km. The samples include best quality matches in 10th Plan districts, single-habitations within the 150 bandwidth only. Share of agricultural workers and distance to nearest town not available in 2001. SC and ST are Scheduled Castes and Scheduled Tribes (low castes).

Table B.12: Summary Statistics - DISE Dataset by proximity to town

	2005		2011		2017	
	close 10k	far 10k	close 10k	far 10k	close 10k	far 10k
A. Students per village						
Lower Primary, Boys	33.66	30.48	30.77	27.68	28.58	20.54
	(85.94)	(63.67)	(74.87)	(63.87)	(77.50)	(49.90)
Lower Primary, Girls	32.14	29.99	28.67	26.82	26.68	19.95
	(80.90)	(57.74)	(69.51)	(57.93)	(69.92)	(44.82)
Upper Primary, Boys	9.21	9.14	10.43	11.30	11.24	9.50
	(60.04)	(33.36)	(45.45)	(38.80)	(46.29)	(35.59)
Upper Primary, Girls	8.09	8.20	10.52	11.03	10.79	9.06
	(48.36)	(27.07)	(45.59)	(35.54)	(44.49)	(33.07)
Total, Boys	42.87	39.62	41.19	38.99	39.83	30.04
	(138.21)	(93.40)	(114.98)	(99.25)	(120.96)	(82.98)
Total, Girls	40.23	38.19	39.19	37.85	37.47	29.01
	(123.88)	(80.28)	(109.04)	(89.85)	(111.43)	(75.64)
Total	83.11	77.82	80.38	76.84	77.29	59.05
	(261.03)	(173.00)	(222.02)	(188.16)	(231.12)	(158.06)
Appeared at the C5 test	0.00	0.00	0.00	0.00	8.47	7.45
	(0.00)	(0.00)	(0.00)	(0.00)	(32.19)	(20.04)
Appeared at the C8 test	0.00	0.00	0.00	0.00	5.72	6.17
	(0.00)	(0.00)	(0.00)	(0.00)	(34.07)	(33.26)
Passed the C5 test	0.00	0.00	0.00	0.00	8.29	7.37
	(0.00)	(0.00)	(0.00)	(0.00)	(30.81)	(20.14)
Passed the C8 test	0.00	0.00	0.00	0.00	5.67	6.19
	(0.00)	(0.00)	(0.00)	(0.00)	(33.91)	(33.30)
Passed the C5 test with more than 60%	0.00	0.00	0.00	0.00	5.64	4.80
	(0.00)	(0.00)	(0.00)	(0.00)	(21.31)	(12.77)
Passed the C8 test with more than 60%	0.00	0.00	0.00	0.00	3.09	3.19
	(0.00)	(0.00)	(0.00)	(0.00)	(18.55)	(15.34)
B. School amenities						
Schools per village	1.30	1.36	1.38	1.51	1.44	1.49
	(1.83)	(1.24)	(2.01)	(1.56)	(2.10)	(1.54)
Number of Computers	0.56	0.42	0.33	0.23	0.49	0.36
	(15.23)	(9.03)	(1.83)	(1.90)	(2.58)	(1.62)
Classrooms need major repair (0/1)	0.10	0.12	0.13	0.13	0.16	0.16
	(0.25)	(0.27)	(0.28)	(0.27)	(0.29)	(0.31)
Electricity access (0/1)	0.22	0.13	0.71	0.51	0.82	0.71
	(0.40)	(0.32)	(0.44)	(0.48)	(0.37)	(0.44)
Drinking water in the premises (0/1)	0.71	0.77	0.95	0.90	0.98	0.93
	(0.45)	(0.41)	(0.21)	(0.27)	(0.14)	(0.24)
Residential Schools (0/1)	0.01	0.02	0.01	0.01	0.02	0.01
	(0.11)	(0.13)	(0.09)	(0.07)	(0.11)	(0.10)
Private (0/1)	0.02	0.01	0.03	0.03	0.04	0.03
	(0.11)	(0.08)	(0.14)	(0.12)	(0.16)	(0.12)
Co-educational (0/1)	0.99	0.99	1.00	1.00	0.99	1.00
	(0.09)	(0.10)	(0.06)	(0.04)	(0.06)	(0.04)
Approachable by all-weather roads (0/1)			1.00	1.00	0.85	0.73
			(0.00)	(0.00)	(0.34)	(0.42)
Computer-Aided Learning (0/1)			0.05	0.04	0.05	0.05
			(0.19)	(0.17)	(0.20)	(0.19)
Observations	1585	1601	1566	1635	1446	1551

Note: This table allows to compare the two subsamples between each other and over time. Far 10k villages are those further from town than the median village. The median distance to 10k town is 10.7km. The samples include best quality matches in 10th Plan districts, single-habitations within the 150 bandwidth only. Lower Primary school includes classes from 1 to 5, Upper Primary school from 6 to 8. Co-education schools are those where both girls and boys are educated. Computer-Aided Learning is defined as the integration of technology in the education process.

B.2 Validity Tests

Table B.13: RD Validity - Pre-treatment outcomes and baseline covariates

Pre-treatment Variables	RD Coefficient	SE	bandwidth	Outcome Mean	Obs.
A. 2001 Nighttime Brightness					
2001 Max Nighttime Brightness	-0.538	(0.784)	105	5.517	2,488
B. 2005 DISE Enrolments					
2005 Total Enrolments	3.389	(2.575)	100	45.09	1,918
2005 Lower Primary	1.600	(2.046)	100	40.36	1,918
2005 Upper Primary	1.952	(1.323)	102	4.760	1,962
C. 2001 Baseline Covariates					
2001 share of literate population	0.007	(0.013)	104	0.518	1,999
Distance to nearest 10k town (km)	0.866	(0.839)	100	15.77	1,918
2001 Share of ST/SC population	0.000	(0.031)	92	0.381	1,752
D. 1991 Population					
1991 Total Population	7.433	(7.052)	126	277.168	2,830

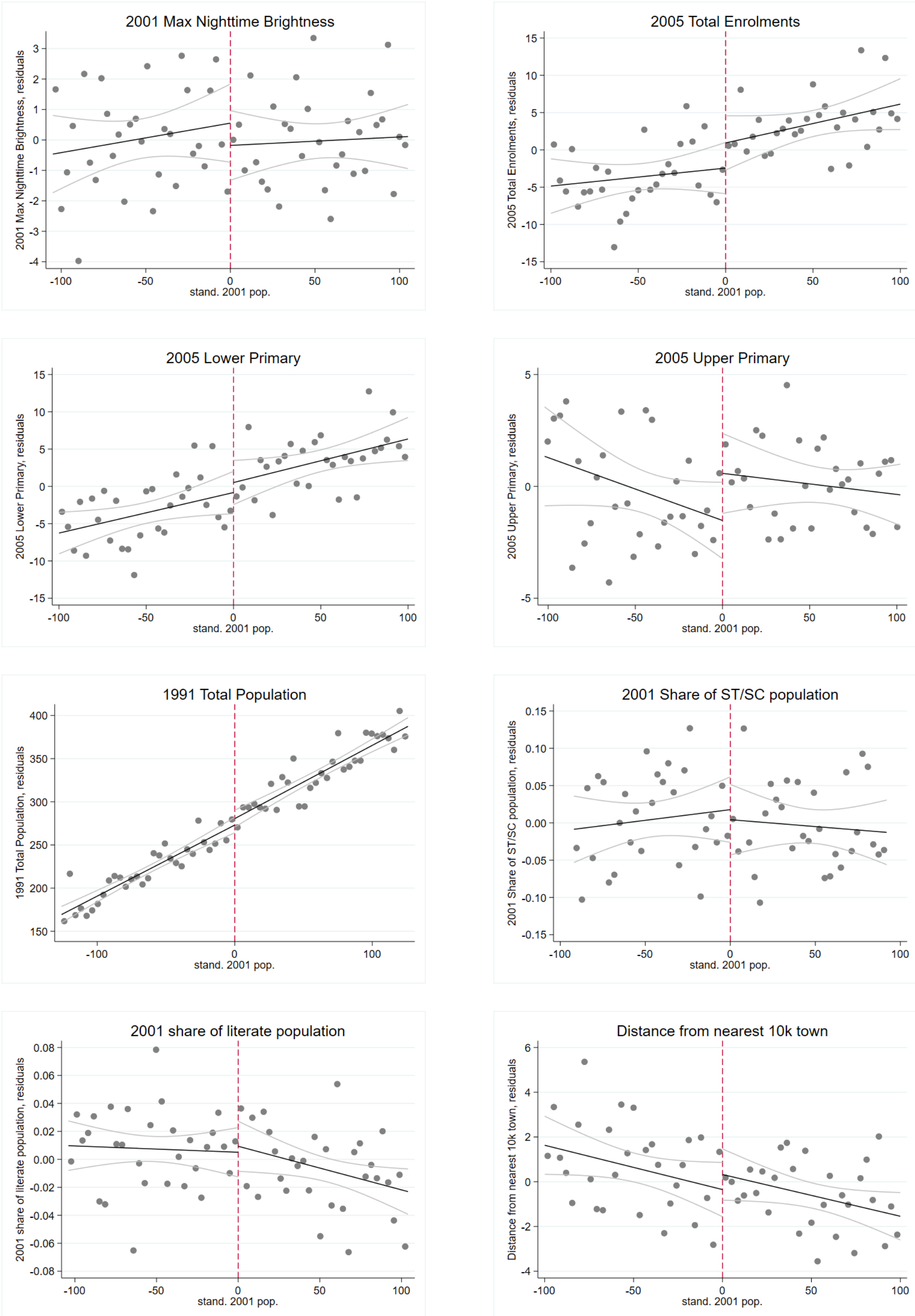
Note: Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. These specifications do not control for previous levels because data not available. 2005 data on exam results also not available.

Table B.14: RD Validity - Pre-treatment outcomes by subsample

Outcome variables	RD coef.	sd	bw	outcome mean	Obs.	RD coef.	sd	bw	outcome mean	Obs.
	Far					Close				
A. 10k town										
2005 Total Enrolments	2.439	(2.922)	127	46.09	1,555	4.068	(3.494)	152	44.65	1,025
2005 Lower Primary	1.002	(2.193)	130	40.81	1.596	2.564	(3.046)	145	40.68	977
2005 Upper Primary	1.484	(1.732)	123	5.373	1,520	0.863	(1.220)	165	3.725	1,091
A. 100k town										
2005 Total Enrolments	6.612	(4.302)	68	46.08	780	2.675	(3.161)	207	46.38	1,419
2005 Lower Primary	3.433	(2.892)	75	41.17	889	0.947	(2.727)	198	41.22	1,353
2005 Upper Primary	2.934	(2.588)	64	5.233	743	1.470	(1.402)	219	4.767	1,489

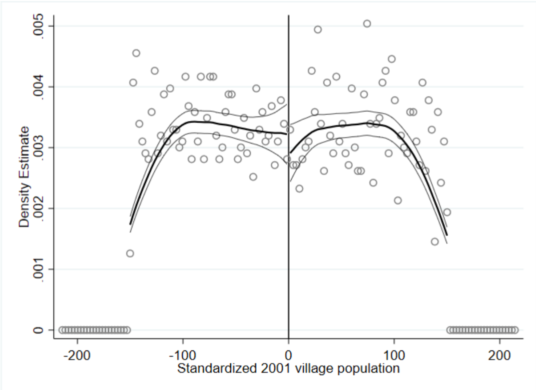
Note: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. These specifications do not control for previous levels because data not available. 2005 data on exam results also not available. Median distance from nearest 10k, 100k and 500k town are of 10.7km, 36km and 94.6km, respectively

Figure B.9: RD Validity tests - Pre-treatment outcomes and covariates



These figures show the numerical results obtained in Table B.13. The dots represent the conditional average residuals obtained from regressing the outcome variable on all covariates but the running variable. These specifications do not include lagged outcome variables because not available. Each dot contains on average 40 observations, averaged in 5-person population bins. Lines are estimated separately on each side of the 300-person cutoff. The sample contains only unique-habitation villages located in districts receiving funds under the 10th Plan and with 2001 population within the optimal bandwidth chosen for estimation based on Calonico et al. (2014a).

Figure B.10: McCrary test for manipulation



The McCrary test (McCrary, 2008) tests for discontinuity of the density of the running variable at the cutoff. The sample includes only single-habitation villages located in districts which received funds under the 10th Plan and with 2001 population within the average bandwidth of 150. The point estimate is -0.11, standard error of 0.12.

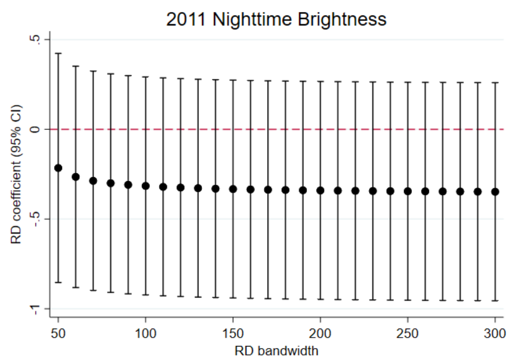
B.3 Robustness checks

Table B.15: RD Sensitivity - Controls and Fixed Effects

Variables	2011 Max Nighttime Brightness			
1[2001 pop. >= 300]	-1.056	-0.164	-0.128	-0.102
	(1.001)	(0.244)	(0.209)	(0.197)
stand. 2001 pop.	0.011	0.001	0.001	-0.000
	(0.013)	(0.003)	(0.002)	(0.002)
1[2001 pop. >= 300]*stand. 2001 pop.	-0.008	-0.001	0.000	0.001
	(0.018)	(0.004)	(0.003)	(0.003)
Observations	2,523	2,941	3,361	3,361
R-squared	0.001	0.943	0.953	0.957
2001 Control	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Baseline covariates	No	No	No	Yes
Bandwidth	106	125	143	143
Mean of dependent variable	7.269	7.466	7.436	7.436

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table allows to compare different specifications among each others. Each column adds controls to the specification.

Figure B.11: RD Sensitivity to bandwidth size



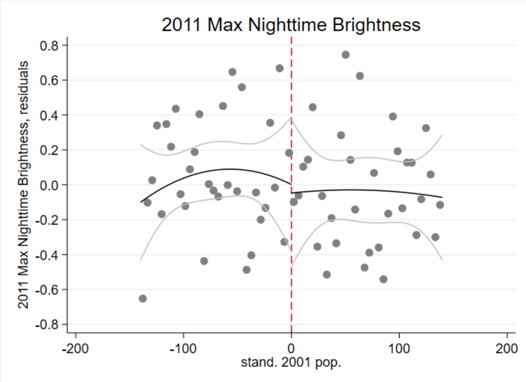
Note: This figure plots point estimates and 95% confidence intervals of the RD coefficient when eq. (1) is estimated with bandwidth sizes ranging from 50 to 300.

Table B.16: RD Sensitivity - Second order polynomial

	Max Nighttime Brightness	Total Enrolments		Lower Primary		Upper Primary		Passed C5 test with high grades	Passed C8 test with high grades
Variables	2011	2011	2017	2011	2017	2011	2017	2017	2017
1[2001 pop. >= 300]	-0.151	3.657	-1.135	3.241	-0.482	1.523	-1.051	-0.985	-0.572
	(0.292)	(6.659)	(8.222)	(4.575)	(5.984)	(3.026)	(2.763)	(0.625)	(0.489)
stand. 2001 pop.	0.001	-0.100	-0.086	-0.023	0.007	-0.132	-0.101	-0.022	-0.039**
	(0.007)	(0.219)	(0.241)	(0.125)	(0.142)	(0.090)	(0.080)	(0.021)	(0.015)
1[2001 pop. >= 300]*stand. 2001 pop.	0.002	-0.143	-0.012	-0.158	-0.054	0.117	0.128	0.081***	0.066***
	(0.010)	(0.279)	(0.296)	(0.161)	(0.185)	(0.130)	(0.104)	(0.031)	(0.022)
(stand. 2001 pop.)^2	0.000	-0.001	-0.001	-0.000	-0.000	-0.001	-0.001*	-0.000	-0.000***
	(0.000)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
1[2001 pop. >= 300]*(stand. 2001 pop.)^2	-0.000	0.003	0.002	0.002	0.001	0.001	0.001	-0.000	0.000
	(0.000)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Observations	3,289	2,686	2,662	2,877	2,824	2,798	2,838	2,414	2,838
R-squared	0.957	0.496	0.382	0.422	0.283	0.453	0.403	0.169	0.177
2001 Control	Yes								
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	140	140	152	151	163	146	164	137	164
2005 Control		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table shows the estimates obtained with a second order polynomial. Specifications add a quadratic form of the running variable as well as the interaction with the eligibility dummy.

Figure B.12: RD Robustness - Second Order Polynomial, Nighttime Brightness



This figure shows the numerical results obtained in the first column of Table B.16. The dots represent the conditional average residuals obtained from regressing the outcome variable on all covariates but the running variable. Each dot contains on average 40 observations, averaged in 5-person population bins. Second order polynomials are estimated separately on each side of the 300-person cutoff. The sample contains only unique-habitation villages located in districts receiving funds under the 10th Plan and with 2001 population within the optimal bandwidth chosen for estimation based on Calonico et al. (2014a).

Table B.17: RD - 2011 Alternative Outcomes

Variables	Non-projected maximum Nighttime Brightness	Average Total Nighttime Brightness	Average Calibrated Nighttime Brightness	Share of schools with electricity
1[2001 pop. >= 300]	-0.060	-0.018	0.001	-0.010
	(0.258)	(0.234)	(0.178)	(0.025)
stand. 2001 pop.	-0.004	-0.004	-0.003*	0.000**
	(0.003)	(0.003)	(0.002)	(0.000)
1[2001 pop. >= 300]*stand. 2001 pop.	0.006	0.007*	0.005*	0.000
	(0.004)	(0.004)	(0.003)	(0.000)
Observations	2,576	2,596	2,643	3,134
R-squared	0.743	0.759	0.754	0.521
2001 Control	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Baseline covariates	Yes	Yes	Yes	Yes
Bandwidth	109	110	112	168
Mean of dependent variable	5.344	4.829	5.630	0.603

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Average total nighttime brightness measures the yearly average brightness of each pixel contained within a village's boundaries. Average calibrated nighttime brightness are a consistent estimation of night lights over time obtained by Elvidge et al. (2014). The share of schools with electricity in each village is obtained from the DISE dataset. All specifications include pre-treatment outcome levels, village characteristics (share of literate population, share of ST/SC population, distance from 10k town) and district fixed effects.

Table B.18: RD Sensitivity - Controls and Fixed Effects

Variables	2011 Total Enrolments			
1[2001 pop. >= 300]	8.073	0.400	0.159	0.480
	(8.059)	(5.442)	(5.452)	(5.441)
stand. 2001 pop.	-0.127	-0.054	-0.052	-0.059
	(0.093)	(0.099)	(0.100)	(0.099)
1[2001 pop. >= 300]*stand. 2001 pop.	0.178	-0.026	-0.012	-0.012
	(0.151)	(0.134)	(0.134)	(0.132)
Observations	2,079	1,832	1,832	1,832
R-squared	0.002	0.517	0.532	0.538
2005 Control	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Baseline covariates	No	No	No	Yes
Bandwidth	107	95	95	95
Mean of dependent variable	62.04	62.41	62.41	62.41

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table allows to compare different specifications among each others. Each column adds controls to the specification.

Table B.19: RD Sensitivity - Controls and Fixed Effects

Variables	2017 Total Enrolments			
1[2001 pop. >= 300]	3.269	-3.212	-4.495	-4.098
	(8.070)	(6.486)	(6.898)	(6.837)
stand. 2001 pop.	-0.087	0.018	0.025	0.013
	(0.106)	(0.107)	(0.114)	(0.113)
1[2001 pop. >= 300]*stand. 2001 pop.	0.183	-0.014	-0.017	-0.010
	(0.153)	(0.140)	(0.149)	(0.146)
Observations	1,956	1,872	1,835	1,835
R-squared	0.001	0.353	0.377	0.385
2005 Control	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Baseline covariates	No	No	No	Yes
Bandwidth	110	105	103	103
Mean of dependent variable	51.98	51.88	52.17	52.17

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table allows to compare different specifications among each others. Each column adds controls to the specification.

Table B.20: RD Sensitivity - Controls and Fixed Effects

Variables	2011 Lower Primary			
1[2001 pop. >= 300]	3.500	-1.077	-0.631	-0.395
	(4.914)	(3.464)	(3.472)	(3.464)
stand. 2001 pop.	-0.025	0.003	-0.001	-0.004
	(0.060)	(0.041)	(0.040)	(0.040)
1[2001 pop. >= 300]*stand. 2001 pop.	0.080	0.008	0.009	0.006
	(0.091)	(0.057)	(0.055)	(0.054)
Observations	2,145	2,524	2,524	2,524
R-squared	0.002	0.394	0.422	0.429
2005 Control	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Baseline covariates	No	No	No	Yes
Bandwidth	111	131	131	131
Mean of dependent variable	46.77	46.91	46.91	46.91

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table allows to compare different specifications among each others. Each column adds controls to the specification.

Table B.21: RD Sensitivity - Controls and Fixed Effects

Variables	2017 Lower Primary			
1[2001 pop. >= 300]	0.626	-2.429	-3.094	-2.815
	(5.431)	(4.398)	(4.445)	(4.406)
stand. 2001 pop.	0.025	0.039	0.043	0.037
	(0.072)	(0.048)	(0.044)	(0.043)
1[2001 pop. >= 300]*stand. 2001 pop.	0.042	0.005	-0.005	-0.005
	(0.102)	(0.070)	(0.065)	(0.064)
Observations	1,909	2,281	2,399	2,399
R-squared	0.002	0.239	0.275	0.285
2005 Control	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Baseline covariates	No	No	No	Yes
Bandwidth	107	129	136	136
Mean of dependent variable	38.08	37.67	37.82	37.82

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table allows to compare different specifications among each others. Each column adds controls to the specification.

Table B.22: RD Sensitivity - Controls and Fixed Effects

Variables	2011 Upper Primary			
1[2001 pop. >= 300]	5.289	1.001	1.084	1.174
	(3.945)	(2.663)	(2.618)	(2.602)
stand. 2001 pop.	-0.148***	-0.082	-0.083	-0.086
	(0.050)	(0.052)	(0.053)	(0.053)
1[2001 pop. >= 300]*stand. 2001 pop.	0.151*	0.078	0.085	0.085
	(0.080)	(0.076)	(0.078)	(0.077)
Observations	1,997	1,673	1,660	1,660
R-squared	0.004	0.458	0.470	0.474
2005 Control	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Baseline covariates	No	No	No	Yes
Bandwidth	103	87	86	86
Mean of dependent variable	15.64	15.30	15.37	15.37

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table allows to compare different specifications among each others. Each column adds controls to the specification.

Table B.23: RD Sensitivity - Controls and Fixed Effects

Variables	2017 Upper Primary			
1[2001 pop. >= 300]	2.538	0.665	0.096	0.125
	(3.293)	(2.519)	(2.631)	(2.608)
stand. 2001 pop.	-0.096**	-0.067	-0.076	-0.079
	(0.042)	(0.061)	(0.064)	(0.064)
1[2001 pop. >= 300]*stand. 2001 pop.	0.117*	0.050	0.075	0.077
	(0.064)	(0.075)	(0.081)	(0.080)
Observations	1,973	1,537	1,507	1,507
R-squared	0.002	0.368	0.382	0.388
2005 Control	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Baseline covariates	No	No	No	Yes
Bandwidth	111	87	85	85
Mean of dependent variable	13.76	13.28	13.31	13.31

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table allows to compare different specifications among each others. Each column adds controls to the specification.

Table B.24: RD Sensitivity - Controls and Fixed Effects

Variables	Students passed the C5 test with high grades			
1[2001 pop. >= 300]	-0.339	-0.701	-0.897*	-0.884*
	(0.470)	(0.505)	(0.471)	(0.468)
stand. 2001 pop.	-0.003	-0.009	-0.009	-0.009
	(0.004)	(0.008)	(0.007)	(0.008)
1[2001 pop. >= 300]*stand. 2001 pop.	0.029***	0.043***	0.043***	0.044***
	(0.008)	(0.012)	(0.012)	(0.012)
Observations	2,568	1,898	1,898	1,898
R-squared	0.010	0.098	0.184	0.189
2005 Control	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Baseline covariates	No	No	No	Yes
Bandwidth	146	106	106	106
Mean of dependent variable	4.171	4.072	4.072	4.072

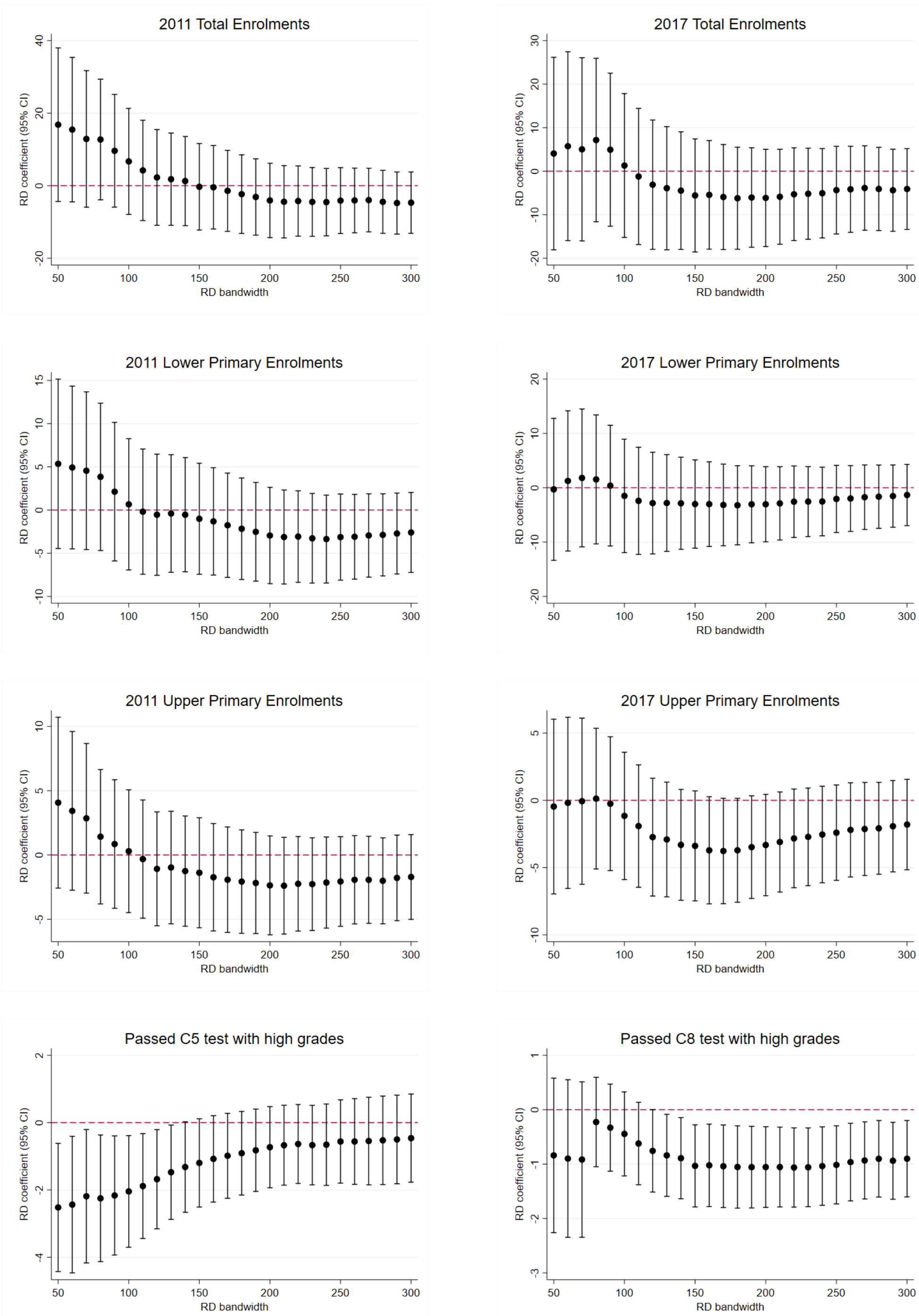
Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table allows to compare different specifications among each others. Each column adds controls to the specification. 2005 Control refers to the level of total enrolments in Class 5 in 2005

Table B.25: RD Sensitivity - Controls and Fixed Effects

Variables	Students passed the C8 test with high grades			
1[2001 pop. >= 300]	-0.576	-0.786**	-0.918**	-0.914**
	(0.455)	(0.390)	(0.391)	(0.390)
stand. 2001 pop.	-0.007*	-0.006	-0.006	-0.006
	(0.004)	(0.005)	(0.005)	(0.005)
1[2001 pop. >= 300]*stand. 2001 pop.	0.021***	0.021***	0.023***	0.023***
	(0.007)	(0.007)	(0.007)	(0.007)
Observations	2,635	2,380	2,380	2,380
R-squared	0.003	0.154	0.180	0.182
2005 Control	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Baseline covariates	No	No	No	Yes
Bandwidth	150	135	135	135
Mean of dependent variable	2.119	1.991	1.991	1.991

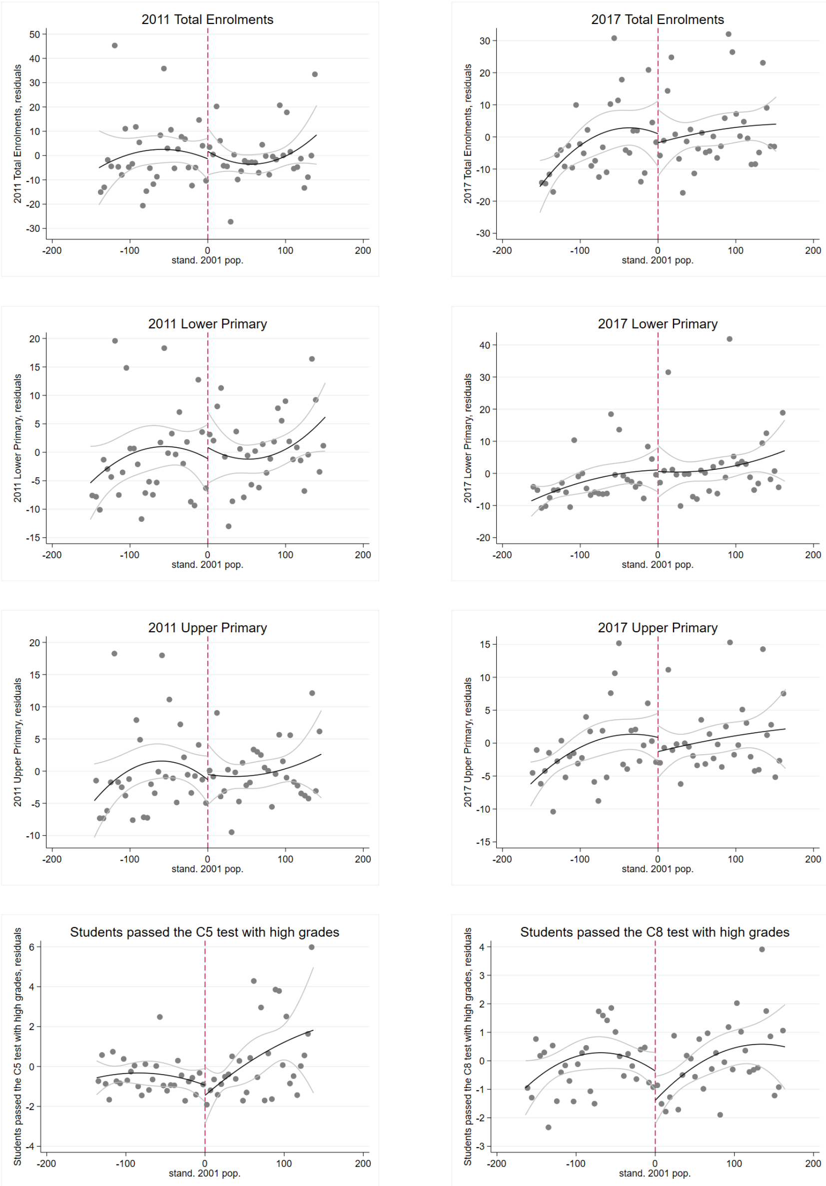
Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table allows to compare different specifications among each others. Each column adds controls to the specification. 2005 Control refers to the level of total enrolments in Class 8 in 2005

Figure B.13: RD Sensitivity to bandwidth size



Note: These figures plot point estimates and 95% confidence intervals of the RD coefficient when eq. (1) is estimated with bandwidth sizes ranging from 50 to 300.

Figure B.14: RD Robustness - Second Order polynomial, Education variables



These figures show the numerical results obtained in columns 2 to 9 of Table B.16. The dots represent the conditional average residuals obtained from regressing the outcome variable on all covariates but the running variable. Each dot contains on average 40 observations, averaged in 5-person population bins. Second order polynomials are estimated separately on each side of the 300-person cutoff. The sample contains only unique-habitation villages located in districts receiving funds under the 10th Plan and with 2001 population within the optimal bandwidth chosen for estimation based on Calonico et al. (2014a).

Table B.26: RD Robustness - Log Transformation

Variables	Total Enrolments 2011	Total Enrolments 2017	Lower Primary 2011	Lower Primary 2017	Upper Primary 2011	Upper Primary 2017	Passed C5 test with high grades 2017	Passed C8 test with high grades 2017
1[2001 pop. >= 300]	0.002	0.008	-0.015	0.031	-0.119	-0.118	-0.029	-0.077
	(0.069)	(0.079)	(0.061)	(0.081)	(0.082)	(0.097)	(0.078)	(0.056)
stand. 2001 pop.	0.000	0.001	0.002**	0.002*	-0.000	0.000	-0.001	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
1[2001 pop. >= 300]*stand. 2001 pop.	0.001	0.000	-0.000	-0.002	0.003*	0.001	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Observations	2,018	2,153	2,319	1,805	2,255	2,173	1,987	1,786
R-squared	0.423	0.332	0.383	0.287	0.644	0.567	0.319	0.340
2005 Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	104	122	120	101	117	123	112	100
Mean of dependent variable	3.701	3.429	3.504	3.224	0.976	0.971	0.993	0.350

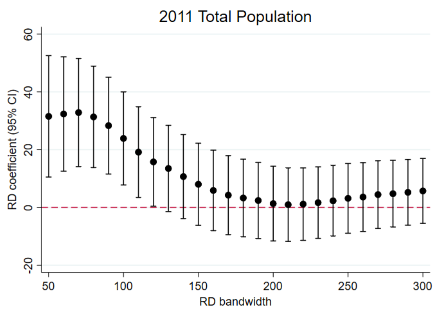
Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table B.27: RD Sensitivity - Falsification Tests

Outcome variables	10th Plan, single-hab	10th Plan, multi-hab	11th Plan, single-hab	11th Plan, multi-hab
A. 2011 Nighttime Brightness				
2011 Max Nighttime Brightness	-0.101	0.025	-0.089	-0.153
	(0.197)	(0.126)	(0.121)	(0.099)
B. DISE Enrolments				
2011 Total Enrolments	0.479	-2.660	0.055	-0.633
	(5.441)	(3.895)	(3.134)	(2.556)
2011 Lower Primary	-0.394	-1.640	0.527	0.207
	(3.464)	(2.618)	(2.107)	(1.669)
2011 Upper Primary	1.173	-0.317	-0.912	-0.846
	(2.601)	(1.776)	(1.505)	(1.312)
2017 Total Enrolments	-4.090	-0.571	-2.900	-8.340*
	(6.836)	(4.201)	(4.132)	(4.367)
2017 Lower Primary	-2.810	-0.814	-1.520	-2.110
	(4.405)	(2.906)	(2.616)	(2.098)
2017 Upper Primary	0.125	1.307	-1.540	-1.300
	(2.607)	(1.916)	(1.866)	(1.629)
C. DISE Examination Results				
Students passed the C5 test 2017 with high grades	-0.884*	-0.294	-0.451	-0.174
	(0.467)	(0.607)	(0.561)	(0.408)
Students passed the C8 test 2017 with high grades	-0.913**	0.263	0.128	1.054
	(0.390)	(0.447)	(0.330)	(1.168)

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table shows the RD coefficients and standard errors from the falsification tests. Each regression is estimated within the optimal bandwidth chosen following Calonico et al. (2014a). The second column contains the results from the actual sample of estimation. The third column contains villages which are located in districts which received funds uniquely under the 10th Plan and that contain more than one habitation within their boundaries. The fourth and fifth columns contain villages located in districts which received funds uniquely under the 11th Plan. The fourth column includes only single-habitation villages, while the fifth only multiple-habitation villages.

Figure B.15: RD sensitivity to bandwidth size



Note: These figures plot point estimates and 95% confidence intervals of the RD coefficient when eq. (1) is estimated with bandwidth sizes ranging from 50 to 300.

B.4 Heterogeneity of effect

Table B.28: Heterogeneity of effect - Power availability and Gender

Outcome variables	RD coef.	sd	bw	outcome mean	Obs.	RD coef.	sd	bw	outcome mean	Obs.
A: Heterogeneity by power availability										
	High Power Deficit					Low Power Deficit				
2011 Max Nighttime Brightness	0.071	(0.294)	126	7.991	1,802	-0.245	(0.241)	126	6.695	1,158
B: Heterogeneity by gender										
	Boys					Girls				
2011 Total Enrolments	0.199	(2.806)	103	31.26	1,997	1.070	(2.813)	88	30.85	1,695
2011 Lower Primary	0.056	(1.898)	126	23.88	2,416	-0.472	(1.660)	140	23.22	2,686
2011 Upper Primary	0.265	(1.170)	105	7.591	2,038	0.837	(1.452)	80	7.894	1,542
2017 Total Enrolments	-2.170	(3.831)	115	26.20	2,040	-1.030	(3.122)	94	25.93	1,672
2017 Lower Primary	-1.460	(2.548)	148	19.23	2,597	-1.340	(1.948)	123	18.60	2,173
2017 Upper Primary	-0.263	(1.283)	98	6.939	1,733	0.417	(1.378)	78	7.012	1,376

Note: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. The specifications include 2005 control, district fixed effects and baseline covariates. States with above-average power availability are Chhattisgarh, Orissa, Karnataka, West Bengal, Gujarat, Haryana, Rajasthan (Central Electricity Authority, 2011).

Table B.29: Heterogeneity of effect - Distance from nearest town

Outcome variables	RD coef.	sd	bw	outcome mean	Obs.	RD coef.	sd	bw	outcome mean	Obs.
	Far					Close				
A. 10k town										
2011 Total Enrolments	4.370	(6.296)	134	61.42	1,683	-17.862**	(7.133)	157	63.89	1,033
2011 Lower Primary	2.002	(5.036)	103	45.08	1,313	-6.173	(4.949)	171	50.70	1,104
2011 Upper Primary	2.914	(2.905)	120	15.86	1,521	-10.268***	(3.836)	139	15.47	932
2017 Total Enrolments	-2.737	(7.600)	118	48.29	1,395	-20.244**	(9.147)	191	59.01	1,081
2017 Lower Primary	-1.606	(5.266)	140	35.36	1,625	-8.582	(6.384)	170	43.63	985
2017 Upper Primary	-0.779	(2.617)	111	13.15	1,323	-8.457**	(4.308)	149	14.37	881
Students passed the C5 test 2017 with high grades	-0.625	(0.576)	100	3.850	1,196	-1.709**	(0.798)	184	4.477	1,052
Students passed the C8 test 2017 with high grades	-0.739	(0.501)	142	2.179	1,644	-2.436***	(0.715)	182	2.010	1,041
B. 100k town										
2011 Total Enrolments	0.709	(6.357)	139	63.39	1,661	-12.295**	(6.217)	174	63.42	1,216
2011 Lower Primary	-0.524	(4.755)	121	46.64	1,445	-5.515	(4.624)	182	48.61	1,256
2011 Upper Primary	1.886	(3.349)	114	15.98	1,373	-7.871***	(2.871)	170	14.80	1,194
2017 Total Enrolments	-7.427	(9.027)	113	49.44	1,251	-10.569	(6.875)	171	55.81	1,112
2017 Lower Primary	-4.273	(7.044)	111	35.87	1,235	-3.787	(4.554)	160	40.92	1,061
2017 Upper Primary	-1.058	(2.911)	132	13.42	1,437	-5.941**	(3.023)	130	12.98	876
Students passed the C5 test 2017 with high grades	-0.801	(0.587)	109	3.876	1,219	-0.459	(0.786)	160	4.026	1,061
Students passed the C8 test 2017 with high grades	-0.474	(0.519)	109	2.191	1,219	-1.576***	(0.573)	175	1.831	1,130
C. 500k town										
2011 Total Enrolments	-5.080	(7.272)	165	62.97	1,559	-0.711	(5.183)	113	62.91	1,085
2011 Lower Primary	-4.051	(5.080)	173	47.76	1,628	0.122	(3.499)	145	47.39	1,387
2011 Upper Primary	0.056	(3.051)	159	15.63	1,516	-0.556	(3.412)	107	15.42	1,034
2017 Total Enrolments	-8.748	(8.982)	177	49.61	1,477	-4.617	(4.779)	131	52.74	1,196
2017 Lower Primary	-3.851	(7.072)	166	36.67	1,400	-3.597	(3.401)	152	39.16	1,370
2017 Upper Primary	-5.244*	(3.091)	169	13.24	1,417	3.446	(2.765)	94	14.18	858
Students passed the C5 test 2017 with high grades	-0.734	(0.547)	166	3.462	1,400	-0.922	(0.652)	123	4.523	1,118
Students passed the C8 test 2017 with high grades	-1.080*	(0.612)	131	1.983	1,115	-1.170**	(0.510)	133	2	1,209

Note: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. The specifications include 2005 control, district fixed effects, baseline covariates excluding distance to town. Median distance from nearest 10k, 100k and 500k town is of 10.7km, 36km and 94.6km, respectively.

Table B.30: Heterogeneity of effect - Road connection

Outcome variables	RD coef.	sd	bw	outcome mean	Obs.	RD coef.	sd	bw	outcome mean	Obs.
	Connected by road					Not connected by road				
2011 Total Enrolments	-5.113	(5.530)	123	62.72	1,137	-1.182	(6.979)	140	62.98	1,416
2011 Lower Primary	-3.056	(4.222)	133	47.77	1,212	1.119	(5.602)	115	46.01	1,161
2011 Upper Primary	-2.431	(2.756)	141	15.11	1,277	2.206	(4.159)	95	16.19	967
2017 Total Enrolments	-9.354	(6.391)	135	56.59	1,093	-4.085	(9.864)	115	47.80	1,097
2017 Lower Primary	-6.421	(4.266)	138	41.63	1,115	-1.324	(7.879)	115	34.94	1,097
2017 Upper Primary	-3.431	(3.181)	123	14.88	1,010	-0.358	(3.123)	125	12.79	1,174
Students passed the C5 test 2017 with high grades	-0.595	(0.624)	154	4.031	1,223	-1.229*	(0.709)	100	3.884	966
Students passed the C8 test 2017 with high grades	-1.153*	(0.603)	176	2.026	1,337	-0.227	(0.523)	114	2.003	1,088

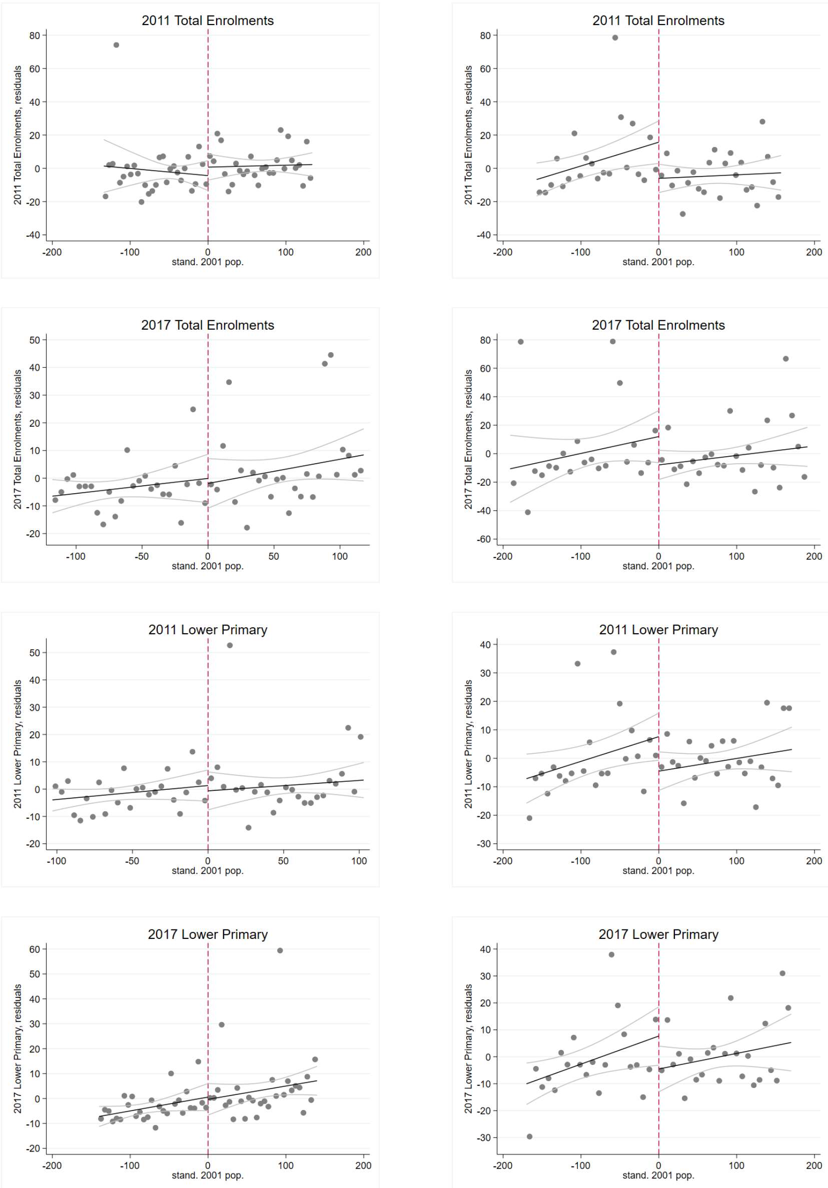
Note: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. The specification includes 2005 control, district fixed effects and baseline covariates. 48% of the villages in the sample (Plan 10, single-habitation villages, best quality match) was connected by tar road in 2001.

Table B.31: Heterogeneity of effect - Distance from nearest town

Outcome variables	RD coef.	sd	bw	outcome mean	Obs.	RD coef.	sd	bw	outcome mean	Obs.
	Far					Close				
A. 10k town										
2011 Total Population	28.65***	(11.12)	75	341.2	1,079	27.99**	(11.50)	132	343.4	1,075
A. 100k town										
2011 Total Population	44.90***	(12.05)	62	340.1	879	-1.83	(10.10)	182	340.0	1,424
A. 500k town										
2011 Total Population	45.63***	(13.93)	72	340.9	815	6.77	(8.176)	148	342.5	1,617

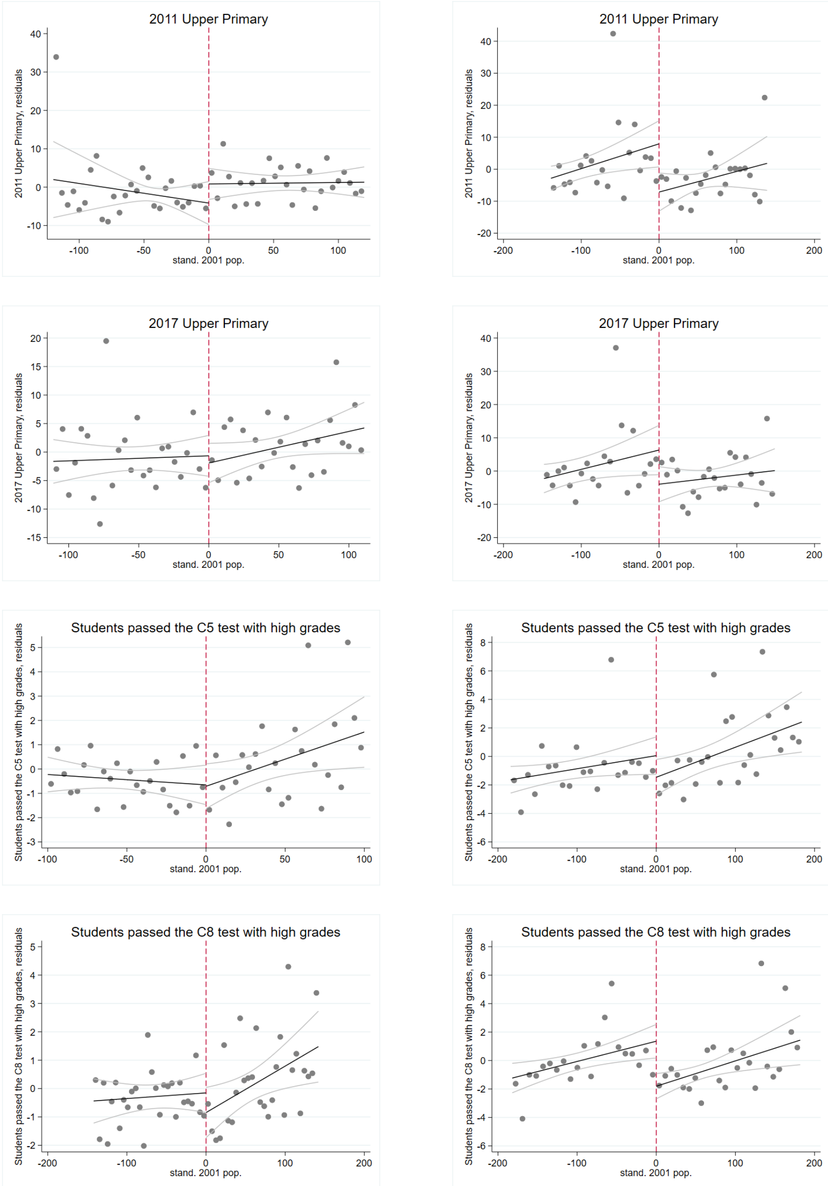
Note: Robust standard errors in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The specification includes district fixed effects and baseline covariates excluding distance to town. Median distance from nearest 10k, 100k and 500k town are of 10.7km, 36km and 94.6km, respectively.

Figure B.16: RD Heterogeneity by proximity to town



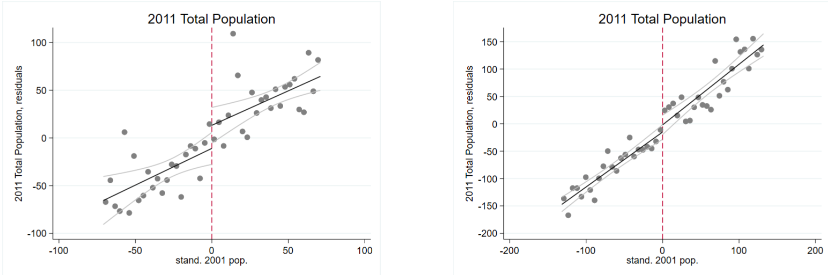
These figures show the numerical results obtained in Table B.29. The dots represent the conditional average residuals obtained from regressing the outcome variable on all covariates but the running variable. Each dot contains on average 40 observations, averaged in 5-person population bins. Lines are estimated separately on each side of the 300-person cutoff. The sample contains only unique-habitation villages located in districts receiving funds under the 10th Plan and with 2001 population within the optimal bandwidth chosen for estimation based on Calonico et al. (2014a). The first column contains villages at least 10.7km far away from the nearest 10k town. The second column contains villages which are relatively closer to the nearest 10k town.

Figure B.17: RD Heterogeneity by proximity to town



These figures show the numerical results obtained in Table B.29. The dots represent the conditional average residuals obtained from regressing the outcome variable on all covariates but the running variable. Each dot contains on average 40 observations, averaged in 5-person population bins. Lines are estimated separately on each side of the 300-person cutoff. The sample contains only unique-habitation villages located in districts receiving funds under the 10th Plan and with 2001 population within the optimal bandwidth chosen for estimation based on Calonico et al. (2014a). The first column contains villages at least 10.7km far away from the nearest 10k town. The second column contains villages which are relatively closer to the nearest 10k town.

Figure B.18: RD Heterogeneity by proximity to town



These figures show the numerical results obtained in Table B.31. The dots represent the conditional average residuals obtained from regressing the outcome variable on all covariates but the running variable. Each dot contains on average 40 observations, averaged in 5-person population bins. Lines are estimated separately on each side of the 300-person cutoff. The sample contains only unique-habitation villages located in districts receiving funds under the 10th Plan and with 2001 population within the optimal bandwidth chosen for estimation based on Calonico et al. (2014a). The first column contains villages at least 10.7km far away from the nearest 10k town. The second column contains villages which are relatively closer to the nearest 10k town.