Jobs, Schools and Roads: The Long Run Economic Development of India

by

Gaurav Khanna

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2016

Doctoral Committee:

Professor John Bound, Co-Chair
Associate Professor Dean C. Yang, Co-Chair
Assistant Professor Achyuta R. Adhvaryu
Professor Jeffrey A. Smith

For my family who taught me to argue incessantly.

And for MM, who has read and edited this dissertation many more times than I have. Please direct complaints accordingly.

ACKNOWLEDGMENTS

I'd like to thank my chairs, John Bound and Dean Yang, and my advisors, Achyuta Adhvaryu and Jeff Smith for their patience and sound advice. I am grateful to Austin Davis for tolerating me and supporting me, to Gretchen Lay for inspiration, to Katherine Lim for keeping me on track, to Ben Thompson for helping me re-evaluate my purpose in life, and to Meera Mahadevan being a part of the purpose. I'd also thank my family who let me disappear, and only re-emerge once a year.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Jobs, Schools and Roads:
The Long Run Economic Development of India

by

Gaurav Khanna

Chair: John Bound and Dean Yang

My dissertation uses novel empirical methods to explore large, national-level issues that affect developing economies across several dimensions. I study a variety of issues such as unequal educational access and attainment, labor market consequences of educational investments, affirmative action programs, nationwide infrastructure programs and regional development. One aspect of my dissertation focusses on measuring the causal effects of large-scale government initiatives, which, unlike smaller researcher-led interventions, come with a unique set of challenges like addressing significant general equilibrium effects, imperfect implementation, and sporadic or poor-quality data. I consolidate conventional data from household surveys and census data, with data from more innovative sources like satellite imagery, geospatial data, state-level legislation and news reports. I also employ innovative econometric techniques and various instrumental variable strategies in order to overcome data and identification difficulties and comprehensively explore the efficacy of large-scale policies in developing countries.

CHAPTER 1

Large-scale Education Reform in General Equilibrium: Regression Discontinuity Evidence from India

The economic consequences of large-scale government investments in education depend on the general equilibrium (GE) effects in both the labor market and the education sector. I develop a novel general equilibrium model and derive sufficient statistics that capture the economic consequences of a massive countrywide schooling initiative implemented by the Indian government. I provide unbiased estimates of the sufficient statistics using a Regression Discontinuity design. The earnings returns to a year of education are 13.4%. The general equilibrium labor market effects are substantial: they depress the returns to skill and dampen the increase in economic benefits. These GE effects have distributional consequences across cohorts and skill groups, where as a result of the policy unskilled workers are better off and skilled workers are worse off. In the education sector, more private schools enter these markets negating concerns of crowd-out. These results indicate that researchers and policymakers need to consider the GE effects when scaling up micro-interventions.

1.1 Introduction

Large-scale educational expansions represent substantial investments of public resources and benefit households by increasing productivity in the local economy. However, since they impact both individual behavior and labor market outcomes, convincing causal estimates of their overall economic benefits are hard to generate. While small-scale, carefully controlled, researcher-led experiments provide promising evidence about which educational investments are effective, for a variety of reasons these estimates may not be valid for large scale policies. Importantly, large-scale education programs may have sizable general equilibrium (GE) effects in the education sector and the labor market that may undermine the effectiveness of the intervention. In the labor market, an expansion in schooling makes skilled workers more abundant and lowers their equilibrium earnings. At the same time, an increase in the size of the skilled workforce may attract capital and technology, benefiting skilled workers. In the education sector, additional public schools may simply crowd-out private schools, diminishing the benefits from schooling expansions. I causally estimate and take into account these GE effects in determining the overall economic consequences and benefits of a nationwide education program.

I build a new framework to analyze the consequences of a large-scale educational expansion program in India with an explicit focus on issues inherent to nationwide government policies: the persistence of effects and the consequences of lost funding, and GE effects in the markets for both education and labor. I model the labor market and education sector and decompose wage changes into the individual returns to education and the GE effects. I then estimate the sufficient statistics generated by the model using a Regression Discontinuity (RD) approach. Further, I exploit variation in cohort exposure to the program and skill levels to identify the GE effects, by estimating how the earnings skill-premium changes across local economies. I use the estimated parameters to comprehensively measure the overall benefits of the policy and its distributional consequences across skill levels and age cohorts. Not only do I find substantial GE effects in the labor market, but I am also able to precisely estimate their size—in my context these effects depress the returns to skill by 32% and dampen the increase in labor market benefits by 23%. By increasing the size of the skilled workforce, the policy makes skilled workers worse off and unskilled workers better off. At the same time, the GE effects in the education sector suggest a crowd-in of private schools, negating concerns of crowd-out.

From a novel model of households, public schools, private schools, and firms, I derive sufficient statistics that measure the effect of the program on welfare. In the education sector, I model the entry and exit decisions of private schools, the spending decisions of public schools, and household decisions to attend school. On the labor market side, I combine two sets of canonical models. The first is an education choice model that determines one individual's optimal level of education for a given set of wages (Becker, 1967; Mincer, 1958; Willis, 1986). The second is a labor supply and demand framework that determines the set of wages for a given distribution of education, where changes in the skill-distribution of the workforce affect the skill-premium in wages (Card and Lemieux, 2001).[1] Combining these models allows me to study how the distribution of earnings affects education choices, and consequently how changes in education choices affect the distribution of earnings.

The returns to education and the change in the returns due to the GE effects are among the model's important sufficient statistics. While a well implemented policy can effectively increase the supply of schooling, equilibrium schooling may not change much if the returns to education are low (Jensen, 2010, 2012).[2] If overall education levels do rise because of the interventions, we expect earnings and therefore the returns to be affected in a few ways. First, a more educated worker is more productive and will earn a higher wage. Second, a more educated worker may reside in a region where there are fewer educated workers, making her relatively more valuable in the labor market. But, if large numbers of people receive additional education, there is also a GE effect in the labor market: an increase in the abundance of high-skill labor puts downward pressure on their wages. At the same time, as more skilled workers join the labor force, skill-biased capital may flow into these regions, raising their earnings. Last, as workers switch to more productive skill groups, overall output may increase to the benefit of all workers. I, therefore, estimate all components of the GE effects to better quantify the distributional impacts and the overall increase in labor market benefits.

The policy I study is the District Primary Education Program (DPEP), which expanded public schooling in half the country by targeting low-literacy regions. It was the India's flagship scheme in the 1990s and early 2000s. At that time it was the largest program for primary education in the world, in terms of geography, population and funding, suggesting that its effects would be similarly broad (Jalan and Glinskaya, 2013). The policy primarily built schools, hired teachers and upgraded infrastructure in low-literacy districts. Such schooling expansions reduce the marginal cost of attaining education by improving access to schools (Behrman et al., 1996; Birdsall, 1985). This will induce some students who have potentially high returns to schooling but could not previously afford it to get more education. Duflo (2001) shows that a similar program in Indonesia increased education levels and earnings for eligible cohorts. Similarly, I examine not only the educational outcomes, but also earnings for different cohorts and skill groups long after their exposure to the policy. Additionally, I measure how educational inputs such as teachers and classrooms respond to decreases in funding in the long run.[3]

The allocation rule under which districts receive the funding allows me to estimate the parameters of the model using a Regression Discontinuity (RD) approach. Since districts that had a female literacy rate below the national average were more likely to receive the program, I can compare regions on either side

[1]There is a long literature that shows how the changes in the skill distribution affect returns in the labor market (Bound and Johnson, 1992; Katz and Murphy, 1992).

[2]Changes in returns will induce changes in education - Khanna (2014) finds that, in India, an increase in the probability of getting a government job because of affirmative action policies incentivizes low caste students to not drop out of school. At higher education levels, Khanna and Morales (2015) studies how an increase in returns to IT sector jobs in the US and India affects graduation from engineering schools in the 1990s.

[3]In an accompanying paper, I complement these results by studying whether more decentralized decision making in such policies is more effective (Khanna, 2015).

of the literacy-rate cutoff to determine the causal impact of the policy. The RD design allows me to tackle biases that may arise when estimating the individual returns to education, and when comparing earnings in two different local economies. At the individual level these biases may arise if more competent workers are also more competent at school. The RD allows me to compare students who were induced into getting more education to similarly competent students that were not. Furthermore, some regions may have a large number of skilled workers or industries that require skilled work, and are therefore not comparable to other regions. At the regional level, therefore, the RD tackles biases that arise due to differences in the local economy and labor market.

To support each piece of the general equilibrium model, I create a comprehensive dataset by combining three waves of a household survey, a census of firms, school-level data, and the Indian Census. I use the data to estimate the returns to education and the GE effects, exploiting not just the RD, but also the variation in cohort exposure and skill levels. Younger cohorts can change their educational attainment in response to the policy, whereas older cohorts cannot. Both the young and the old are, however, affected by changes in the labor market skill distribution. Furthermore, since the young and the old are not perfect substitutes in the labor market, the GE effects on the two cohorts must be estimated separately. Using the estimated parameters, I measure the GE effects and the overall impact of the policy on household welfare for the different types of workers and cohorts.[4]

Given evidence from other contexts, it is crucial for researchers to address these labor market GE effects. In the US, Abbott et al. (2013); Heckman et al. (1998a,b); Lee (2005) show how changes in taxes or tuition and financial aid may have large general equilibrium effects. In developing countries as well, an increase in the size of the skilled workforce have been found to depress wages for high-skill workers (Angrist, 1995; Duflo, 2004).[5] I both flexibly model and causally estimate the GE effects on different cohorts and on different skill groups, allowing me to determine distributional consequences across both dimensions. I estimate the earnings skill-premium by age group separately on either side of the RD cutoff. The difference in the earnings skill-premium for older workers allows me to measure the GE effects that affect all cohorts. At the same time, there is an often ignored, additional impact on younger workers which I estimate by looking at the additional change in the skill-premium for young workers.

There are already a substantial number of micro-interventions in India that can help guide policy-makers in supply-side interventions.[6] These micro-interventions are, however, inherently different from large

[4] I corroborate my results with a Difference-in-Differences (DID) analysis similar to Duflo (2001), where I compare treated to untreated districts and the young to older cohorts. Using a DID design, however, it is hard to recover the entire extent of the labor market GE effects as the portion of the GE effects that affect all cohorts are differenced out by the DID estimator. The advantage of the RD is that it allows me to estimate the entire extent of the GE effects, and disentangle them into the portion that affects all cohorts and any additional impact on treated cohorts.

[5] There are other types of labor market GE effects in other contexts: Crepon et al. (2013) highlight the possibilities of negative externalities in job assistance programs, Albrecht et al. (2009) calibrate a macro model of the change in the labor market equilibrium due to the Swedish Knowledge Lift program, and more recently Bianchi (2016) shows how college major choice in Italy can affect returns.

[6] These studies cover a wide gamut of programs like library programs (Borkum et al., 2010), teacher training (Kingdon and Teal, 2010), teacher incentives (Muralidharan and Sundararaman, 2010), computer-aided programs (Linden, 2008), remedial

school expansion programs since they do not have GE effects. While the evidence on smaller changes of inputs within schools is mixed (Muralidharan, 2013), large-scale investments in schooling expansions like the one studied here, have been found to be relatively more successful across the world.[7]

A concern with an expansion in public schooling is that it may crowd out private supply. On the other hand, a crowd-in could also have occurred if the program increased the overall size and the demand for a skilled workforce. Understanding how private schools respond is, therefore, essential to identifying the overall benefits, as a large amount of crowd-out implies that the funds are essentially wasted and could have been spent on other programs. In my context, I model and estimate this change in private supply. In line with other work (Andrabi et al., 2013), I find an influx of private schools when public schooling is expanded.

To track long-run outcomes, I assemble a 10 year long panel of districts that allows me to follow local labor and education markets over time. While studies have found that policies that lower the costs of schooling have positive impacts in the short run, the existing evidence on the persistence of impacts is mixed (Angrist et al., 2006; Das et al., 2013a). I find that while there was a net increase in the number of new schools built over this period, only a few of these schooling inputs last in the long run. Once the funding is phased out, the physical infrastructure upgrades remain but the differential effects on more qualified teachers dissipates.[8]

I find that the program increased both education and earnings for students in targeted regions. There are large overall economic benefits to households that are driven by reductions in the costs of education and an increase in the overall output of the region. However, general equilibrium effects substantially mitigate the rise in labor market earnings for those who acquire more skill. Increases in the supply of educated workers dampened earnings for skilled workers and put upward pressure on the earnings of unskilled workers. The returns to skill are 13.4%, but the estimated labor market GE effects are substantial – for a 17 percentage point increase in the fraction of skilled workers, the GE effects depress the returns by 6.5% points and dampen the increase in benefits to students by 23%. These GE effects have distributional consequences, with a transfer of labor-market benefits from skilled to unskilled workers, particularly among the younger cohorts. High-skill workers who did not change their educational levels under the policy are adversely affected in the labor market, whereas low-skill workers benefit. Importantly, mobility

---

education (Banerjee et al., 2010, 2007) and class sizes (Banerjee et al., 2007; Jacob et al., 2008; Muralidharan, 2013). Some often cited reasons for low educational outcomes are teacher quality and high levels of teacher absence (Das et al., 2013c; Duflo et al., 2012; Muralidharan, 2013).

[7] Some examples are in Indonesia (Duflo, 2001), Burkina Faso (Kazianga et al., 2013), Zimbabwe (Aguero and Bharadwaj, 2014), Nigeria (Osili and Long, 2008), Sierra Leone (Cannonier and Mocan, 2012), Uganda (Deininger, 2003), Zambia (Ashraf et al., 2015), Kenya (Bold et al., 2013b), Tanzania (Sifuna, 2007), West Bank & Gaza (Angrist, 1995), and India (Afridi, 2010; Chin, 2005).

[8] There is also a large literature in the US that studies whether spending on education affects educational outcomes (Card and Payne, 2002; Hanushek, 1997, 2003, 2006; Hoxby, 2001; Krueger, 2003). One crucial result from the US literature is that not all schooling inputs have similar impacts, and so it is necessary to understand which inputs matter (Grogger, 1996; Hanushek, 1986, 2008; Krueger, 1999; Loeb and Bound, 1996). This is why I extensively study the changes in a whole host of schooling inputs, from teachers to physical infrastructure.

of skill-biased capital does play a role, however small, in mitigating the GE effects. But consistent with the other literature in this context (Munshi and Rosenzweig, 2015), I find no evidence of labor mobility.[9]

The methodology developed in this paper accounts for the general equilibrium effects of large-scale government spending and finds them to be substantial. In doing so, it improves upon the literature estimating the private returns to education that exploits changes to tuition reductions, compulsory schooling laws, schooling expansions or other large-scale policy reforms, but largely ignores the broader adjustments in the labor market and education sector. Consistent with the theory I build in this paper, I find in the Indian context, that labor market GE effects dampen private benefits to students that attain more education and have substantial distributional consequences. The results in this paper indicate that these wage responses may undermine some of the effectiveness of micro-interventions when they are scaled up (Acemoglu, 2010; Deaton, 2010).[10] On the other hand, the crowd-in of private schools indicate that large-scale public schooling expansions may have other unintended benefits in the education sector. I also show that once the funding was phased out certain crucial inputs, such as well qualified teachers, no longer remain. These empirically important consequences are vital considerations for both researchers and policy-makers who examine or implement large-scale interventions.

1.2 The District Primary Education Project (DPEP)

I use exogenous variation generated by a large schooling expansion policy (the District Primary Education Project (DPEP)) implemented by the Indian government. The government selected districts based on the prevailing female literacy rate, which allows for a RD design. I compare districts that should have received the policy to those that should not have, on either side of the RD cutoff, to causally estimate the parameters of the model. This was also a time of rapid growth and development in the Indian economy, which is why the RD is necessary to isolate the impact of the policy from other changes.

In 1992, the Indian Parliament updated their National Policy on Education with a renewed focus on primary and upper primary education. Based on recommendations from the Central Advisory Board of Education, the Parliament amended the constitution and transferred education-related decisions to local bodies, and stressed the decentralization of decision making by helping districts plan and manage both primary and upper primary education.[11]

[9] In an accompanying paper (Khanna, 2015), I compare this policy to more decentralized policies implemented in the following decade. The decentralized policies targeted sub-districts in a way that allows me to use a Multi-Dimensional Regression Discontinuity (MRD) framework. Since MRD is new to the literature, I provide Monte Carlo evidence about the best possible estimators in such contexts. I find that the decentralized policies were relatively more effective in raising literacy rates, suggesting that there are benefits to using local administrative bodies that may have better information.

[10] This point has often been made outside the realm of Development Economics as well (Heckman et al., 1999).

[11] Primary is usually grades 1 through 4 or 5, and upper primary is grades 5 or 6 through 8.

In 1994, the District Primary Education Project (DPEP) was introduced in seven states and 42 districts, and was over time expanded to 271 of approximately 600 districts in the country. The project spanned four phases, the last of which were implemented in the mid-2000s. While a portion of the funds were released under DPEP through the mid-2000s, the bulk of the funding ended in 2005 when other policies under the newer Sarva Shiksha Abhiyan (SSA) were growing in strength.[12] In 2006, only 2 states received any money, and after 2007 none did.[13]

DPEP grew to consist of seven projects, with funding from the World Bank, the European Commission (EC), the U.K. Department for International Development (DFID) and Official Development Assistance (ODA), the Royal Government of the Netherlands, and UNICEF, making it one of the largest donor assisted programs in the world (Jalan and Glinskaya, 2013). The Central government contributed 85% of the amount using these donor funds.[14] The remaining 15% came from states, which had to maintain the level of expenditure that existed before the program was implemented in an attempt to ensure that there was no crowd-out of state funds.[15] However, states did have the ability to re-allocate funds across districts.

Figure 1.1 shows foreign aid earmarked for primary and upper primary education only, and also the amount spent specifically on DPEP. Both amounts rose steadily in the 1990s till they peak in 2002-3. When foreign aid was cut in 2003, so was DPEP expenditure, which fell steadily over the next few years. Foreign aid spiked up again in 2005-6 as the next policy – the SSA – was ramped up. Figure 1.2 shows how much the Central government transferred to the states for social sector spending (health and education). This amount, and the share of total transfers, rose steadily during this period. The figures make it clear that this was a period of a large increase in externally-financed expenditure on education, most of which was concentrated in less than half the districts of the country, allowing for a valuable policy experiment.

The broader program claims to have covered about 271 low literacy districts, and served approximately 51.3 million children and 1.1 million teachers in about 375,000 schools (Jalan and Glinskaya, 2013).

[12] SSA was similar to the DPEP, but covered the entire country. There were, however, certain programs under SSA that targeted certain sub-districts.

[13] The phase-out was fairly rapid. In the 2002-3 financial year, the government spent approximately $345 mn on DPEP, whereas in the 2006-7 financial year, they spent only $24 mn on it. Even though taxes were not raised to fund the DPEP, when the shift to the newer SSA program happened, the government levied a 2% education tax to fund an expansion to all districts.

[14] India has received aid on various social and infrastructure programs, and in 2005-6 alone it received $4 bn (Colclough and De, 2010). By 2002 the World Bank alone had committed about $1.62 bn on DPEP, whereas the other donors concentrated on certain states. For example, in the first few years of the program, the EC spent ECU 150mn in Madhya Pradesh, the Netherlands spent $25.8 mn in Gujarat, DFID spend 80 mn pounds in Andhra Pradesh and West Bengal, whereas UNICEF spent $ 153 mn in Bihar (GOI, 2000). WorldBank (1997) claims that in 1993, the EC provided a grant of ECU 150 mn, whereas the World Bank approved credits of $265 mn in 1994 and $425 mn in 1996. At the time of the transfer to the wider SSA program in 2004, the World Bank's contribution consisted of less than half of the external aid funds, with DFID and the EC being the other major donors. Between 2004 and 2007 alone, about $7.8 bn was spent on the expanded SSA program, including the Government's contributions (Ayyar, 2008).

[15] Varghese (1994) claims that states had to maintain their educational expenditures at at least their 1992 values, and World-Bank (1997) guidelines claim states had to maintain the same growth rate in educational expenditure.

These districts were geographically dispersed all over the country as can be seen in the map in Appendix Figure 1.12. It created 63,000 new schools, including more than 50,000 'alternative' or 'community schools', and trained about 1 million teachers and 3 million community members.[16] Within the state, there was major inter-district variation in planning and management as the districts had the flexibility to allocate funds. In the project states, it increased the average allocation of funds for primary school education by between 17-20% (Jalan and Glinskaya, 2013).[17]

The objectives of the project were two-fold. The first was to improve student access to and retention in primary and upper primary education by building schools, supporting school and community organizations, constructing new classrooms and improving existing school facilities. The guidelines of the program stipulated that the "*project would be a reconstruction of primary education as a whole in selected districts instead of piecemeal implementation of schemes*" (GOI, 1994). While most of the funds were directed towards the government schools, some were used towards a training drive for teachers of private and government-aided schools.[18]

The second objective was to improve the access to primary and upper primary education by establishing district institutions to decentralize planning.[19] The programs were developed by each participating district and appraised by the Bureau that also provided implementation support. The programs were evaluated and the poorly performing subprojects are dropped.

There are numerous World Bank and Government of India briefs and media reports that refer to the program's success.[20] The most in-depth investigation, however, is a working paper by Jalan and Glinskaya (2013) that uses a difference-in-differences methodology to compare the enrollment rates for students in the 42 districts in the first of the four phases to other districts. They find that five years after the program started, enrollment and grade progression of minority groups only in some specific states improved. Furthermore, grade progression for boys in certain states was higher, but there were little to no impacts on girls. Over the entire period, districts were not allowed to receive more than $8 million, which came to approximately $9.1 per student. Jalan and Glinskaya (2013) estimate that this intervention lowered the

[16] Alternative or community schools are part of the non-formal schooling system. They provide the basic schooling infrastructure to remote areas and disadvantaged groups with the help of the local community.

[17] In this period, DPEP was the flagship education program, despite being restricted to less than half the country. For example, in 2001 alone, the Ministry of Human Resource Development, estimates spending $275 mn on DPEP for the limited number of districts. The second and third largest expenditures were on schemes that covered all districts like the Mid-day Meal Scheme ($232 mn), and Operation Blackboard ($130 mn).

[18] The guidelines of the policy also discussed the local community initiatives in promoting enrollment and retention. For example, Village Education Committees and local bodies like the Mother-Teacher Associations were tasked with creating local awareness campaigns and getting more children into schools and preventing them from dropping out of schools.

[19] Specifically, this was to be done by managing the delivery of education, including teacher support and materials development through Block Resource Centers (BRC) and Cluster Resource Centers (CRC), and strengthening the District Institutes of Education and Training (DIET). This also included targeted interventions for girls and minority groups, and the expansion of Early Childhood Education (ECE). The program established a DPEP Bureau in the Ministry of Human Resource Development that served as a financial and technical intermediary. They appraised, monitored and supervised the district programs.

[20] See World Bank Report (2003), "World Bank praises India for DPEP" Economic Times, (Sep 2005) and Government of India (2011).

private household costs of schooling by between 20 to 40%. Their paper uses two repeated cross sections of enrollment to look at the short-run impacts on the few districts in the first phase of the program. In contrast, I use the RD design and look at the longer run effects fifteen years after the program started, and after all the phases were implemented. Other descriptive studies examine the outcomes for DPEP districts without comparing them to other districts (Aggarwal, 2000; Menon, 2001; Pandey, 2000), and hence cannot distinguish between the changes in overall education taking place all across the country driven by a robust economic growth, and the changes specifically attributable to the program.

1.3 The Model, Comparative Statics and Economic Benefits

I set up a model that captures the salient features of the local economy and the market for education, including the general equilibrium effects. The model will identify sufficient statistics that determine the effect of schooling expansion policies on economic benefits. On the labor market side, I combine two sets of canonical models. The first is a returns-to-education model (Becker, 1967; Card, 1999; Mincer, 1958), which determines one individual's optimal level of education for a given distribution of wages. The second is a labor market model (Card and Lemieux, 2001) that determines the equilibrium distribution of wages for a given distribution of educational skill levels. By combining them, I can study how changes in the education distribution affect the distribution of wages, and vice versa, allowing me to identify the general equilibrium effects in the labor market.[21]

The demand for education is determined by students' optimization decisions, whereas the supply depends on the choices made by both public and private schools. Furthermore, how students react to changes in the education sector depends on the labor market returns and the general equilibrium effects on earnings. Building new schools and increasing access to schools will reduce the marginal cost of schooling (Behrman et al., 1996; Birdsall, 1985); directly affecting household welfare, and increased schooling will lead to payoffs in the labor market.[22]

1.3.1 Economic Production and the Labor Market

A final good is produced in the local economy using capital and labor as inputs, where the labor input varies by cohort and skill level. Capital is perfectly elastically supplied across districts at rental rate R^*.[23] Effective labor supply in district d is L_d, and depends on the labor aggregate at each skill level. The production function is a modification of the nested Constant Elasticity of Substitution (CES) function

[21] Here and elsewhere, I will be using the terms education and skill interchangeably.

[22] New schools reduce transportation costs, and lower the market price by expanding supply, whereas improvements in quality make it easier for students to complete the grade.

[23] The perfectly elastic capital assumption is not essential. The results are unaffected by assuming a fixed capital stock (see Section 1.12.2).

proposed by Card and Lemieux (2001). Labor aggregates at each skill level L_{sd} are represented by their skill level s. This production function helps me disentangle the various components that contribute to the returns to a higher level of skill. Aggregate output Y_d in district d depends on L_d (effective labor) and K_d (capital):[24]

$$Y_d = L_d^{\varrho} K_d^{(1-\varrho)} \tag{1.1}$$

$$L_d = \left(\sum_s \theta_{sd} L_{sd}^{\frac{\sigma_E - 1}{\sigma_E}} \right)^{\frac{\sigma_E}{\sigma_E - 1}} , \tag{1.2}$$

where $0 < \varrho < 1$ is the share of output accruing to labor, $\theta_{sd} > 0$ is the productivity of workers with education or skill level s, and $\sigma_E > 0$ is the elasticity of substitution across education or skill groups. The productivity parameter θ_{sd} captures the productivity of each skill level, and increases with an increase in skill-biased capital in the district k_{sd}, such that $\theta'_{sd}(k_{sd}) > 0$.[25] The value of θ_{sd} therefore varies across districts only because of the variation in skill-biased capital k_{sd}. The aggregate supply of workers at skill level s depends on the aggregate effective supply of workers in each skill level ℓ_{asd} in a given age cohort a:

$$L_{sd} = \left(\sum_a \psi_a \ell_{asd}^{\frac{\sigma_A - 1}{\sigma_A}} \right)^{\frac{\sigma_A}{\sigma_A - 1}} \tag{1.3}$$

Here, σ_A is the elasticity of substitution across age cohorts, and ψ_a is the productivity of a specific cohort. The effective supply ℓ_{asd} may depend on the ability of workers ϵ_i.[26] A worker gets paid their marginal product. The average log earnings are therefore:[27]

$$\log w_{asd} = \log \left(\frac{\partial Y_d}{\partial \ell_{asd}} \right) = \log \widetilde{\varrho} + \log \theta_{sd} + \log \psi_a + \frac{1}{\sigma_E} \log Y_d + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E} \right) \log L_{sd} - \frac{1}{\sigma_A} \log \ell_{asd} \ , \tag{1.4}$$

where $\log \widetilde{\varrho} \equiv \left[\left(1 - \frac{1}{\sigma_E} \right) \left(\frac{1-\varrho}{\varrho} \right) \log \left(\frac{1-\varrho}{R^*} \right) \right]$ is common across all districts and workers.[28]

---

[24] As explained in Section 1.4.1.1 the estimation procedure and measured GE effects do not depend on the functional form of the production function. Furthermore, adding non-tradables like land into the aggregate production function Equation (1.1) does not directly affect the estimation strategy. The policy will theoretically change the value of non-tradables; however, I will be concentrating on the earnings of workers, and not be examining the returns to owners of capital and land.

[25] For completeness, in Appendix 1.12.3 I explicitly model skill-biased capital within the nested CES framework and show how flexible ways of incorporating it do not affect the estimation or results.

[26] For instance, the effective supply $\ell_{asd} = \sum_i \epsilon_i \ell_{asdi}$.

[27] This is at the optimal value of K_d^*, so that $Y_d = \left(\frac{1-\varrho}{R^*} \right)^{\frac{1-\varrho}{\varrho}} L_d$.

[28] For convenience, I have ignored the role played by changes in prices. It is easy to include a $\log P_d$ that will be clubbed

There are a few components that drive the differences in average earnings when comparing two different types of people in two different labor markets represented in Equation (1.5):

$$\log\left(\frac{w_{asd}}{w_{a's'd'}}\right) = \underbrace{\log\left(\frac{\theta_{sd}}{\theta_{s'd'}}\right)}_{\text{productivity}} + \underbrace{\log\left(\frac{\psi_a}{\psi_{a'}}\right)}_{\text{cohort}} + \underbrace{\frac{1}{\sigma_E}\log\left(\frac{Y_d}{Y_{d'}}\right)}_{\text{output}} + \underbrace{\left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right)\log\frac{L_{sd}}{L_{s'd'}}}_{\text{skill distribution}} \underbrace{-\frac{1}{\sigma_A}\log\frac{\ell_{asd}}{\ell_{a's'd'}}}_{\text{skill-cohort distribution}} \tag{1.5}$$

This equation is crucial in that it captures why earnings are systematically different across people and across labor markets. The first component – 'productivity' – θ_{sd} is the higher productivity associated with more education. Not only are skilled workers more productive, but variation in the supply of skill-biased capital across districts will affect earnings as well. The second component – 'cohort' – captures the age-specific productivities and returns to experience ψ_a. The third – 'output' – is the difference across labor markets related to different overall output Y_d. The fourth – 'skill-distribution' – is the difference in earnings due to differences in the supply of more educated workers L_{sd}. This influences the labor market general equilibrium effects that affect all cohorts. Last – 'skill-cohort distribution' – affects the earnings due to differences in the supply of skilled workers within each cohort ℓ_{asd}, and drives an additional GE effect on cohort a. Changes in the skill cohort distribution by age cohort will therefore have important GE effects on the earnings of workers.

Furthermore, how much the skill distribution affects the difference in earnings also depends on the elasticities of substitution σ_E and σ_A. For example, if the young and the old are perfect substitutes, then the skill-cohort distribution should not affect earnings.

The increase in earnings for a person who goes from being unskilled u to skilled s will be defined as the returns to education β_{asd}:

$$log\frac{w_{asd}}{w_{aud}} = log\frac{\theta_{sd}}{\theta_{ud}} + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right) log\frac{L_{sd}}{L_{ud}} - \frac{1}{\sigma_A} log\frac{\ell_{asd}}{\ell_{aud}} \equiv \beta_{asd} \tag{1.6}$$

These returns depend on the difference in the productivity parameters θ_{sd} and θ_{ud}, the skill distribution L_{sd} and L_{ud}, and the cohort specific skill distribution ℓ_{asd} and ℓ_{aud}. For instance, in regions that have relatively more skilled workers, the returns to acquiring skill will be relatively lower. Whereas for regions with more skill-biased capital, the returns to skill are higher.

with the $\frac{1}{\sigma_E}\log Y_d$ term, and not affect the returns to skill.

1.3.2 Students' Decisions

Students, in my model, choose the optimal level of education given their marginal costs of going to school and the returns to education (Becker, 1967; Mincer, 1958; Willis, 1986). Given how earnings are determined in section 1.3.1, these choices will also eventually affect earnings.

Individuals i in district d and age a choose their optimal consumption stream, C_{it}, and years of schooling, s_{id}, to maximize utility $u(C_{it})$, where $u'(C_{it}) > 0$ and $u''(C_{it}) < 0$. For a given subjective discount rate δ, an internal rate of interest r_{id} and a constant stream of earnings $w_{aid}(s_{id})$, the optimization problem can be set up as:

$$\max_{\mathbf{C_{it}},\mathbf{s_{id}}} \int_0^\infty u(C_{it})e^{-\delta t}dt \tag{1.7}$$

$$s.t. \int_0^\infty w_{aid}(s_{id})e^{-r_{id}(\kappa(s_{id})+t)}dt \geq \int_0^\infty C_{it}e^{-r_{id}t}dt \ , \tag{1.8}$$

where $\kappa(s_{id})$ captures the costs of schooling. For example, if $\kappa(s_{id}) = s_{id}$, then it only captures the opportunity cost of foregone earnings for each additional year of schooling.[29]

In the absence of incomplete markets and uncertainty, this problem is separable into individuals first choosing s_{id} to maximize their stream of earnings, and then choosing $\mathbf{C_{it}}$ to maximize utility. Student i chooses their optimal years of education s_{id} to maximize the present discounted value of their lifetime earnings $w_{aid}(s_{id})$ given the costs of going to school $\kappa_i(s_{id})$. Since the linear form of $\kappa(s_{id})$ only captures the opportunity costs, Card (1999) suggests a more general formulation of the cost function, which may also help capture credit and other monetary constraints (Becker, 1967):[30]

$$\max_{s_{id}} \ \log \ w_{aid}(s_{id}) - \left(\log \ r_{id} + r_{id}s_{id} + \frac{1}{2}\Gamma s_{id}^2\right) \ , \tag{1.10}$$

where Γ is the quadratic cost parameter. Equations (1.5) and (1.6) determine the form of the individual earnings function. The benefits from education for individual i can be represented by the following function, where β_{asd} captures the returns to schooling that may differ across districts, cohorts and skill-groups:

[29] This specific opportunity-cost only formulation leads to the familiar form (Mincer, 1958; Willis, 1986):

$$\log \ \left(\int_0^\infty w_{aid}(s_{id})e^{-r_{id}(s_{id}+t)}dt\right) = \log \ w_{aid}(s_{id}) - (\log \ r_{id} + r_{id}s_{id}) \tag{1.9}$$

[30] Becker (1967) justifies the quadratic costs from the observation that each subsequent year of education is even more expensive than before, because (a) fees are higher for higher levels (and in many cases early education is subsidized), and (b) students first exhaust easy sources of funds (parents, relatives) before using more expensive sources (loans).

$$\log\ w_{aid}(s_{id}) = \gamma_d + \nu_a + \beta_{asd}s_{id} + \log\ \epsilon_i\ , \tag{1.11}$$

where ϵ_i is the ability of the worker that is not observable to researchers. This ability will be correlated with the marginal costs of schooling r_{id} and lead to biases in standard OLS regressions ($corr(\epsilon_i, r_{id}) \neq 0$). For example, high-ability workers earn high wages but also have lower costs of performing in school. Also crucial to notice is that the returns to education β_{asd} differ across districts and skill-groups due to differences in relative skills in the local labor force, and across cohorts due to the cohort-specific differences in the skill distribution.

In Equation (1.11), average earnings also differ across districts γ_d due to differences in the overall output and capital across regions, and differ across age cohorts ν_a due to the returns to experience or other cohort-specific productivities discussed in Equation (1.5).

Given this setup, from the first order conditions one can obtain the optimal years of education for person i:

$$s^*_{id} = \frac{\beta_{asd} - r_{id}}{\Gamma} \tag{1.12}$$

The variation in s^*_{id} within a district d is driven entirely by the variation in the marginal cost parameter r_{id}. Notice, however, that the distribution of earnings in district d are driven both by the costs of education r_{id}, and by ϵ_i abilities.[31]

The marginal cost parameter for person i in district d is a function of the district-level costs of going to school, and an individual component η_i that captures individual heterogeneity in the costs of schooling. The district-level costs depend on the access to schooling A_d (like distance to the nearest school) and the monetary price of going to school p_d (like school fees).[32]

$$r_{id} \equiv -\Psi A_d + p_d + \eta_i\ , \tag{1.13}$$

where Ψ represents how aggregate access to education affects each i individual.[33] An increase in the number of schools in regions that did not initially have many will directly lower the transportation costs of going to school, but may also lower the competitive equilibrium fees, even for private schools.[34] These

[31]Smith (1775) highlights the importance of educational capabilities r_{id} when arguing that "The difference between the most dissimilar characters, between a philosopher and a common street porter, for example, seems to arise not so from nature, as from habit, custom and education." On the other hand, early formal models of variation in earnings (Roy, 1950) discuss the importance of 'abilities' ϵ_i, like "health, strength, skill, and so on."

[32]Restricting the cost parameter to simply depend on either only the monetary costs of going to school (p_d) or only the non-monetary costs (A_d) will not change the qualitative predictions of the model. This is because an expansion in public schooling will lower both types of costs in equilibrium.

[33]For example, one simple form is $\Psi = \frac{1}{N_d}$ where N_d is the number of potential students in the district.

[34]Here, students choose the lowest cost school regardless of whether they are privately or publicly owned.

education decisions are a nested portion of the problem where individuals maximize their lifetime utility based on their consumption stream.[35]

1.3.3 Schools

In this section I model the decisions made by private and public schools. The major modeling difference between public and private schools is that for public schools a district planner makes the decisions for all schools, whereas in the private sector each school decides separately. Furthermore, while public schools are meant to increase the access to schooling to citizens, private schools care about profits. Both types of schools can have heterogeneous costs or efficiency, but they provide the same output. Hence, students merely chose the school that is less costly for them, where the costs not only depend on the school fees p_d, but also transportation costs and non-monetary costs A_d.

1.3.3.1 District Level Public School Administrator's Decisions

District administrators for public schools maximize the access to schooling for students by investing in inputs. The total supply of public schooling depends on inputs such as newer schools, more teachers or better infrastructure. As access to schooling is increased, this reduces the marginal costs of going to school for students. For example, one crucial aspect of access to schooling could be the distance to the nearest school. By building more schools, public officials may reduce this distance and increase access to schools.

I model this in the following way: district administrators want to maximize the overall access to education A_d for the students in the entire district d. Greater access to schooling will lower the marginal costs for households. Public schools are not directly concerned with revenues from fees, and many are meant to be free (Kremer and Muralidharan, 2007). They do, however, have a budget constraint that restricts their spending.[36] The district d receives R_d from the government, and spends p_m for each input x_m into the schooling production function. Any funds received under government-backed schemes will increase the value of R_d. The vector of inputs at the district level $\mathbf{x_m}$ can consist of new schools, better qualified teachers, better infrastructure, more resource centers, etc. This setup, therefore, reduces the district's maximization problem to the following:[37]

$$\max_{\mathbf{x_m}} \; A_d(\mathbf{x_m}) \tag{1.14}$$

[35]On the consumption side, the inter-temporal consumption stream can be represented by the Euler equation: $\frac{u'(C_{i,t+1})}{u'(C_{i,t})} = \frac{\delta}{r_{id}}$, where overall consumption C_{dt}, must equal overall production by the firms Y_{dt} for the product market to clear.

[36]In equilibrium, the fees will be affected by the decisions of public and private schools, since they compete in the same market.

[37]The set-up is agnostic about heterogeneity in public schools – some may be more efficient than others.

$$s.t. \sum_{m=1}^{M} p_m x_m \leq R_d \,, \tag{1.15}$$

where $\frac{\partial A}{\partial x_m} > 0, \frac{\partial^2 A}{\partial x_m \partial x_m} < 0, \frac{\partial^2 A}{\partial x_m \partial x_n} > 0$. From the first order conditions, it is easy to derive the optimal amount of inputs of type m: $x^*_{md}(R_d, \mathbf{p_m})$, where $\frac{\partial x^*_m}{\partial R_d} \geq 0$ and $\frac{\partial x^*_m}{\partial p_m} \leq 0$. An increase in government funding R_d thus increases the amounts of each input in the schooling-access production function, increases the overall access to education A_d and reduces the marginal costs of schooling for the students in the district.[38]

1.3.3.2 Private schools

Private schools are profit maximizers and have heterogeneous costs.[39] An increase in the demand for schooling from households will increase their profitability, and more private schools will enter. Building more public schools affects the entry of private schools and determines the extent of crowd-in or crowd-out. If private schools are merely crowded out one-for-one, then the funds may have been better spent elsewhere.

Private schools are price takers in the competitive market and charge a fee p_d. Muralidharan and Sundararaman (2015) are among the first to use causal evidence to show that students in private schools have the same test scores as public school students for the subjects that are taught in school. They are, however, more cost-effective. Private schools, in my model, therefore, have the same output as public schools, but may do so at a different cost. Furthermore, there is heterogeneity in private school efficiency (Kremer and Muralidharan, 2007).

The total educational output (in student-years) Q_{jd} by school j is a function of its aggregate inputs X_{jd} in the following way: $Q_{jd} = \overline{\theta_d} X_j$. Here the aggregate output of the private schools depends on the average skill level of the district $\overline{\theta_d}$. This is meant to capture demand externalities. For instance, Birdsall (1982) models the demand for schooling from households, as a function of the aggregate supply of public schools. An expansion of public schooling, will then increase the overall demand for all schools, including privately owned ones. Another alternative comes from peer effects in school participation. If certain students are encouraged to go to school, then the demand from their neighbors may also rise (Bobonis and Finan, 2009).[40] There are, however, quadratic costs associated with using inputs $Z(X_j)$. The school chooses inputs to maximize profits:

$$\max_{X_j} \;\; p_d \overline{\theta_d} X_j - Z(X_j) \tag{1.16}$$

[38] See Appendix 1.12.1 for a parametric example of this set-up.

[39] Alternatively, they could have been modeled as having heterogeneous productivities, with the same result.

[40] Output in the public schools may also depend on $\overline{\theta_d}$, without a qualitative change in the results.

The costs $Z(X_j) = z_{1j}X_j + \frac{1}{2}z_{2d}X_j^2$ have a simple quadratic formulation.[41] There is a heterogeneity in costs z_{1j} across schools, where some schools use their inputs more effectively than others, and a heterogeneity in costs z_{2d} across districts, where certain districts have better infrastructure for setting up a school. This is meant to capture the fact that in some districts it is cheaper to hire teachers (Andrabi et al., 2013) and some have better physical infrastructure like electricity, drinking water, roads, and Block Resource Centers (BRCs) than others (Jagnani and Khanna, 2016). The demand for inputs can, therefore, be found from the first order conditions, and the supply curve of schooling is:

$$Q_{jd} = \overline{\theta_d}X_j^* = \overline{\theta_d}\frac{p_d\overline{\theta_d} - z_{1j}}{z_{2d}}\,, \tag{1.17}$$

whereas, the school's profits are:

$$\pi_{jd} = \frac{(p_d\overline{\theta_d} - z_{1j})^2}{2z_{2d}}\,, \tag{1.18}$$

Since there is free entry of private schools into these regions, schools will enter until $\pi_{jd} = 0$. The marginal school, therefore, will have a cost parameter $\widetilde{z_{1d}} = \overline{\theta_d}p_d$. If costs are drawn from a distribution $F(z_{1j})$, then the fraction of schools that enter is given by: $F\left(\overline{\theta_d}p_d\right)$.

Notice what guides the entry and exit decision of schools is the average productivity level in the district $\overline{\theta_d}$, the price p_d, and consequently the cost z_{2d} which depends on the infrastructure levels. If we see a fall in the supply of private schools along with a fall in the equilibrium price, then it is clear that the strongest driving force is that an increase in the supply of public schooling drives down the equilibrium price and crowds-out private school.

Alternatively, if we see a rise in the supply of private schools in the light of an expansion in public schools, there are two possible reasons. The first is that demand externalities and peer effects – captured by θ_d – drive up the equilibrium price and induces private schools to enter. The second is that infrastructure upgrades and the presence of more teachers lowers the operating costs – captured by z_{2d} – and lead to more private school entry and further lowers the equilibrium price. The best evidence for how private schools respond comes from Andrabi et al. (2013), who show that how an expansion in public schooling increased education for girls, and these girls became teachers in Pakistani districts. This increase in the number of teachers allowed private schools to enter the market soon after. Similarly, Jagnani and Khanna (2016); Pal (2010) find that physical infrastructure upgrades can induce private-school entry in India.[42]

[41] While it is easy to hire the first few teachers or administrators, it is more costly to hire the next few as the pool of potential candidates dwindles.

[42] See Appendix 1.12.1 for a derivation for the overall supply of private schooling in the region – in the parametric formulation, it is easy to see that $\frac{\partial p_d^*}{\partial z_{2d}} > 0$. The intuition is simple a fall in the operating costs will increase the supply of private schools and lower the equilibrium price. On the other hand, an increase in the demand externality will increase demand for schooling and raise the equilibrium price. Seeing how prices change will allow us to distinguish between the various potential

Last, Kremer and Muralidharan (2007) argue that private schools may exist in regions where public schooling has failed, suggesting that they may end up serving different sub-regions within a district. Public-school officials may then target regions where there are fewer private schools in the first place, and dissipate any concerns of crowd-out or any likelihood of crowd-in. Together, these predictions on equilibrium quantities and prices will allow us to determine which of these mechanisms are stronger than others.

1.3.4 Definition of an Equilibrium

The exogenous elements of equilibrium are the student utility, schooling-cost functions, educational access functions, private firms' production functions, and the amount of exogenous government spending on schooling. What is endogenous is the years of education, the earnings-returns to education, the optimal inputs in the schools, the output of firms, the fraction of private schools that enter, and the equilibrium price and quantity of schooling.[43]

Appendix 1.12.1 characterizes and derives the education-sector equilibrium. For the product market to clear, the amount of consumption C_{td} must equal the amount of output Y_{td}. For the labor market to clear, the demand for workers ℓ_{asd} with education level s must equal the supply. The equilibrium amount of schooling determines the supply of labor at skill level s, and the demand for labor is determined by Equation (1.4).

Proposition 1 (Equilibrium) *Given the following dimensions of the model: A student utility function $U(C)$, returns to education function $\log\ w(s)$ and cost functions $\kappa(s, r, \Gamma)$; access to schooling function $A(\mathbf{x_m})$, and prices of inputs p_m; exogenous revenues from the government R_d; distribution of private school costs $F(z_{1j})$, and cost functions for private schools $Z(X_j)$; firms' production functions Y, different productivities for each education level θ_{sd}, the elasticity of substitution between education groups σ_E, and age groups σ_A; there exists an equilibrium that determines: The returns to an additional year of schooling β_{asd} that varies by district, age-cohort and skill level; the distribution of the optimal years of schooling S^*_d, and the monetary price of going to school p^*_d; the optimal inputs into the access function $x^*_m(R_d, \alpha_{\mathbf{m}}, \mathbf{p_m})$; optimal private school inputs $X^*_j(p_d, z_{1j})$; and equilibrium earnings w_{asd} and quantities of each type of worker ℓ_{asd}.*

mechanisms.

[43] In line with the literature, so far I have assumed perfect foresight. When there are general equilibrium effects, students know exactly what the earnings will be *including* the general equilibrium effects. If expectations were adaptive, and students did not take into account the GE effects, they would get "too much" education, causing the skilled wage to fall even further. The subsequent cohort would then need to adjust its expectations as well, and the equilibrium is approached very slowly as each cohort revises its expectations. For a cobweb style model see Freeman (1975).

1.4 Identification, Policy Outcomes and Economic Benefits

I use this model to generate sufficient statistics which determine the overall economic consequences of the policy, including the general equilibrium effects. To estimate these sufficient statistics, I use a RD based on a policy which expanded the public-school system. Improving access to schooling by building newer schools or upgrading its infrastructure will reduce the marginal costs of schooling and induce some students to get more education. Since the policy was large, the GE effects will affect the benefits to not just the induced students, but all others in the local economy as well. In order to determine the labor market earnings of all the different types of workers, I use the policy variation to isolate the GE effects.

1.4.1 Using Policy Changes to Estimate Parameters

The variation in s^*_{id} is driven entirely by the variation in the marginal costs r_{id}. Since the costs of schooling are likely to be correlated with the ability of the worker $Cov(\eta_i, \epsilon_i) \neq 0$, a simple OLS regression of earnings on education will give us biased estimates of the parameters. Moreover, comparisons in the cross-section across different labor markets will provide biased estimates due to underlying baseline differences in the skill distribution across these markets.

The equilibrium amount of schooling is affected by the expansion of public schooling:[44]

$$S^*_d = \phi_1 \beta_{asd} + \phi_2 R_d - \frac{\eta_d}{\Gamma} \tag{1.19}$$

There are a few crucial components to this equation – the $\phi_2 R_d$ portion captures how more government spending increases equilibrium schooling by making public schools more accessible, and making (via adjustments in the market price) private schools more affordable (Appendix 1.12.1). The term $\phi_1 \beta_{asd}$, captures how changes in the returns to education will affect equilibrium schooling. If, for example, the labor-market general equilibrium effects substantially lower the returns to education β_{asd}, then there may be no detectable increase in the equilibrium amount of schooling. The final term $\frac{\eta_d}{\Gamma}$ is unaffected by the schooling expansion. We would, however, expect it to be correlated with other district-level characteristics causing biased estimates in standard estimation frameworks.

The districts that received the schooling expansion policy were selected based on a criteria that leads to a fuzzy Regression Discontinuity (RD) design. Any district that had a female literacy rate below the national average (based on the 1991 Census) was made eligible to receive the policy. Therefore, it is

[44] See Appendix 1.12.1 for a parameterization of ϕ_1 and ϕ_2, where $\phi_1 \equiv \left(\frac{\overline{\theta^2_d}}{\Gamma\overline{\theta^2_d}+z_{2d}}\right)$ and $\phi_2 \equiv \left(\frac{\left(z_{2d}+\Psi\overline{\theta^2_d}\right)\left(\prod_m \frac{\alpha_m}{p_m}{}^{\alpha_m}\right)}{\Gamma\overline{\theta^2_d}+z_{2d}}\right)$, and $\eta_d = \mathbb{E}[\eta_i | i \in d]$..

possible to compare districts just above and below this cutoff to determine the causal impact of the policy on equilibrium schooling. Furthermore, we should expect that η_d is not different for districts that just fall on either side of the cutoff.

Let us define $D_d = 1$ to be districts that just fall on the side of the cutoff that receives the policy, and $D_d = 0$ districts that fall on the other side. In the neighborhood of the cutoff, we should therefore expect:

$$S_d = \phi D_d + \frac{\eta_d}{\Gamma} \quad and \quad \mathbb{E}[\eta_d | D_d = 1] = \mathbb{E}[\eta_d | D_d = 0] \tag{1.20}$$

If the direct effects of increasing access to schooling outweigh any negative labor market general equilibrium (GE) effects that depress returns, then we should expect $\phi > 0$.

1.4.1.1 Returns to Education and Disentangling Earnings

The policy changes the distribution of earnings across the RD cutoff. In this section, I disentangle the labor market GE effects of the policy with the help of Equation (1.4). In Equation (1.4), ψ_a captures the cohort effect. θ_{sd} captures the pure productivity effect and a change in the amount of skill-biased capital in response to the policy will change its value. The term $\frac{1}{\sigma_A} \log \ell_{asd}$ is crucial for the cohort specific labor-market general equilibrium effect, and $\frac{1}{\sigma_E} \log Y_d + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right) \log L_{sd}$ determines the general equilibrium effect that affects all cohorts:[45]

$$\log w_{asd} = \log \widetilde{\varrho} + \log \theta_{sd} + \log \psi_a + \frac{1}{\sigma_E} \log Y_d + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right) \log L_{sd} - \frac{1}{\sigma_A} \log \ell_{asd} \tag{1.4}$$

I exploit variation along various dimensions to disentangle the components of the change in earnings across the RD cutoff. These dimensions include age cohorts, skill levels and treatment status. By restricting comparisons to be within cohorts, the cohort effect on earnings Ψ_a is differenced out. Cohorts that were too old to change their years of education when the policy was implemented will be affected by the labor-market general equilibrium effects if they work in the treated districts. The general equilibrium effects that affect all cohorts can thus be isolated by looking at the impact on the older cohort. If we restrict comparisons to people of the same skill level across the cutoff, one can difference out the productivity differences not related to skill biased capital and thus back out the cohort-specific general equilibrium effect. This process allows me to completely identify all the components that affect earnings.

Since older cohorts participate in the same labor market, they are affected by the general equilibrium

[45] While this equation is represented in terms of production function parameters, the estimated GE effects will not depend on the specific functional form of the production function as long as workers can be disaggregated into skilled and unskilled; young and old.

effects that affect all cohorts. There is thus a difference in the skill premium – the difference between the skilled and unskilled wage – that exists among the older cohorts across the RD cutoff. By looking at the earnings of both the skilled and unskilled workers in older cohorts separately, I can back out the GE effects that affect all cohorts. Earnings for younger cohorts, however, will additionally be affected by cohort-specific general equilibrium effects since there are more highly educated people in the younger cohorts.[46]

For ease of exposition I restrict the analysis to two skill levels – skilled s and unskilled u workers. For example, the fraction of each among the young y are represented by ℓ_{sy} and ℓ_{uy} respectively. For any two-skill groups: $\Delta\ell_{sy} \equiv (\ell_{sy,D=1} - \ell_{sy,D=0}) = -\Delta\ell_{uy} \equiv (\ell_{uy,D=1} - \ell_{uy,D=0})$.

Let $D = 0$ represent the local economies that do not receive the program, and $D = 1$ the districts that do. If only a single individual was to acquire skill and change status from unskilled u to skilled s, there would be no general equilibrium effects. If the person lives in the untreated region $D = 0$, then that person's earnings would change in the following manner:

$$log\frac{w_{as,D=0}}{w_{au,D=0}} = \underbrace{log\frac{\theta_{s,D=0}}{\theta_{u,D=0}}}_{\text{Productivity}} + \underbrace{\left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right) log\frac{L_{s,D=0}}{L_{u,D=0}}}_{\text{Aggregate skill distribution}} \underbrace{-\frac{1}{\sigma_A} log\frac{\ell_{as,D=0}}{\ell_{au,D=0}}}_{\text{Cohort specific skill distribution}} \equiv \beta_{as,D=0} \ , \tag{1.21}$$

where $\beta_{as,D=0}$ is defined as the earnings returns to changing ones skill from u to s in a district where $D = 0$. If however, the individual lived in a treated region $D = 1$, where there are a lot more educated people or a lot more skill-biased capital because of the policy, the change in earnings would be:

$$log\frac{w_{as,D=1}}{w_{au,D=1}} = log\frac{\theta_{s,D=1}}{\theta_{u,D=1}} + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right) log\frac{L_{s,D=1}}{L_{u,D=1}} - \frac{1}{\sigma_A} log\frac{\ell_{as,D=1}}{\ell_{au,D=1}} \equiv \beta_{as,D=1} \ , \tag{1.22}$$

where $\beta_{as,D=1}$ is defined as the earnings returns to changing ones skill from u to s in treated regions $D = 1$. These returns differ because of the differences in the skill distribution of the workforce and the amount of skill-biased capital across regions. The difference in the returns to acquiring skill between these two regions is $\Delta\beta_{as} \equiv \beta_{as,D=1} - \beta_{as,D=0}$. Across the RD cutoff these returns will be different because of a change in the skill composition of the workforce and the flow of skill biased capital. These are the GE effects on the returns to education:

[46]In the estimation exercise, there are a few rules that need to be followed in order to get unbiased estimates. First, wage comparisons must always be made across the RD cutoff. Second, the same cohorts must be compared across the cutoff, and last the same skill group must be compared across the cutoff.

$$\Delta\beta_{as} = \underbrace{\left(log\frac{\theta_{s,D=1}}{\theta_{u,D=1}} - log\frac{\theta_{s,D=0}}{\theta_{u,D=0}}\right) + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right)\left[log\frac{L_{s,D=1}}{L_{u,D=1}} - log\frac{L_{s,D=0}}{L_{u,D=0}}\right]}_{\text{GE effects on all cohorts}}$$
$$\underbrace{-\frac{1}{\sigma_A}\left[log\frac{\ell_{as,D=1}}{\ell_{au,D=1}} - log\frac{\ell_{as,D=0}}{\ell_{au,D=0}}\right]}_{\text{Additional GE on young}} \quad (1.23)$$

Since the policy also changed inputs within schools, we may expect there to be an increase in the quality of schooling. While some quality improvements lower the marginal cost of schooling, others may increase the returns in the labor market. Such improvements would therefore attenuate the change in the returns to education.

In order to disentangle the general equilibrium effects on each cohort, one can look at the discontinuity in the skill premium of the younger and older cohorts separately. By restricting the population to a specific skill level (and cohort) one can ensure that the differences in earnings across the RD cutoff are only due to differences in the skill distribution and the amount of skill-biased capital.

The change in returns in Equation (1.23) can be split up into two components. The first is the GE effect that affects all cohorts. To estimate this effect, I look at the change in the skill premium for the older cohort o. Empirically, this is the earnings differential between the skilled older population and the unskilled older populations:[47]

$$\underbrace{log\frac{w_{so,D=1}}{w_{so,D=0}} - log\frac{w_{uo,D=1}}{w_{uo,D=0}}}_{\text{GE effects on all cohorts}} = \underbrace{\left(log\frac{\theta_{s,D=1}}{\theta_{u,D=1}} - log\frac{\theta_{s,D=0}}{\theta_{u,D=0}}\right)}_{\text{Skill biased capital}} + \underbrace{\left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right)\left[log\frac{L_{s,D=1}}{L_{u,D=1}} - log\frac{L_{s,D=0}}{L_{u,D=0}}\right]}_{\text{Aggregate skill distribution}} \quad (1.24)$$

Notice that we would expect that these two portions of the GE effects on all cohorts counteract each other. On the one hand, an increase in the skilled workforce will attract more skill-biased capital and raise the skill premium. On the other hand, increasing the relative supply of skilled workers makes them less valuable in production, lowering the skill premium. If there is differential migration, and skilled workers migrate out of the treated districts in search of work, then it will weaken the strength of the 'Aggregate skill distribution' component of the GE effects.

The second component of the GE effects is the additional GE effect on the young y that is driven solely by the change in the age-specific skill distribution. This component can be measured by estimating the

[47]The change in the skill premium for older cohorts will be the GE effects on all cohorts regardless of the specific formulation of the production function. Similarly, all estimates of the returns and GE effects do not rely on the production function.

earnings differential between the skilled young and unskilled young, and differencing out the earnings differential between the skilled old and unskilled old:[48]

$$\underbrace{\left[log\frac{w_{sy,D=1}}{w_{sy,D=0}} - log\frac{w_{uy,D=1}}{w_{uy,D=0}}\right] - \left[log\frac{w_{so,D=1}}{w_{so,D=0}} - log\frac{w_{uo,D=1}}{w_{uo,D=0}}\right]}_{\text{Additional GE on young}} = \underbrace{-\frac{1}{\sigma_A}\left[log\frac{\ell_{ys,D=1}}{\ell_{yu,D=1}} - log\frac{\ell_{ys,D=0}}{\ell_{yu,D=0}}\right]}_{\text{Age specific skill distribution}} \tag{1.25}$$

To estimate the two different returns $\beta_{as,D=0}$ and $\beta_{as,D=1}$, I use discontinuities in the average earnings of the young, and the wages of the skilled young, and unskilled young separately:[49]

$$log\frac{w_{y,D=1}}{w_{y,D=0}} = \ell_{sy,D=1}log\frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=1}log\frac{w_{uy,D=1}}{w_{uy,D=0}} + \Delta\ell_{sy}\underbrace{log\frac{w_{sy,D=0}}{w_{uy,D=0}}}_{\beta_{ys,D=0}} \tag{1.26}$$

$$log\frac{w_{y,D=1}}{w_{y,D=0}} = \ell_{sy,D=0}log\frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=0}log\frac{w_{uy,D=1}}{w_{uy,D=0}} + \Delta\ell_{sy}\underbrace{log\frac{w_{sy,D=1}}{w_{uy,D=1}}}_{\beta_{ys,D=1}} \tag{1.27}$$

The change in the average earnings for the younger cohorts is a weighted average of how the skilled and unskilled wages change, and the shift from unskilled to skilled work multiplied by the returns to skill. These relationships can be used to derive the returns to skill in both the treated and untreated districts separately. At the same time, the average years of education in the districts change across the cutoff in the following manner:

$$\begin{aligned}\Delta S &= (\ell_{sy,D=1}s_1 + \ell_{uy,D=1}s_0) - (\ell_{sy,D=0}s_1 + \ell_{sy,D=0}s_0) \\ &= \Delta\ell_{sy}s_1 + \Delta\ell_{uy}s_0 = \Delta\ell_{sy}(s_1 - s_0)\ ,\end{aligned} \tag{1.28}$$

where s_0 is the years of education for the skilled group, and s_1 are the years for the unskilled group, and $\Delta\ell_{sy}$ is the fraction of students induced into getting more skill.

It is important to remember that the shift in the skill-distribution will change overall output as well. If an individual that has a skill level s were to switch districts from $D = 0$ to $D = 1$, that person's earnings

[48] Notice that if $\sigma_A < \sigma_E$ then the two components may be of opposite signs.
[49] See Appendix 1.12.4 for detailed derivations of these equations.

would be different not only because of the skill distribution, but also because of the differences in overall output Y_d and skill-biased capital across the regions:[50]

$$log\frac{w_{s,D=1}}{w_{s,D=0}} = \frac{1}{\sigma_E}log\frac{Y_{D=1}}{Y_{D=0}} + log\frac{\theta_{s,D=1}}{\theta_{s,D=0}} + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right)log\frac{L_{s,D=1}}{L_{s,D=0}} - \frac{1}{\sigma_A}log\frac{\ell_{as,D=1}}{\ell_{as,D=0}} \tag{1.29}$$

1.4.2 Outcomes and Economic Benefits

The model predicts that when the district receives more funds for expenditure on public schooling, the following happens: First, public administrators build more schools, increasing the access to schooling (Section 1.3.3.1). This lowers the marginal cost of schooling for households, and induces certain students to get more education (Section 1.3.2). At the same time, private schools decide whether to enter or exit the education sector, leading to either a crowd-in or crowd-out of private schools (Section 1.3.3.2). When the newly skilled workforce joins the labor market they earn the higher skilled wage (Section 1.3.1). There is, however, a distributional impact on the earnings of skilled and unskilled workers (Section 1.4.1). If skilled workers are more productive and attract more skill biased capital, then there is an increase in overall output, productivity and consumption (Section 1.3.1).

The changes in overall benefits will depend on the reduction in the costs of schooling for younger cohorts, the increase in overall output due to the acquisition of skills by workers, and the labor market returns. The increase in total output depends on the productivity parameters and the change is the skill distribution. At the same time, the general equilibrium effects will have distributional consequences.[51]

The welfare for a young high-skill person that would have acquired skill even in the absence of the policy rises by the reduction in the total costs of education, but is dampened by the GE effect that affects all cohorts and the additional GE effect on younger cohorts. The welfare for older cohorts is unaffected by the reduction in the costs of schooling. The skilled old however are adversely affect by the GE effects that affect all skilled workers, whereas the unskilled old benefit from the increase in their earnings. All these groups additionally benefit from any overall increases in output, and all young workers that get education benefit directly from a reduction in the total costs of schooling.

Let $\beta_{as,D=0}$ be the returns to education in untreated districts, and $\beta_{as,D=1}$ be the returns including the general equilibrium effects. $\Delta\beta_{as}$ is thus the change in the returns due to the GE effects. The labor-market welfare change for those who were always going to acquire skill even in the absence of the policy

[50] If the capital that flows in to these regions is skill-biased, then we just need to interpret the GE effects as including not just the change in the aggregate skill distribution but also an inflow of skill-biased capital. See Appendix 1.12.2. If aggregate output prices change, then the skill-premium is unaffected since both the skilled and unskilled within the same district face the same price change.

[51] Aggregate profits for private schools has a closed form solution and will change due to the policy. The extent of this will depend on the increase in productivity $\bar{\theta}_d$ and the decrease in the equilibrium price of schooling p_d). Furthermore, the returns to capital and land may change as well, depending on the ease of mobility and transaction costs. However, my analysis only concentrates on the earnings of workers.

is just the difference in the skilled wage at the cutoff: $log\frac{w_{as,D=1}}{w_{as,D=0}}$. Similarly, for those workers who would never acquire more skill even in the presence of the policy, the difference in the unskilled wage at the cutoff captures their labor market welfare: $log\frac{w_{au,D=1}}{w_{au,D=0}}$. For the younger cohorts, who are induced into getting more skill, the labor market welfare change depends on the skilled wage in the treated districts and the unskilled wage in the untreated districts: $log\frac{w_{ys,D=1}}{w_{yu,D=0}}$. To estimate this component of welfare, I use the returns to schooling $\beta_{ys,D=0}$, since:

$$log\frac{w_{ys,D=1}}{w_{yu,D=0}} = \quad log\frac{w_{ys,D=0}}{w_{yu,D=0}} + log\frac{w_{ys,D=1}}{w_{ys,D=0}} = \quad \beta_{ys,D=0} + log\frac{w_{ys,D=1}}{w_{ys,D=0}} \tag{1.30}$$

The change in labor market benefits for those induced into getting more skill therefore consists of two components – the returns to skill in the untreated districts and the change in economic benefits to the 'always skilled.' In the absence of any GE effects, the change in earnings for a person induced into getting more education would simply be $\beta_{ys,D=0}$. Lastly, let δ be the discount rate, and τ be the years between when the labor market returns are realized and the costs of schooling are borne. To compare the labor market benefits to the reduction in total costs of schooling to get a measure of aggregate welfare change, I discount the labor market gains by the real interest rate δ, over the time period τ. For a student induced into getting more education, the costs include tuition costs and the opportunity cost of a foregone unskilled wage.[52] The benefits, however, include the present discounted value of a skilled worker's earnings stream.

1.5 Data

In order to study the policy in a comprehensive manner, I put together a number of large datasets that have not been combined in this manner before. The data are merged at the district level since districts are the relevant local economy and labor market in this context (Duflo and Pande, 2007). I combine data on school-level inputs, household level data on years of education, migration decisions and amount spent on schooling, labor market data on earnings and occupations, and firm-level data on types of manufacturing in the different regions. Doing so allows me to study how the policy affects the entire district rather than just individual households.

Furthermore, it is important to study the dynamic consequences of the policy. For this purpose, I assemble a yearly panel data set that allows me to track schools, firms and other characteristics of the local economy over time. Given the changes in district boundaries over time, this panel is particularly challenging to create. Details of all the data sets can be found in Data Appendix 1.13.

Data for inputs into schools comes from the District Information System for Education (DISE). Table 1.1 summarizes the variables of interest in the year 2005, twelve years after the DPEP started. The top panel

[52]Empirically, I will not measure the non-monetary costs of schooling.

classifies schools by ownership (government vs private), and when they were built (before 1993, the first year of the earliest program or after). 27% of all schools existing in 2005 were built post 1993, and while 20% are government schools, the remaining 7% are private schools. There is large variation in the level of infrastructure - while 84% of the schools have drinking water, only 31% have constant electric supply. DISE also collects information on Block and Cluster Resource Centers, which are often used for training teachers, and have computerized facilities, which are used to access other teaching materials. On average, schools are about 13 kms away from the closest Block Resource Center (BRC), and are visited by a BRC official about 1.5 times a year. The data also has information on various sources of funding, one of which is the Teacher's Learning Material (TLM) grant that all schools are eligible for irrespective of other programs. On average, schools got about Rs. 1517 ($38) in total from the TLM grant every year.

In order to study educational outcomes, household surveys and Census data were used. The Census has detailed tables at all three of the administrative levels - states, districts and sub-districts. A panel of districts can be created using the 1991, 2001 and 2011 Census years, all of which include district-level statistics. The 1991 Census determines the running variable for the RD, since the 1991 female literacy rate was used to determine which districts are eligible for DPEP funds.

I use three different rounds of a household survey to study the impact on education, earnings, expenditures, migration and other labor-market characteristics. The National Sample Survey (NSS) is a nationally representative survey used by many researchers studying India. It is the largest household survey in the country, asks questions on weekly activities for up to five different occupations per person, and records earnings during the week for each individual in the household. Summary statistics for the 2009 NSS round are presented in Table 1.2. In 2009, only about 60% of the population had finished primary school, and on average people had about 6 years of education and earned about Rs 1466 ($30) a month.

Last, to study the behavior of firms, I use the Annual Survey of Industries (ASI), which is a census of all manufacturing firms in the country that employ more than ten persons. These data are available at the establishment level, and have information on the type of products produced, wages paid, and number of employees. One can then use these data to study whether changes in the skill level of the population can affect firm mobility and production decisions.

1.6 Estimation

In order to target the DPEP program to districts that were worst off in terms of educational outcomes, a selection criterion was used. Districts that had a female literacy below the national average (based on the previous 1991 Census) were eligible for the program, but not all such districts were selected. Further, in some states that had no low-literacy districts, a few districts were selected at the state's discretion. This imperfect assignment requires a fuzzy regression discontinuity design using the 1991 female literacy rate as a running variable. The fuzzy design allows for imperfect assignment, since not all low-literacy districts were selected, and for states with no low-literacy districts, some high-literacy districts were selected. To my knowledge, there are no other programs that use the district-level 1991 female-literacy rate as cutoff. I empirically test and show that cohorts that were too old to change their schooling decisions by the time the policy was implemented have no discontinuity in educational attainment.

The RD allows me to compare districts just above the literacy cutoff to those just below. Since we should not expect any discontinuity in the distribution of individual labor-market abilities or individual-specific costs of going to school around the cutoff at baseline, the RD estimator is consistent. Furthermore, at the cutoff, we should expect no discontinuity in pre-policy labor market characteristics and regional outputs that would otherwise bias the estimated parameters. In order to estimate the GE effects, I further exploit variation in cohort exposure and skill levels.

OLS estimates of the returns to education suffer from a few sources of concern. Since more able workers are also more capable students, OLS estimates can suffer from an omitted variable bias. The variation generated by the policy can be used to overcome this bias. The policy induces certain students to go to school, whereas identical students in non-policy regions do not. Following students into the labor market, it is therefore possible to compare wages in the two regions to determine the returns to schooling for the subpopulation that was induced into getting more education. At the same time, local labor markets may differ widely across regions in terms of their skill distributions and skill premiums. This will confound OLS estimates of the GE effects. The RD allows me to compare similar local economies that differ only on the access to the DPEP policy.[53]

The first stage is presented in Figure 1.3. It is clear that the more literate amongst the eligible districts (i.e. amongst the districts with lower than average female literacy) were selected for the program, leading to a discontinuity at the cutoff. There is also visible imperfect assignment at both ends, with not all eligible districts being selected, and not all selected districts being eligible. Since it is clear that policy makers selected the most literate of the low-literacy districts, there is a high likelihood of political manipulation that is correlated with a whole host of unobserved characteristics in these regions. Given such a set-up, regression specifications that do not allow for these differences in unobserved characteristics will be biased. A RD specification can, therefore, provide a causal estimate of the impact of this program. This will be the Local Average Treatment Effect (LATE) for districts near the cutoff value (Imbens and Angrist, 1994). The parameters estimated should be thought of as being representative only for districts near the RD cutoff. Furthermore, as will be discussed below, parameters like the estimated returns to education are for the students who were induced into getting more schooling and lived in the districts near the cutoff. The GE effects as well depend on what type of students get induced into getting more

[53] One additional concern with OLS estimates is the measurement error in the independent variable (Griliches, 1977). This can especially be an issue in developing countries where the levels of education may not be clearly defined. The RD helps tackle the attenuation bias that arises from classical measurement error.

skill, as this may affect the amount of skill biased capital flowing in and the change in the effective supply of labor. These general equilibrium effects, however, will also affect sub-populations that were not induced into getting more education.

Estimating causal impacts requires that there is no perfect manipulation of the running variable or the cutoff, which is likely in this case since the cutoff chosen was the national average of the female literacy rate from the previous 1991 Census. Furthermore, the McCrary (2008) tests indicate that the density of districts and of households around the cutoff is not discontinuous (Figure 1.4), since the p-value of the change in density at the cutoff is 0.71. Other falsification tests will be discussed below that solidify the RD assumptions that there were no other discontinuities at the same cutoff.[54]

While RD results will be represented graphically, the coefficients of interest will also be calculated using a two-stage least squares procedure where the optimal bandwidth will be calculated using two different methods. The optimal bandwidths determine what fraction of the sample is used from the regions near the cutoff to estimate the effect of the policy in the neighborhood of the cutoff. I use the Calonico et al. (2014b) method, and the Imbens and Kalyanaraman (2012) method to select these bandwidths. The Imbens and Kalyanaraman (2012) method uses a data-driven bandwidth selection algorithm to identify the optimal bandwidth for a local linear regression given a squared loss function, whereas the Calonico et al. (2014b) method also performs a bias-correction and develops robust standard errors for such a procedure.[55] Results using both the optimal bandwidth procedures are presented, and are robust to using more parametric approaches like local linear and quadratic control-function approaches as suggested by Hahn et al. (2001) and Imbens and Lemieux (2008).[56]

1.7 Results and Discussion

The household level analysis can also be split up by the age groups that should and should not have been directly affected by the DPEP program. In Appendix Figure 2.8.5 one can see a sharp drop in schooling enrollment at the age of 19, because by that age students have usually finished schooling. Since the household survey was conducted 16 years after the start of the program, anybody above the age of 35 should not be directly affected by the program. Those under the age of 35 in treated districts, however, should be directly affected.

I present RD figures showing the discontinuity along the running variable for the year 2005, which was

[54]Cattaneo et al. (2015) offers an alternative test for manipulation at the cutoff that does not rely on the selection of binning parameters. The p-value of a discontinuity in the density using their method is 0.97, indicating almost no likelihood of discontinuity in densities.

[55]I use the code written by Calonico et al. (2014a) to estimate the parameters.

[56]The results are robust to using various alternative procedures that are as yet unpublished (Appendix Table 1.15). The first, described in Bartalotti and Brummet (2015) allows for computing the standard errors at an aggregated level. I thank Quentin Brummet for sharing his code. The second method allows for different optimal bandwidths on either side of the cutoff and for standard errors at an aggregated level.

the last year before the phase-out of funds begin. The 2005 figures can be thought of as capturing the cumulative effects of the last twelve years of funding increases. Alongside the RD figures, the 2SLS coefficient over time will be presented. How the 2SLS coefficient changes over time shows what happens to the outcomes of interest once the funding is cut in 2006, and stopped entirely in 2007. The RD figures that will be shown are Intent-to-Treat (ITT) estimates, so to scale up the effects by the probability of treatment, the two-staged least squares (2SLS) coefficient will be shown alongside. These 2SLS coefficients and standard errors are calculated using the Calonico et al. (2014b) optimal bandwidth selection procedure and bias-correction methods.

1.7.1 School Building and Survival

One of the primary objectives of the DPEP program was to build new schools. New schools should increase access to schooling in areas that did not have schools. This will lower the marginal cost of schooling for students who no longer need to travel as far for a school. At the same time, we may expect private schools to respond to this governmental schooling expansion as well.

I present RD figures for the number of schools and school-level inputs using the DISE data. Figure 1.5 shows the effect of the program on schools built once the program was underway in 1994. While the top panel shows the fraction of all schools that were built post 1993, the middle panel shows the discontinuity in government schools. Both panels show that DPEP districts had a substantially larger fraction of new schools than non-DPEP districts. The Intent-to-treat (ITT) estimates indicate a 4.9 percentage point increase in the fraction of government schools that were new.

The size of the discontinuity and trend over time are robust to the choice of the bandwidth. In Appendix Figure 1.13 I show alternative versions where I plot the total schools per capita. The lower panels show the RD coefficients for different types of optimal bandwidths (CCT - Calonico et al. (2014b), I&K - Imbens and Kalyanaraman (2012)), and by also restricting the bandwidths to be the same as in the first year of the data.[57]

Studying how the coefficient in Figure 1.5 changes over time allows me to trace out the longer term effects of the program. In all of the figures, the first coefficient plotted for the year 2005 shows a large discontinuity in the fraction of new schools, whereas the other coefficients in later years show a smaller difference among the districts on either side of the cutoff. While large sources of funding were still being received by these districts in 2005, more schools were being built. However, once the funding was whittled down there was no longer any differential impact on the fraction of new schools built. This is because in the absence of funds, regions on the other side of the cutoff catch up by building schools at an even more rapid rate.

[57] This was a rapid period of school building, affecting the sample size from year to year. As the sample grows, the bandwidths get smaller, so restricting to the same bandwidth as the first year, approximately checks the results for a balanced panel of districts.

As a falsification test, I can also look for any differential impacts on the fraction of schools that were built in the twenty year period prior to the program (1973 to 1993) out of all schools built before 1993. If schools have a short lifespan, then more funding may allow schools to last for longer. However, this was a time when more schools were being opened rather than old schools being shut down. Therefore, we should not expect any discontinuities in the number of older schools, and they would serve as a clean falsification test in this context. Appendix Figure 1.14 shows the lack of a discontinuity in older schools, both for government and private schools.

1.7.2 Private Schooling Response

How private schools respond to such interventions, however, is also crucial for determining the overall benefits. Private schools may be crowded out of the local education sector with the entry of new government schools. An expansion in public schooling may lower the competitive price that private schools can charge and price out the less efficient private schools. However, it is also possible for them to enter given the likelihood of peer effects, and changes to the local economy and infrastructure with such a large-scale program. In the bottom panel of Figure 1.5, there is no evidence of crowd-out. If anything, there is mild evidence of crowd-in in 2005, which declines rapidly in the later years as funding is phased out.

What then drives the crowd in? On the one hand, the demand externality could raise the equilibrium price and pulls in private schools; on the other, the fall in operating costs may induce private school entry and lower the equilibrium price. In Section 1.3.3.2 I discuss how we can determine which of these mechanisms is stronger by seeing how the price changes. In Section 1.7.6 – and specifically Table 1.7 – it is clear that household expenditure on schooling falls suggesting that the cost-reduction mechanism is stronger.

In Section 1.7.8 I look at what drives this cost reduction. Figures 1.9 show that the initial increase in the supply of female teachers and teachers with a college degree may be an early driver of these effects – but they die out quickly. Infrastructure upgrades, like electricity and drinking water may help drive the cost reductions (Figures 1.9, 1.22 and 1.21).

The school building results, therefore, indicate that more government schools were built in DPEP areas, and there is no evidence of a crowd-out of private schools. We should expect that this would then increase the access to schooling for households in treated districts, and lower the marginal costs of going to school. A lower marginal cost will then lead to more years of schooling attained by the households.

1.7.3 Education and Earnings

The DPEP program was specifically targeted towards the primary and upper primary levels, and we may expect the largest impacts at those levels. Since the survey was collected 16 years after the start of the program, there has been enough time for many of these students to become part of the labor force. However, any student who was 19 at the time of the program (or 35 at the time of the survey) should be unaffected by the program since student enrollment sharply drops at that age, and child-labor laws prevent many workplaces from hiring children below eighteen.[58]

I check for a lack of a discontinuity in schooling attainment at the cutoff for persons above the age of 35, using the same RD methods. The left panel of Figure 1.6 shows how the older populations do not have any discontinuities in literacy, probability of finishing primary school, or upper primary school. The tables discussed below will also show there is no economically or statistically significant discontinuity in educational outcomes for older populations.

Looking at the younger population in the right panel of Figure 1.6, one can see discontinuities in different levels of education. Appendix Figure 1.15 shows the analogous figures for the full sample, rather than the sub-sample of those who reported earnings. The 2SLS-LATE numbers are shown in Table 1.3. The young have 0.57 more years of schooling in regions that were just eligible for the program. There is no statistically significant discontinuity in the older population.[59] These results indicate that the policy directly affected cohorts that were young enough to change their schooling attainment, and had no impact on the education of older cohorts.

Figure 1.7 and Table 1.4 show the RD impacts on education and log earnings for those who reported earnings, across the different bandwidth selection procedures and age groups. After scaling up the ITT estimates by the probability of treatment, the 2SLS increase in the years of education is 1.6 years, and younger students in regions eligible for the program had a 0.129 percentage point increase in the likelihood of finishing primary school.

In terms of earnings, there was an increase of about 0.25 log points for the younger cohorts (Table 1.4). Even though older populations had no discontinuity in education, their earnings are lower by about 0.21 log points due to the GE effects. As we would expect, I find that the general equilibrium effects are largest for close substitutes, like cohorts that were close to the treated cohorts. In Appendix Table 1.9, the sample is broken up into more age groups, and even though they are imprecisely estimated, there do seem to be larger effects on the 36 to 45-year-olds which is the closest age group to the treated cohorts.

Comparing Tables 1.3 and 1.4, it is clear that the impact on education is higher for the sub-sample that reported earnings.[60] However, as the top panel of Appendix Table 1.10 shows, there is no discontinuity in the probability that earnings are reported at the cutoff. This suggests that the policy did not lead to any differential selection into the group of persons that reported earnings.

The difference in the educational impacts between those that reported earnings and the full sample can be

[58] The Factories Act of 1948 and the Mines Act of 1952, govern the child-labor laws.

[59] As a robustness check, I restrict the sample to only those districts that had DISE school-level data. The results are seen in Appendix Table 1.21 where the estimates are slightly larger, and more precisely estimated.

[60] This is similar to Duflo (2001).

tied to the difference in labor market returns by gender. Only one-fourth of the sample reported earnings. One of the strongest predictors of whether earnings are reported is a person's gender, with males having a higher proportion of reported earnings. Persons who are engaged in domestic work, and this is mostly women, are least likely to report earnings. Individuals who are expected to gain in the labor market are the most likely to change their years of education. Since men are more likely to be in occupations that report earnings, and women are more likely to be engaged in domestic work, we should expect men to gain the most from increased education and therefore be more responsive to these interventions.

I find that the effects are concentrated among males, which is similar to the related literature (Ashraf et al., 2015; Breierova and Duflo, 2003). In Appendix Tables 1.11 and 1.12, one can see that the results are almost entirely driven by males. This is in line with the effects of this policy on enrollment estimated by Jalan and Glinskaya (2013). In the full sample, men increase their years of education by about 0.9 years, whereas women increase theirs only by 0.2 years. For the sub-sample of those who report earnings, however, the impact on education is similar in magnitude, but more precisely estimated for men. There is also little to no change in the earnings of women, even though men's earnings do rise.

To study where in the education distribution the impacts were felt, Table 1.5 looks at the fraction of people who have completed at least a given level of education. Since the program was targeted to the primary and upper-primary sections, the biggest increases are seen here. For the sub-sample that reported earnings, literacy rates are higher by 6 percentage points, and the likelihood of finishing primary school is higher by 12 percentage points.[61]

I support my RD results by also performing a difference-in-differences (DiD) analysis. In Appendix Table 1.13 I compare older cohorts to younger cohorts and DPEP districts to non-DPEP districts. For the full sample, the years of education are higher by 0.389 years, and for the sub-sample of those who reported earnings education is higher by 0.458 years and log earnings increase by 0.065.

1.7.4 Returns to Education: Conventional Instruments

The DPEP set-up is ideal for estimating the returns to education using household survey data. In my sample, a simple OLS regression of log earnings on years of education and a quadratic age profile, yields a Mincerian return of 10%. Instrumental variable (IV) estimates, however, will estimate a 2SLS-LATE weighted by the probability of being induced into getting more education by the instrument. Card and Lemieux (2001); Imbens and Angrist (1994); Oreopoulos (2006) discusses why IV estimates of the returns to education are larger than their OLS counterparts. In general, we may expect this to be the case since a reduction in marginal costs that affects all students equally will induce those with higher

[61] As another robustness check I collapse all the household data into district-age cells, and re-run the regressions. The results do not change, as can be seen in Appendix Table 1.14. Collapsing the data, however, is not recommended, since we are losing valuable information about the variability in the outcomes that may be different on each side of the cutoff – the optimal bandwidth procedure utilizes this variability.

returns into getting more education.[62] Another possibility is that OLS estimates suffer from classical measurement error (Griliches, 1977).

One IV method to estimate the returns to education is to simply use the RD cutoff to first estimate the change in the years of eduction for the younger cohort, and then the corresponding change in log earnings for the same cohort. By taking the ratio of the change in log earnings to the change in years of education, one can find an estimate of the returns to schooling. Under the assumption that the policy only induces some younger workers to get more education, this method will identify the change in earnings due to an additional year of schooling, for this marginal group. On the other hand, as the model shows, the policy should simultaneously affect both the skill premium and the overall output in the district. Since the change in the average earnings is not just driven by the switch in the fraction of students from unskilled to skilled groups, but also by the changes in earnings of skilled and unskilled workers, the estimated returns would be confounded by the changes in output and the skill premium.[63]

The estimates in Table 1.4 can be used to calculate the returns taking the ratio of the change in log earnings and the change in years of education. The ratio of 0.25 log earnings and 1.65 years gives us a return of about 15.5%.[64] However, due to the size of the confidence intervals, this number is not statistically indistinguishable from numbers as low as 7% and certainly not from the OLS estimated Mincerian returns of 10% estimated in this sample. This estimate, therefore, lies reasonably within the range of estimates found in the literature (Banerjee and Duflo, 2005; Psacharopoulos and Patrinos, 2004).[65]

Another IV method is to use a difference-in-differences (DiD) strategy. In Appendix Table 1.22 and Figure 1.17 I compare DPEP districts to non-DPEP districts, and the older cohorts to the younger cohorts.[66] I estimate the difference-in-differences coefficient for three different subsamples. For the full sample, there is an increase in 0.3 years of education, and a 5.5 percentage point increase in the literacy rate. There is also a 3.8 percentage point increase in the likelihood of finishing primary school. The estimates are similar even when restricting the sample to be in the neighborhood of the RD cutoff, and around the cohort-cutoff. For the subsample that reported earnings, there is also a statistically significant increase in earnings. The 2SLS IV-LATE returns can be estimated by taking the ratio of the change in log

[62] See Carneiro et al. (2011) for a nuanced alternative interpretation based on the generalized Roy model.

[63] From equation (1.26) we know: $log\frac{w_{y,D=1}}{w_{y,D=0}} = \ell_{sy,D=1}log\frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=1}log\frac{w_{uy,D=1}}{w_{uy,D=0}} + \Delta\ell_{sy}log\frac{w_{sy,D=0}}{w_{uy,D=0}}$. For changes in partial equilibrium, $log\frac{w_{uy,D=1}}{w_{uy,D=0}} = log\frac{w_{sy,D=1}}{w_{sy,D=0}} = 0$, and the change in average earnings across the RD cutoff recover the returns to skill $log\frac{w_{sy,D=0}}{w_{uy,D=0}}$ for the compliers $\Delta\ell_{sy}$.

[64] Appendix Table 1.18 shows the 2SLS-LATE version of this exercise, where the first-stage is the change in the years of education rather than the probability of receiving DPEP funds.

[65] A survey by Psacharopoulos and Patrinos (2004) finds Mincerian returns higher in low-middle income countries. In Asia these are near 10% and the returns to finishing primary schooling are around 20%. Banerjee and Duflo (2005) update this exercise, and document a range of Mincerian returns from 2.7% to 35.3%.

[66] For person i in age cohort a and district d, the following difference-in-differences regression was estimated:

$$y_{ida} = \beta_{DiD}T_{da} + \mu_d + \varpi_a + \epsilon_{ida} \ , \tag{1.31}$$

where μ_d is a district fixed effect, ϖ_a is a cohort fixed effect, and $T_{da} = 1$ if the individual lives in a DPEP district and is young enough to be affected. Under the usual parallel trends assumption, β_{DiD} is the difference-in-differences parameter.

earnings and the years of education. This 2SLS return is 15.9%, which is statistically and economically indistinguishable from the RD-2SLS return of 15.5%.

The difference-in-differences strategy, however, already accounts for some portion of the GE effect. Portions of the change in average earnings due to an increase in output, and the GE effects that affect older cohorts are differenced out. It is, therefore, impossible to estimate the overall GE effect using the DiD method without additional assumptions. It is, however, possible to measure the 'additional GE on the young' component by looking at how the skill-premium changes differentially for younger rather than older cohorts. This component depresses the returns to being skilled by about 7.9 percentage points (Appendix Table 1.22).

1.7.5 Returns to Education and the Labor Market General Equilibrium Effects

However, as pointed out in the model section, the methods of taking the ratio of the younger cohort's change in earnings and years of education is confounded by the fact that earnings are affected by the general equilibrium effects in the local economy. Average earnings of all persons in treated districts are affected by changes in overall output. At the same time, the change in the skill distribution and the inflow of skill-biased capital affects the earnings skill premium, as captured by Equation (1.23). While older cohorts are affected by the change in the aggregate skill distribution and inflow of skill-biased capital, younger cohorts are additionally affected by the change in the cohort-specific skill distribution for the young.

Given these general equilibrium effects it is necessary to use the method outlined in Section 1.4.1.1 and specifically, Equations (1.26) and (1.27) to derive the returns to education with and without the labor market general equilibrium effects. Table 1.6 estimates the returns by dividing the population into these skilled and unskilled groups. I define skilled workers as those having finished upper primary school as the policy was targeted at getting students through this level of schooling, and because the largest earnings increase in OLS regressions on untreated districts comes when a student finishes upper primary school.[67] There was a 17 percentage point shift into the skilled category across the cutoff.

The estimated returns to shifting into the skilled group in the absence of GE effects is 19.9%. The returns to being skilled with the GE effects, however, are only 13.4%. This constitutes a 32.5% decrease in the returns attributable to the GE effects. This change in the skill-premium can be split up into the portion that affects all cohorts, and the additional impact only on the young. 91.87% of change in the GE effects are explained by the 'additional impact on the young' term. The GE effect that affects all cohorts may be small because the two components that determine this effect may counteract each other – an increase in the relative supply of skilled workers will tend to lower the skill premium, but an inflow in skill biased

[67] In going from literate-below primary to finishing primary schools, average earnings increase by 10%, whereas in going from primary to upper primary school average earnings increase by 20%.

capital may increase this skill premium. Furthermore, the additional impact on the young term may be high because the young and the old may not be close substitutes in production.

1.7.6 Total Output, Consumption, and Educational Expenditure

The change in overall output depends on the productivity of the different skill levels and the shift in the labor force from one skill level to another. As workers acquire more skill, and/or if skill-biased capital flows into the region, overall productivity and output in the region may increase. In the top panel of Figure 1.8, one can see the impact on total output (the District Domestic Product). These regressions are underpowered, and the standard errors are quite large. Between 2000 and 2006, the increase in GDP associated with the policy was between 0.137 and 0.190 log points (Appendix Table 1.24).

The change in total output will lead to a change in total consumption. In the top panel of Table 1.7, one can see that the change in consumption expenditure in the last year of the policy (2004-5) was about 0.17 log points. At the same time, in 2004 the money spent for educational purposes (tuition, fees, books and stationery) falls by about 0.21 log points.[68] In the bottom panel of Figure 1.8 one can see the discontinuity in education expenditure. This fall in educational expenditure is persistent even five years after the program ended in 2009. The decline in total expenditure on education-related items is driven largely by lower expenditures on school tuition and fees. There is, in fact, an increase in expenditure on other education-related items, like books and stationery (Table 1.7). We may expect expenditures on books and stationery to rise when households gain more education, as these are complementary to the consumption of education.

Changes in consumption and the costs of education will directly impact overall economic benefits. The increase in output and consumption benefit all cohorts, whereas the fall in the costs of education benefit those in the younger cohorts who attend school. The fall in the costs of schooling even benefit those who are *not* induced into getting more education – at an extreme, policies that successfully reduce the costs of schooling can have significant economic benefits even if they do not change the equilibrium years of education.

1.7.7 Productivity & Movement of Firms, Capital, Workers

Local economies at the district level that received educational funds for at least a decade witnessed a transition in the skill level for younger cohorts in their workforce. For this to have happened, any combination of the following four things may have taken place. First, skilled workers may have migrated out, dampening the portion of the GE effects that depends on the change in the aggregate skill distribution. Second, existing firms may have switched the composition of their workforce by hiring more skilled

[68] This is consistent with Jalan and Glinskaya (2013) who estimate a 20-40% fall in household educational expenditures.

workers. Third, new firms may enter and hire these skilled workers. Last, workers may have utilized their increase in skill and adopted newer technologies in production. The entry of skill-biased capital, therefore, will increase the returns to skill in treated districts and be a crucial determinant of the GE effects.[69]

To test the first possibility about worker migration, I assemble the 2007-8 round of the NSS household survey which asked detailed questions on migration. Worker migration is extremely low in the Indian context (DasGupta, 1987; Munshi and Rosenzweig, 2009; Topalova, 2005).[70] It is, therefore, unlikely that those who acquired skills migrated out of these districts. By analyzing the NSS 2007-8 waves, we can see that of all the households that reported having any migrants across districts, only 30% of the migration was work related, whereas more than half were for marriage reasons. Appendix Table 1.10 supports the claim that the policy did not impact migration. There are no economically significant changes to the number of out-migrants or the number of households that migrated to that district.[71]

On the other hand, firms are relatively more mobile in India (Ghani et al., 2015). I compile data from the Annual Survey of Industries (ASI), which is a census of all manufacturing firms. The results for these are shown in Appendix Figure 1.18. In Figures 1.18, one can see that even at the manufacturing establishment level, the average wage paid to workers increases as educated workers join the labor market, in and around the year 2004. Furthermore, I classify firms based on their products as ‘high-skill’ firms. The figure shows that there is a steady increase in the fraction of firms that produce more mechanized products. This is suggestive of the fact that either existing firms shifted production and employed more high-skill workers, or newer firms entered and hired these skilled workers. Both mechanisms are suggestive evidence in support of an inflow of skill-biased capital into these regions.

One relevant question is where is this capital flowing from, and in the absence of the policy would it have gone to regions that lie just on the other side of the cutoff. If this is indeed the case, then it would attenuate the GE effects on earnings. It is, however, unlikely that regions *just* above the cutoff receive less capital due to the policy. First, policy regions are geographically dispersed all over the country (Figure 1.12) rather than being neighbors of districts just on the other side of the cutoff. Second, anecdotal evidence support stories of people residing in major cities – that have very high literacy rates – often investing in

[69]Regions around the RD cutoff are geographically dispersed, so it is less likely that the migration of firms or workers happens among regions near the cutoff.

[70]Many studies on India are explicit about ignoring migration (Anderson, 2005; Banerjee et al., 2008; Foster and Rosenzweig, 1996). Munshi and Rosenzweig (2009) show that for a sample of rural males aged 20-30, the permanent migration rate outside their village was 8.7%, a lot of which may have taken place within the same district. Deshingkar and Anderson (2004) also show that rates of rural-urban migration are much lower in India than in comparable countries, and Munshi and Rosenzweig (2015) show that male worker migration is extremely low despite the presence of large wage gaps across regions. One possible reason lies in the uncertainty related to getting work at the destination and the fixed cost of migrating (Bryan et al., 2014). Duflo and Pande (2007) argue that the district is the relevant local labor market in the Indian context, and workers of different skills can find employment elsewhere in their own district.

[71]For example, the total number of out-migrants ranges between 4.2 and 10.9 persons per district - this includes migrants for any purpose (like marriage, education, temporary work, etc.). It is not possible to find RD estimates by finer skill groups or age cohorts since almost nobody is migrating.

villages that they originate from, suggesting that the source of capital are not regions near the cutoff. In Figure 1.24, I look at the density of capital-intensive firms in the early period and the the late period for the part of the country that should not have received the policy. Regions near the cutoff (normalized to 0), if anything, have an increase in the firms involved in mechanized production and providing higher compensation. On the other hand, regions with high female literacy – often the major cities – show a mild decrease. This is suggestive evidence to support the fact that, if at all, that as the workforce changes in DPEP regions, capital may flow out of the major cities but not the regions that lie just above the cutoff.

In general, there are some clear changes to the labor market for the workers in these regions. The bottom half of Appendix Table 1.10 shows that the probability of being paid monthly (as opposed to daily) is higher, and the fraction unemployed is lower in the treated regions. The last possibility, that workers adopted newer technologies given their increased levels of education is, therefore, likely to be true in this context (Foster and Rosenzweig, 1996).

1.7.8 Teachers, Infrastructure, and Other Funding

While the primary focus of the program was to increase educational attainment by building schools and hiring teachers, there may have been improvements in quality given such a large amount of funding. On the one hand, such improvements may have increased the returns to schooling in the labor market, attenuating the GE effects on the returns. In Appendix Table 1.23 I use a relatively recent dataset – known as the Annual Status of Education Report (ASER) data. This is geographically the most comprehensive test-score dataset in the country. I consider six different test score variables, and only one of them shows a significant increase – being able to identify numbers between 1 and 9 has a 5 percentage point increase at the cutoff. This is, at best, mild evidence of increasing test scores, and if better test scores translate into higher returns then it may attenuate the estimated GE effects. On the other hand, better 'quality' in terms of better infrastructure may have made it easier for students to finish a grade and further lower the marginal costs of schooling. In this subsection, I explore how various inputs at the school level were changed around the RD cutoff. Furthermore, I can study what happens when the DPEP funding dissipates over time.[72]

Card and Krueger (1992) show that more qualified teachers and female teachers have important impacts on schooling in the US. In India, female teachers may also encourage female student enrollment and are, therefore, important. In 2005, when program funding was still high, the number of college-educated teachers and the number of female teachers in DPEP districts was higher. However, this discontinuity dissipated over time, showing that once the funding is no longer targeted to DPEP districts, there is a low

[72]Even though funds to DPEP districts declined relative to other regions, overall funding for schools in the country did not fall.

retention rate of such teachers (Figure 1.9). A lack of targeted funds may have led to a lack of incentives for retaining these teachers.[73]

Since, under DPEP, funding was stepped up to districts below the cutoff, there may have been a crowd-out of other funds that schools were supposed to receive. The Teachers Learning Material (TLM) grant is funding that is available to schools regardless of whether they lie in DPEP districts or not. In the top panel of Figure 1.10, one can see that regions that were eligible for DPEP systematically spent less TLM funds, showing the possibility that other funds were actually crowded out when DPEP funds were allocated.[74]

Tangible infrastructure, however, seems to last even when the DPEP funding is reduced (Figures 1.9 and 1.10). Drinking water, electricity, and library books are consistently higher in regions that received the DPEP (Appendix Figure 1.22). Infrastructure upgrades like girls' toilets may be important in getting girls to school. While there was constant funding, there were significantly more facilities for girls, and there is a slight dissipation as the funding is stopped. Other inputs, such as medical checkups, are also consistently higher for schools in DPEP districts.

The condition of the classrooms also seem to have deteriorated over time once funding was stopped. While, in 2005, schools in DPEP regions had a lower number of classrooms needing some repair, over time more of these classrooms broke down (Figure 1.10). These results indicate that a constant source of funding may be needed to keep the rooms in good condition.[75]

One significant change in the DPEP regions is the introduction of pre-primary sections, which was thought to be a good way to get children into schools at a young age. Many more schools in DPEP regions have such pre-primary sections after the policy, and there are more pre-primary teachers and students in these schools (Figure 1.10 and Appendix Figure 1.20).

Under the DPEP regional educational centers called Block Resource Centers (BRCs) and Cluster Resource Centers (CRCs) were built, with facilities for training teachers, and other learning materials that teachers could access. There were also government officials at these centers who would visit the schools, and could assist with teacher training at these schools. In Figure 1.10 and Appendix Figure 1.21, it is clear that the distance to the closest center was lower for DPEP regions, since many more centers were built under DPEP. Over time, however, once the funding was reduced, centers continued to be built in non-DPEP regions, and the differential effect dissipated. In the lowermost panel, however, one can see

[73] The dissipation in the discontinuity does not necessarily imply that teachers left the DPEP districts – it may also be the case that non-DPEP districts hired teachers at a relatively more rapid rate once the DPEP funding was gone.

[74] Appendix Figure 1.19 shows that DPEP regions both received less and spent less of TLM grant money.

[75] While there are clear changes to tangible infrastructure, it is unclear whether these may have contributed to any changes in the quality of instruction. In fact, quality may suffer if new schools are forced to employ teachers they would not have otherwise employed, or if there are negative peer effects because of an increase in low ability students. Furthermore, the policy did lead to a crowd out of other funds (Figure 1.10). There may also have been a crowd-out in extra-curricular work – in Appendix Figure 1.23 we see that treated regions employed fewer teachers for non-teaching assignments and students spent fewer days in non-teaching assignments.

that the number of academic inspections and visits by center officials was, over time, consistently higher in treated areas.[76]

1.7.9 Overall Economic Benefits

Changes in overall output and reductions in the total cost of schooling will directly impact benefits to households. The change in labor market earnings, however, depend on the returns to skill and the GE effects on these returns. Table 1.6 shows the returns by skill group, which helps back out the parameters and the changes in yearly labor market benefits shown in Table 1.8. For these calculations, the average real interest rate during that period was used (5%). Furthermore, a gap of 10 years is assumed between the time the costs of education are borne and the labor market returns are realized.[77]

In the top panel of Table 1.8, I present the results for those in the younger cohort who were induced into getting more skill because of the policy. This is about 17% of the young population. Their welfare increases by 0.121 log points, and the GE effects depress this increase in welfare by 23.3%. At the same time, workers who were always going to acquire skill even in the absence of the policy are worse off by 0.037 log earnings points, whereas workers who were always going to be unskilled are better off by 0.014 log earnings points.[78]

Since unskilled workers are better off and skilled workers are worse off, it is also possible to estimate the transfer in labor-market benefits from the skilled to the unskilled due to the GE effects. Among the older cohorts this transfer is 0.004 log points, and among the young it is 0.05 log points. This indicates that purely when looking at labor-market benefits, those persons who were always going to be skilled even in the absence of the policy, actually lose out, whereas those who were never going to acquire skill even in the presence of the policy benefit.

To measure the change in lifetime welfare for students induced into getting more schooling, I compare the cost of an additional year of schooling to the benefits in the bottom panel of Table 1.8. These costs include not just the tuition fees but also the opportunity cost of a foregone unskilled wage. The benefits, however, are the present discounted value of the skilled earnings stream. All cohorts and skill groups benefit from increases in the overall output. Furthermore, the young cohorts who acquire skill benefit

[76] One last consideration is the medium of instruction in such schools. English-medium education may have greater potential returns in urban labor markets but higher costs for the students who are unfamiliar with the language. At the same time, Hindi-medium education may be valued elsewhere in the country, whereas regional languages are only valued in certain localized areas. In left panel of Figure 1.22, one can see that the schools in DPEP regions are more likely to be Hindi-medium and less likely to be in regional languages. While the discontinuity is slight, there is sharper evidence of a kink at the cutoff indicating that the relationship between the medium of instruction and literacy changes across the DPEP cutoff.

[77] The average real interest rate comes from the World Bank. Changing the interest rate or the gap of 10 years does not affect the percentage change in welfare due to the GE effects, only the levels.

[78] Note that these results focus on labor-market benefits. A policy such as this should also change the prices of non-tradables, like land, affecting the welfare of non-workers as well. Given the scant number of land transactions in the data, there is no discernible effect on land prices.

from the reductions in the total costs of schooling. Even young students who were not induced into getting additional education benefit from the reductions in schooling costs.[79]

1.8 Conclusion

In this paper, I show that large-scale education investments can and do generate substantive general equilibrium effects in the labor market and the education sector. Bringing together a school-level dataset, census data, household surveys, and firm-level data, I perform an intensive analysis of the DPEP program, which measurably increased educational inputs and increased the years of education and earnings for students. With the help of the policy, I estimate the parameters of a general equilibrium model using a RD approach. The estimates imply that the return to acquiring skill is 13.4%, but that is 6.5 percentage points lower than it would be in the absence of general equilibrium effects. There are also large distributional effects, where labor market benefits are transferred from the skilled to the unskilled, especially among the young. High-skill workers who would have acquired skill even in the absence of the policy lose out in terms of labor market earnings. Overall welfare, however, is higher, driven by decreases in the household's costs of education and increases in output in the local economy.

These findings have two important implications. First, we should take care when extrapolating the benefits demonstrated by small-scale schooling interventions, as scaled up versions of such interventions may have GE effects. Second, using large-scale variation to estimate the returns to education may conflate the individual returns and the general equilibrium effects. This is because an experiment where a single individual receives more education is inherently different from the variation induced by changes in policies like nationwide tuition subsidies, schooling expansions, compulsory schooling laws or other regional variation. One limitation of my study, however, is that the estimates are not generalizable to regions further away from the RD cutoff in the absence of stronger assumptions.

While there was a larger fraction of new schools built under these policies, over time other regions caught up once the funding was cut. There were lasting impacts, however, on physical infrastructure such as electricity and drinking water. As the funding was cut, the gap in the number of more qualified teachers across eligible and ineligible districts dissipated, and the condition of classrooms deteriorated over time. In light of these results, it is reasonable to propose that if policy makers wish to retain teachers, then the regions would require a constant source of funds over a long period.

Furthermore, implementation by the government is inherently different from researcher-based experiments. In a companion paper (Khanna, 2015), I compare this policy to more decentralized policies that

[79]These results do not necessarily indicate that the policy was cost effective. I have shown that the directs impacts were concentrated on men, that reported earnings, and only for certain cohorts. In other results I find that the impacts were mostly restricted to the treated districts that had relatively high literacy rates. The interventions had low persistence as well. Given the large amounts of funds invested, the overall cost effectiveness of this policy is questionable, and is left for future research.

were implemented at the sub-district level. Eligible sub-districts were identified on dual criteria, that allows me to use a multi-dimensional RD (MRD) approach to isolate the impacts.[80] I find evidence suggesting that the decentralized policies were more effective at raising literacy rates, especially for girls.

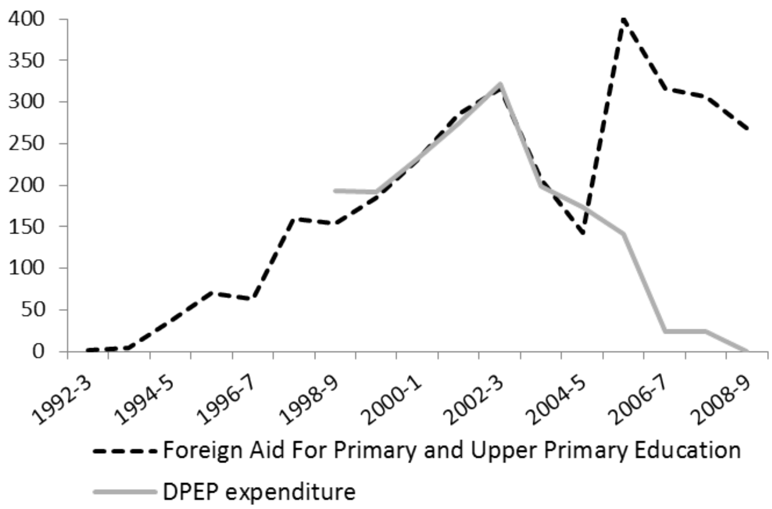
The debates about the role of the government in health and education investments usually center around the economic benefits of the policy. I show that economic benefits to household depend on a few crucial factors — the costs of education, the labor-market returns to education, and the general equilibrium changes in earnings. While these are sufficient in capturing the direct economic benefits, more education can have other welfare consequences as well. For example, more education can lead to better health or more informed political participation (Sen, 1999). Exploring these relationships is left for future research.

Identifying who benefits and who does not in the universe surrounding such a policy, and what works and what does not is key to making such large-scale infrastructure investments more targeted and effective. The results in this paper, however, help explain why scaled up government policies may have different impacts than researcher-led micro interventions (Acemoglu, 2010; Bold et al., 2013a; Deaton, 2010). In light of these results, it is clear that understanding all the consequences of large general equilibrium effects is crucial for both researchers and policy-makers when considering nation-wide interventions in public policy.

[80] The paper is also one of the first to use a new empirical strategy: the fuzzy multi-dimensional regression discontinuity method, and details various estimation procedures that future researchers may apply. Since the MRD approach is new to the literature, I provide Monte-Carlo evidence on the best estimators in such a context. Since many policy interventions may have more than one cutoff with imperfect compliance, this estimator can be used in other contexts as well.

1.9 Figures

Figure 1.1: Foreign Aid and DPEP Expenditure (in USD mn)

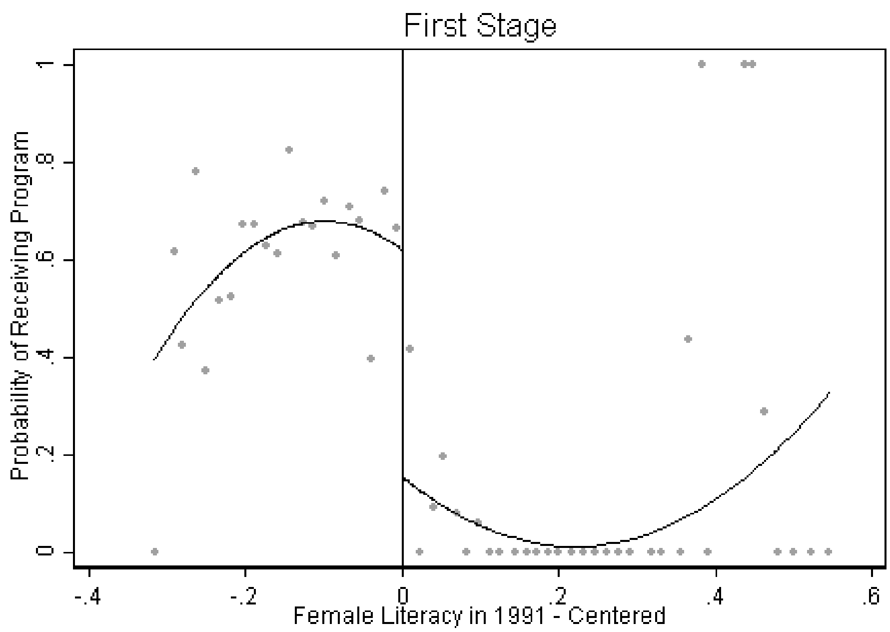


Foreign aid for expenditure on primary and upper primary education, and funds disbursed for DPEP. 1999 Indian rupees converted to USD using the 1999 exchange rate of Rs. 40 to $1. Sources: Foreign Aid from Colclough and De (2010). DPEP expenditures data compiled by author from Ministry of Human Resources and Development Reports, National Institute of Educational Planning and Administration, Lok Sabha Unstarred Question Numbers: 1807- 07.03.2006; 552- 24.02.2009; 55 - 26.02.2008; 267- 22.03.2005; 1320- 10.12.2003; 2018- 4.3.2003, and Rajya Sabha Unstarred Question No. 2855- 19.04.2002.

Figure 1.2: Social Sector (Health and Education) Grants/Loans from Central to State Governments

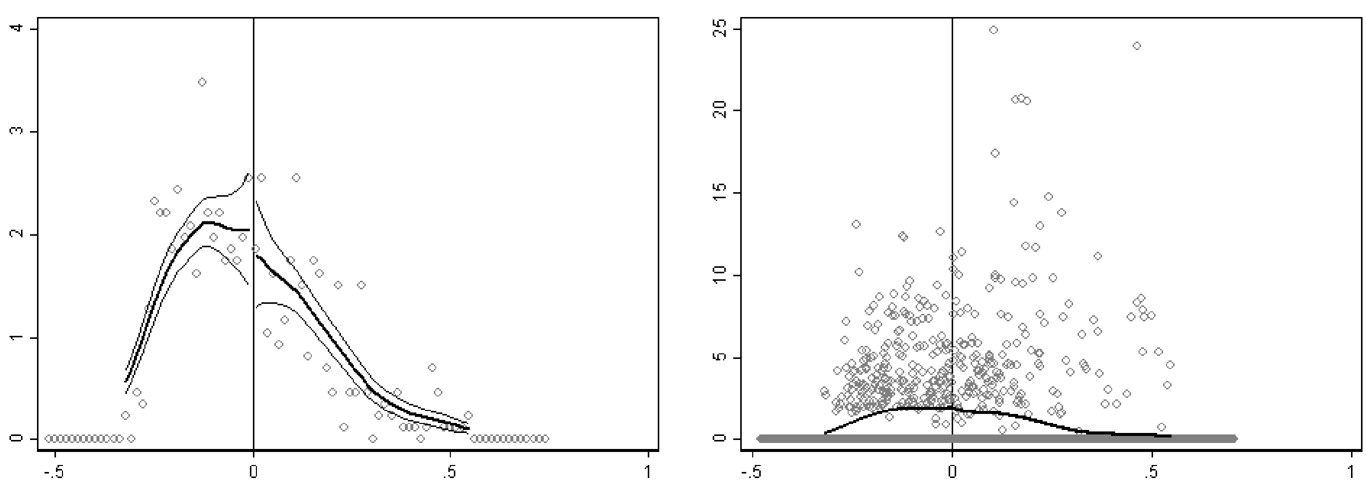
Central government grants and loans to State governments for spending in the social sector (health and education), and as a proportion of total grants/loans. 1999 Indian rupees converted to USD using the 1999 exchange rate of Rs. 40 to $1. Source: External Assistance Brochure of CAA&A, Department of Economic Affairs, Ministry of Finance, Government of India.

Figure 1.3: First Stage of DPEP



First stage graphs showing probability that a district received DPEP funds. Optimal bin sizes calculated using Calonico et al. (2014b) method.

Figure 1.4: McCrary Density Tests

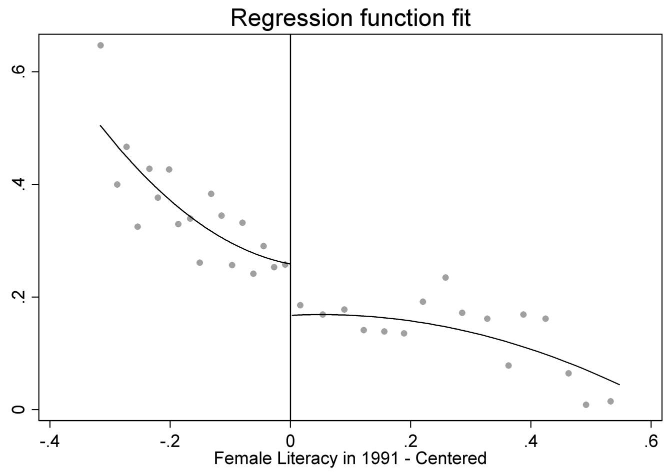


District Level McCrary Test

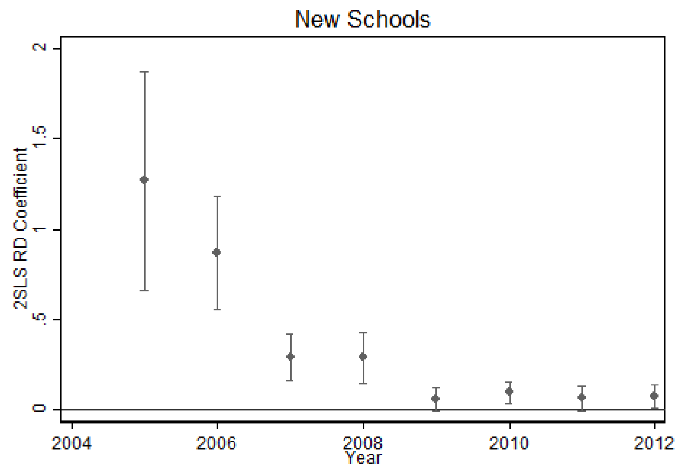
Household Level McCrary Test

McCrary (2008) tests for discontinuity in density at the cutoff. These tests look for evidence of one-sided manipulation of the running variable by testing the discontinuity in the density at the RD cutoff.

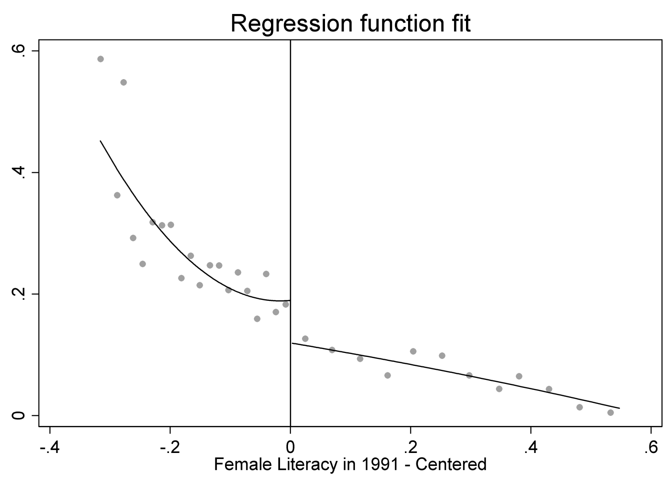
Figure 1.5: Fraction of Schools Built Post 1993



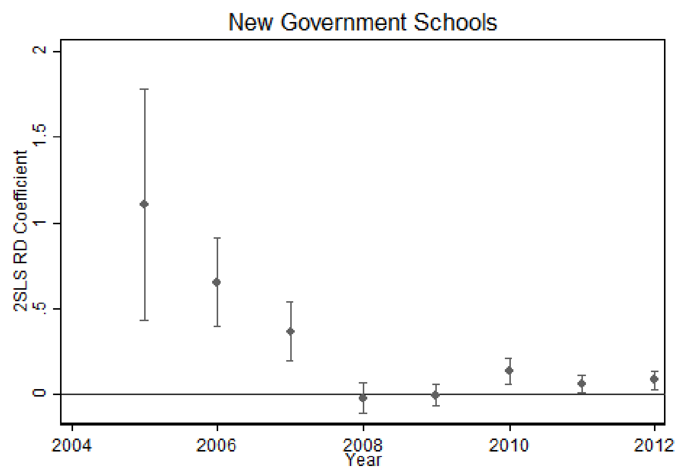
Fraction of All Schools Built Post 1993



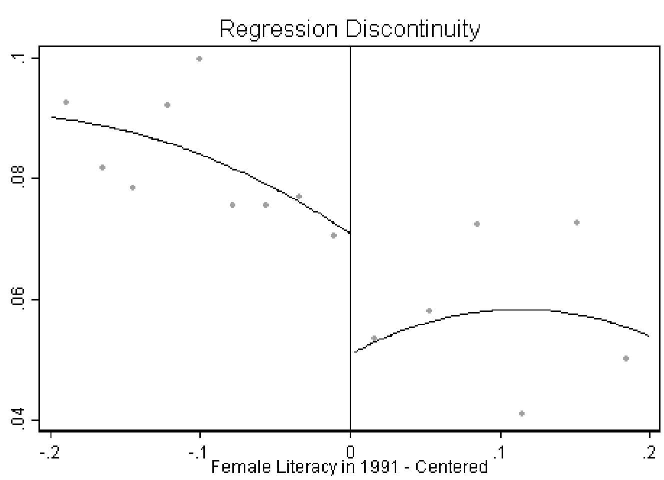
All Schools - RD Coefficient Over Time



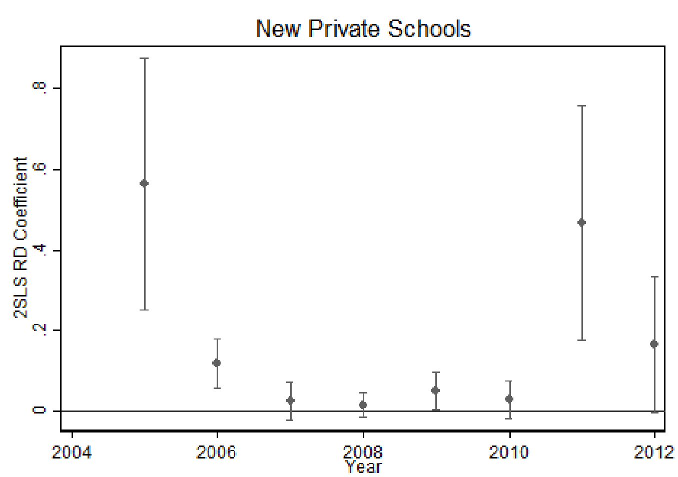
Fraction of Government schools Built Post 1993



Government Schools - RD Coefficient Over Time



Fraction of Private schools built post 1993

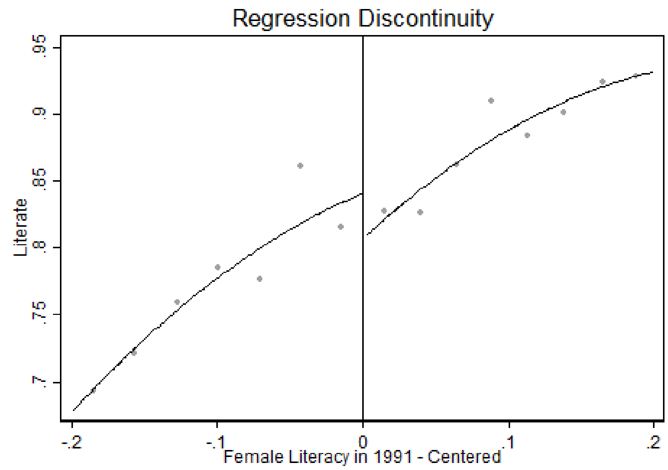


Private schools built post 1993

Source: DISE (District Information System for Education) data. Top panels show RD graphs using the 2005 data. RD graph optimal binning and 2SLS RD coefficients calculated using Calonico et al. (2014b) procedure.

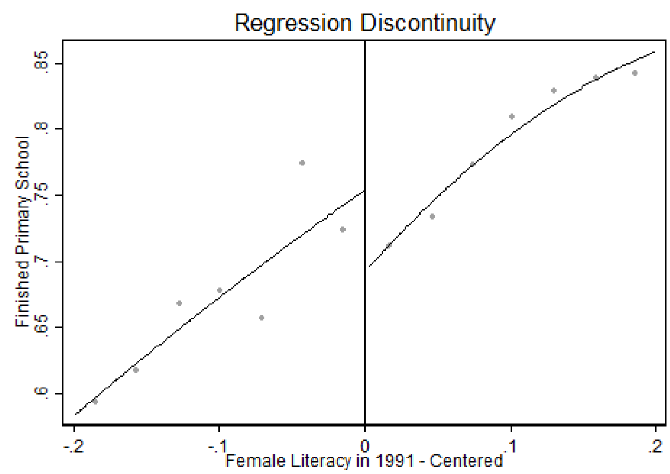
Figure 1.6: RD figures - Levels of Education

Literacy - Older



Literate - Younger

Finished Primary School - Older



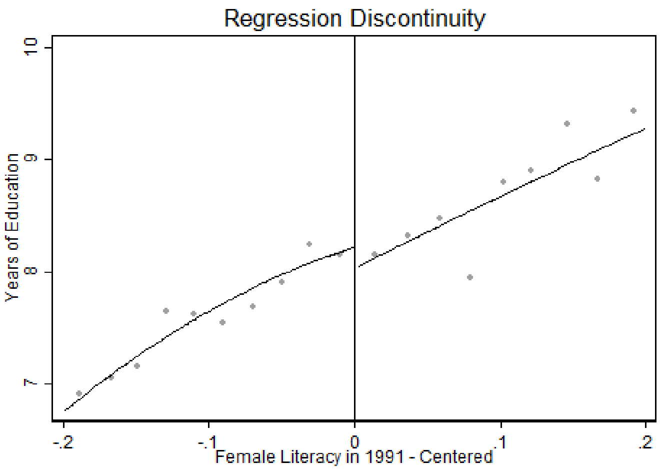
Finished Primary - Younger

Years of Education - Older

Upper Primary - Younger

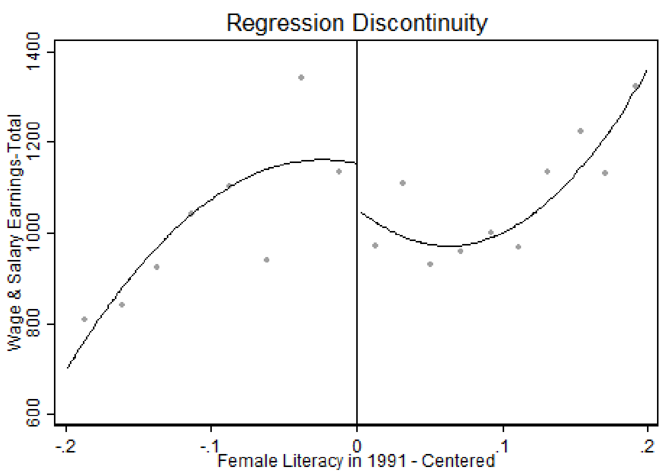
National Sample Survey 2009 for persons who report earnings in primary occupation. Appendix Figure 1.15 shows the analogous graphs for the full sample of persons. Figures made using Calonico et al. (2014b) method of using regression curves to approximate the conditional means on either side of the cutoffs and the equally spaced sample means, and optimally spaced bins.

Figure 1.7: RD Years of Education and Earnings - Young Sample



Years of Education

Years of Education - Earnings Reported

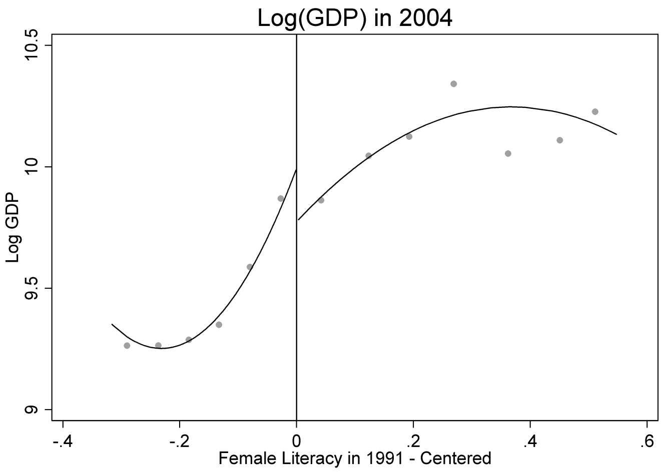


Earnings (Rupees)

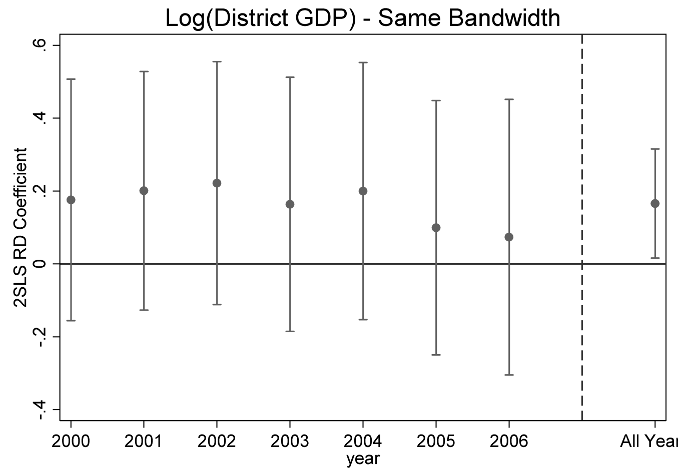
Log(Earnings)

National Sample Survey 2009 for persons below 35 years of age. ‘Earners’ refers to those who report earnings in primary occupation. Figures made using Calonico et al. (2014b) method of using regression curves to approximate the conditional means on either side of the cutoffs and the equally spaced sample means in optimally spaced bins. Average exchange rate in 2009 is Rs. 40 = $1.

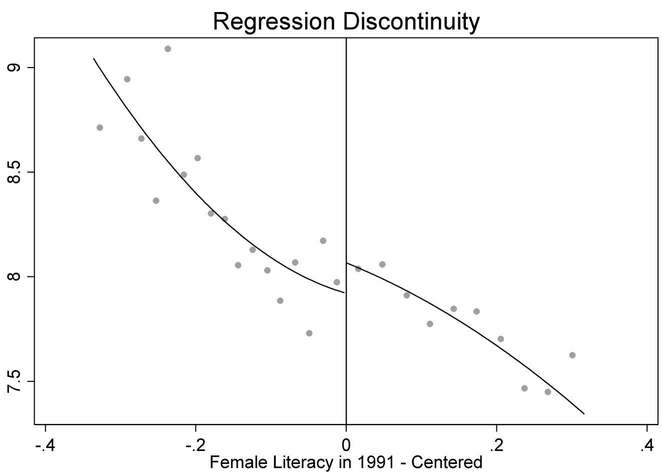
Figure 1.8: Change in Overall Output and Household Expenditure on Education



Log District GDP in 2004



RD 2SLS Coefficients - Same Bandwidth



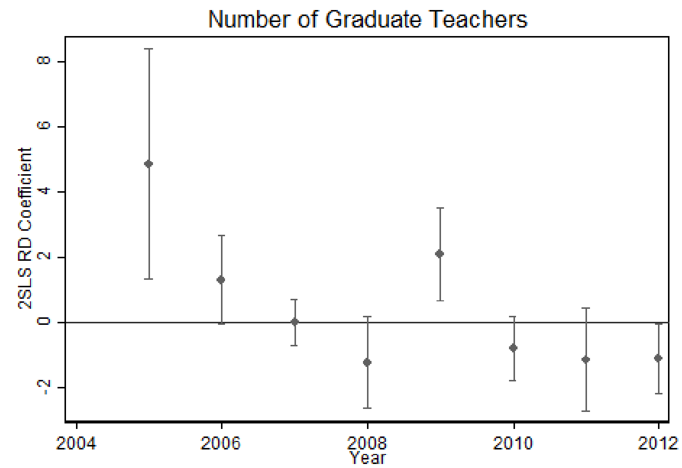
Log(Educational Expenditure)

Log(Expenditure on Tuition/Fees)

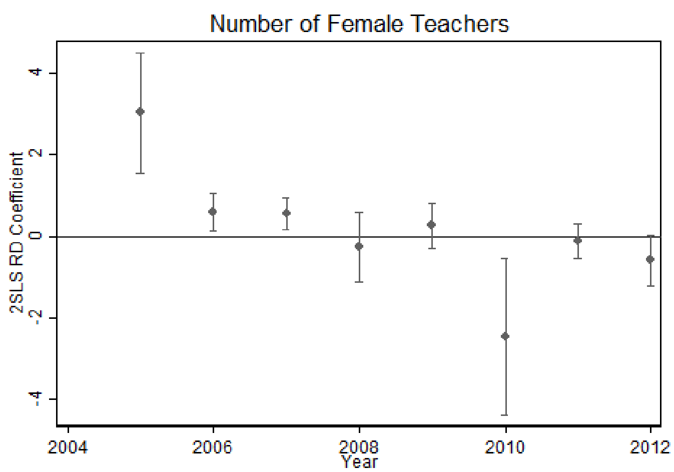
RD graph optimal binning and 2SLS RD coefficients calculated using Calonico et al. (2014b) procedure. ‘Same bandwidth’ restricts bandwidth to be the same as the first year’s optimal bandwidth.
Educational Expenditure Source: National Sample Survey 66th Round.
District Domestic Product Sources: Department of Statistics and Programme Implementation, Government of West Bengal; Planning Commission; Directorate of Economics and Statistics Government of Uttar Pradesh; Department of Economics and Statistics Government of Tamil Nadu; Directorate of Economics and Statistics Government of Rajasthan; Department of Planning Government of Punjab; Planning and Coordination Government of Odisha; Directorate of Economics and Statistics Government of Maharashtra; Directorate of Economics and Statistics Government of Kerala; Planning Programme Monitoring and Statistics Department Government of Karnataka; Directorate of Economics and Statistics Government of Bihar; Directorate of Economics and Statistics Government of Assam; Andhra Pradesh State Portal.

Figure 1.9: Teachers, Drinking Water, Restrooms and Electricity

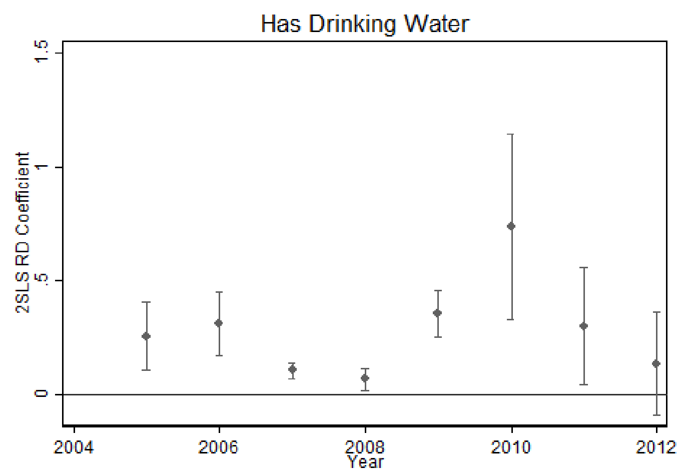
Teachers (per school) with College Degrees



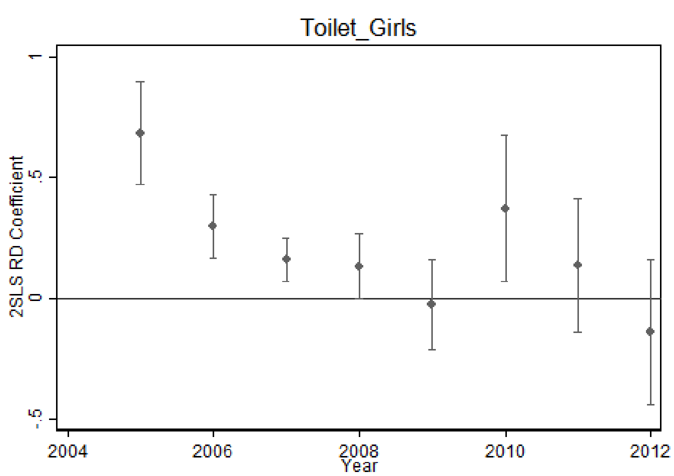
Teachers (per school) with College Degrees



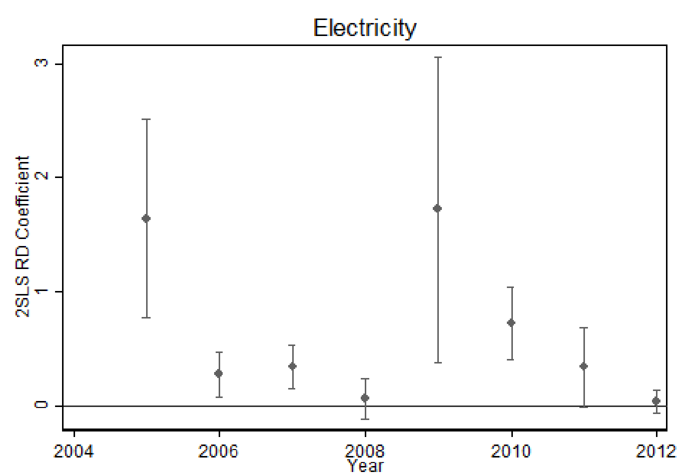
Female Teachers (per school)



Drinking Water



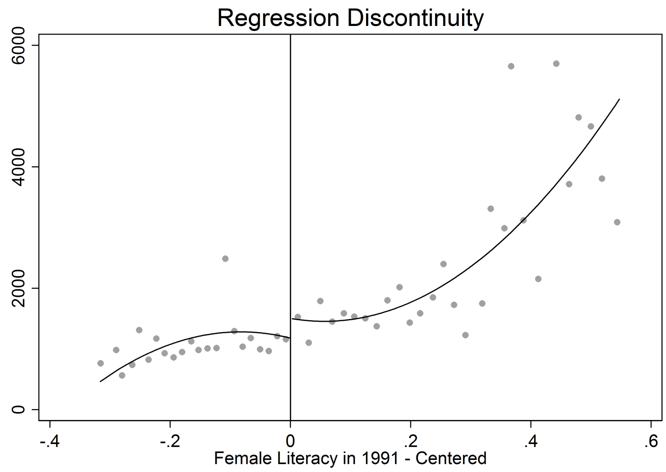
Girls' Restrooms



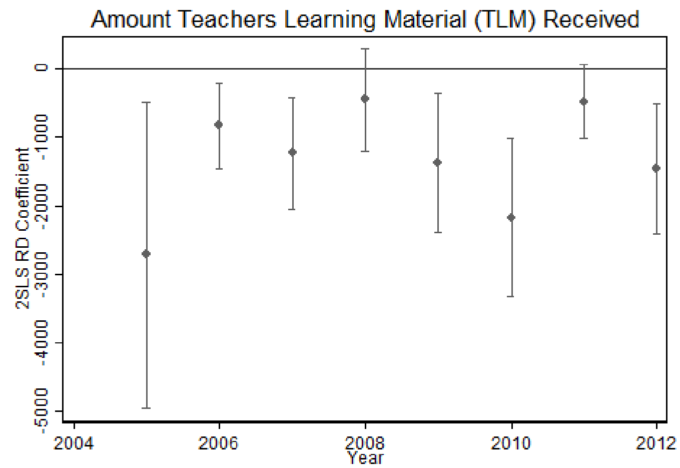
Coefficient Over Time: Electricity

Source: DISE data. RD graphs (Regression Function Fit) use the 2005 data. RD graph optimal binning and 2SLS RD coefficients calculated using Calonico et al. (2014b) procedure.

Figure 1.10: Crowd Out of Other Funds, Classrooms, and Pre-Primary Sections

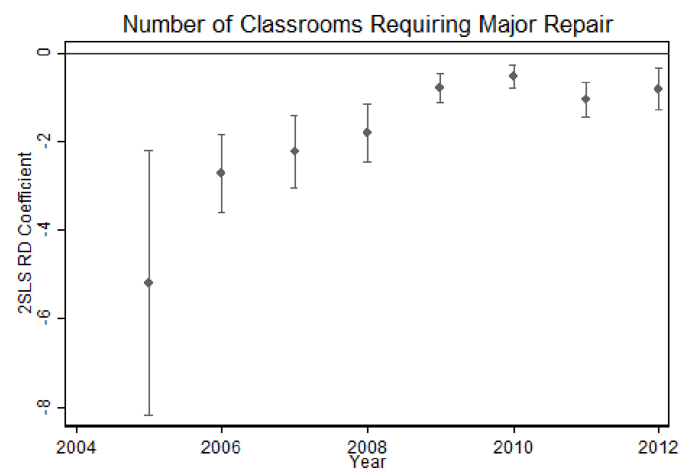


TLM grants Spent (2005)

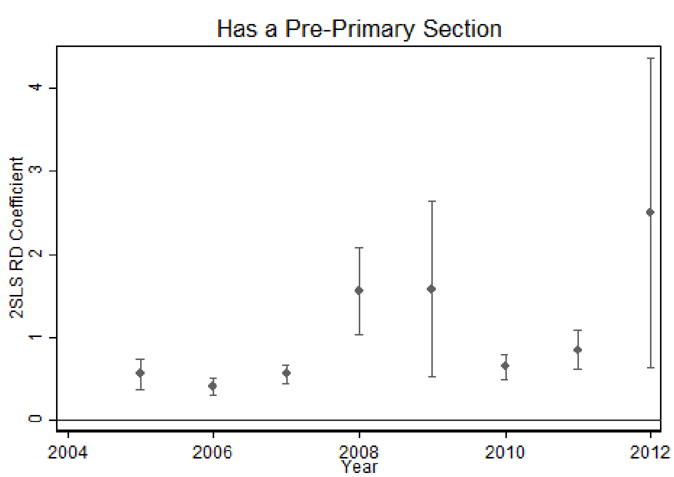


RD Coefficient Over Time: TLM grants Received

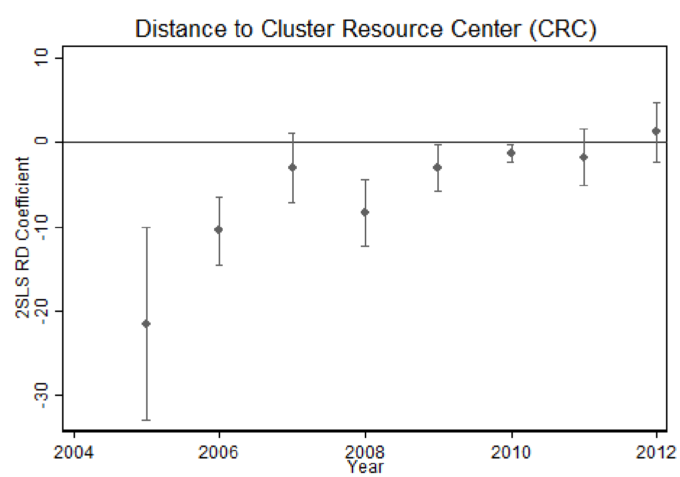
Classrooms Needing Major Repair (2005)



RD Coefficients: Classrooms Need Major Repair



Coefficient Over Time: Pre-Primary Schools



Coefficient Over Time: Distance to CRC

Source: DISE data. All schools, regardless of their which district they are in, are eligible to receive the Teacher Learning Materials (TLM) grant. Cluster Resource Centers (CRCs) provide facilities and training to teachers. RD graphs for TLM grants received (rather than spent) can be found in Figure 1.19. Classrooms needing minor repair and in good condition can be found in Figure 1.19.

1.10 Tables

Table 1.1: Summary Statistics: School Level (2005)

	Mean	SD
Fraction of Schools:		
Built post 1993	0.277	0.447
Gov schools built post 1993	0.200	0.400
Pvt school built post 1993	0.075	0.263
Built between 1973-93	0.227	0.419
Gov schools built 1973-93	0.170	0.376
Pvt Schools built 1973-93	0.055	0.228
Fraction of Schools Having:		
A Girl's Toilet	0.400	0.490
Electricity	0.312	0.463
Playground	0.549	0.498
Medical Checkups	0.541	0.498
Ramps	0.182	0.386
A Boundary Wall	0.506	0.500
Drinking Water	0.846	0.361
A Pre-primary section	0.213	0.410
Block and Cluster Resource Centers:		
Visits by BRC Official	1.485	2.543
Distance to BRC (km.)	13.462	15.936
Visits by CRC Official	4.496	5.612
Distance to CRC (km.)	4.438	8.689
Teacher Learning Materials Grant:		
Amount Received (Rs.)	1517.100	8010.138
Amount Spent (Rs.)	1332.604	7611.869

Source: DISE (2005). Fraction of schools are for schools that still exist in 2005. BRC is Block Resource Center, and CRC is Cluster Resource Center. All schools, regardless of DPEP status, are eligible for Teacher Learning Material Grants (TLM).

Table 1.2: Summary Statistics: Household Level

	Non DPEP Mean	Non DPEP SD	DPEP Mean	DPEP SD	All Mean	All SD
Finished Primary School	0.71	0.45	0.60	0.49	0.67	0.47
Finished Upper Primary	0.59	0.49	0.48	0.50	0.55	0.50
Years of Education	7.40	5.26	6.14	5.38	6.95	5.34
Male	0.50	0.50	0.50	0.50	0.50	0.50
Age	37.75	14.63	37.39	14.59	37.59	14.62
Weekly Earnings	42.17	51.29	31.55	38.50	38.92	47.43

Source: National Sample Survey (2009). Age in years. Earnings in 2005 USD, where Rs. 40 = $1.

Table 1.3: Education Changes - Full Sample

Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.573	0.279	0.571	0.303
	(0.190)***	(0.242)	(0.185)***	(0.224)
Observations	61,787	34,119	65,650	41,893
Fuzzy Conventional p-value	0.00253	0.249	0.00205	0.175
Fuzzy CCT Corrected p-value	0.00175	0.168	0.0246	0.103
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Finished Primary School	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.0365	0.0252	0.0574	0.0263
	(0.0191)*	(0.0226)	(0.0154)***	(0.0218)
Observations	39,326	36,584	68,050	40,068
Fuzzy Conventional p-value	0.0562	0.264	0.000185	0.227
Fuzzy CCT Corrected p-value	0.130	0.419	0.0840	0.404
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10, for all districts, and all persons between the ages of 16 and 75 (including those who did not report earnings).

The sample of ‘Below 35 years’ are of school going age during the policy, whereas those ‘Above 35’ are too old to change their schooling in response to the policy.

Bandwidths: ‘CCT’ is the Calonico et al. (2014b) method. ‘I and K’ is the Imbens and Kalyanaraman (2012) method. ‘CCT corrected p-value’ is the bias-corrected p-values using the method in Calonico et al. (2014b).

Table 1.4: Education and Earnings for those with Reported Earnings

Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	1.654	-0.381	1.569	-0.199
	(0.491)***	(0.590)	(0.417)***	(0.553)
Observations	10,175	7,997	14,277	8,630
Fuzzy Conventional p-value	0.000753	0.519	0.000168	0.719
Fuzzy CCT Corrected p-value	0.00142	0.469	0	0.217
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Finished Primary School	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.129	-0.0403	0.171	-0.0536
	(0.0353)***	(0.0454)	(0.0484)***	(0.0497)
Observations	14,277	8,630	10,175	7,997
Fuzzy Conventional p-value	0.000249	0.375	0.000419	0.280
Fuzzy CCT Corrected p-value	0.00374	0.0291	0.000358	0.241
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Earnings in Rupees	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	217.6	-401.5	306.4	-327.4
	(113.9)*	(187.8)**	(99.61)***	(176.2)*
Observations	10,175	7,997	14,277	8,630
Fuzzy Conventional p-value	0.0561	0.0325	0.00210	0.0632
Fuzzy CCT Corrected p-value	0.580	0.0138	0.706	0.000419
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Log Earnings	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.256	-0.217	0.326	-0.151
	(0.0829)***	(0.105)**	(0.0703)***	(0.0988)
Observations	10,175	7,994	14,277	8,627
Fuzzy Conventional p-value	0.00197	0.0389	0	0.126
Fuzzy CCT Corrected p-value	0.0806	0.00227	0.311	0
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10, for all districts, and all persons between the ages of 16 and 75 that reported earnings. The sample of 'Below 35 years' are of school going age during the policy, whereas those 'Above 35' are too old to change their schooling in response to the policy.

Bandwidths: 'CCT' is the Calonico et al. (2014b) method. 'I and K' is the Imbens and Kalyanaraman (2012) method. 'CCT corrected p-value' is the bias-corrected p-values using the method in Calonico et al. (2014b).

Table 1.5: Fraction of People that Have Finished At Least a Given Level of Education

Literate	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.0623	-0.0643	0.0655	-0.0300
	(0.0368)*	(0.0513)	(0.0289)**	(0.0383)
Observations	9,003	7,413	14,277	11,088
Fuzzy Conventional p-value	0.0906	0.210	0.0236	0.434
Fuzzy CCT Corrected p-value	0.0827	0.104	0.0417	0.00229
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Some pre-primary	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.0622	-0.0617	0.0657	-0.0181
	(0.0373)*	(0.0513)	(0.0293)**	(0.0363)
Observations	9,003	7,413	14,277	12,625
Fuzzy Conventional p-value	0.0956	0.229	0.0250	0.617
Fuzzy CCT Corrected p-value	0.0927	0.121	0.0220	9.12e-05
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Finished Primary	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.121	-0.0616	0.139	-0.0288
	(0.0445)***	(0.0491)	(0.0390)***	(0.0419)
Observations	9,273	7,869	11,972	9,920
Fuzzy Conventional p-value	0.00663	0.209	0.000354	0.493
Fuzzy CCT Corrected p-value	0.00747	0.117	0.000443	0.0815
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Finished Upper-primary	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.167	-0.0557	0.170	-0.0291
	(0.0518)***	(0.0509)	(0.0485)***	(0.0430)
Observations	9,045	7,729	10,175	9,920
Fuzzy Conventional p-value	0.00129	0.274	0.000443	0.499
Fuzzy CCT Corrected p-value	0.000798	0.230	0.000240	0.0250
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10 for persons between 16 and 75 years of age that reported earnings.
The sample of ‘Below 35 years’ are of school going age during the policy, whereas those ‘Above 35’ are too old to change their schooling in response to the policy.
Bandwidths: ‘CCT’ is the Calonico et al. (2014b) method. ‘I and K’ is the Imbens and Kalyanaraman (2012) method. ‘CCT corrected p-value’ is the bias-corrected p-values using the method in Calonico et al. (2014b)

Table 1.6: Returns, and Wage Parameters

	Fraction Switched	Change in Returns $\Delta\beta$	
Estimate	0.173	-0.065	
SE	(0.059)	(0.030)	

	Returns without GE $\beta_{y,D=0}$	Returns with GE $\beta_{y,D=1}$	% Change in returns
Estimate	0.199	0.134	-32.5%
Bootstrapped p-val	[0.055]	[0.098]	

	Change for older cohorts	Additional on Young	% Change on young
	-0.0053	-0.0594	91.87%

National Sample Survey 2009-10. The estimation follows the procedures described in the Model section 1.3, and detailed in Appendix 1.12.4, specifically Equations (1.23), (1.26) and (1.27).
Younger cohorts are those between 17 and 35, whereas older cohorts are between 36 and 50.
P-values for returns with GE $\beta_{y,D=1}$ and returns without GE $\beta_{y,D=0}$ were bootstrapped using 1000 draws of sampling with repetition. The null was created by jointly permutating the RD running variable, treatment status and probability of treatment.
The results in this table further suggest that the elasticity of substitution across age-cohorts is approximately $\sigma_A = 5$, and in the absence of movement in skill-biased capital the elasticity of substitution across skill groups would be $\sigma_E = 4.24$.

Table 1.7: Household Expenditures

	Log(Consumption Expenditure)			
	2004-5		2009-10	
RD Estimate	0.179	0.172	0.405	0.112
	(0.0372)***	(0.0334)***	(0.0682)***	(0.0322)***
Observations	27,372	33,758	12,563	26,420
Fuzzy Conventional p-value	0	0	0	0
Fuzzy CCT Corrected p-value	0	0	0	0
Bandwidth selection procedure	CCT	I and K	CCT	I and K
	Log(Total Educational Expenditure)			
	2004-5		2009-10	
RD Estimate	-0.217	-0.510	-0.191	-0.232
	(0.154)	(0.127)***	(0.135)	(0.118)**
Observations	8,922	11,388	8,205	9,937
Fuzzy Conventional p-value	0.159	0	0.157	0.0489
Fuzzy CCT Corrected p-value	0.0535	0	0.0668	0.0171
Bandwidth selection procedure	CCT	I and K	CCT	I and K
	Log(School Fees and Tuition)			
	2004-5		2009-10	
RD Estimate	-0.504	-0.977	-0.578	-0.616
	(0.204)**	(0.165)***	(0.186)***	(0.150)***
Observations	8,308	12,034	7,608	10,219
Fuzzy Conventional p-value	0.0136	0	0.0018	0
Fuzzy CCT Corrected p-value	0.0029	0	0.0005	0
Bandwidth selection procedure	CCT	I and K	CCT	I and K
	Log(Expenditure on newspapers, books, internet, libraries, stationery)			
	2004-5		2009-10	
RD Estimate	0.140	-0.0572	0.120	0.0189
	(0.121)	(0.101)	(0.0996)	(0.0914)
Observations	8,783	14,068	12,614	14,207
Fuzzy Conventional p-value	0.247	0.573	0.230	0.836
Fuzzy CCT Corrected p-value	0.0591	0.256	0.449	0.885
Bandwidth selection procedure	CCT	I and K	CCT	I and K

Household Expenditure Sources: National Sample Survey 2009-10.
Bandwidths: 'CCT' is the Calonico et al. (2014b) method. 'I and K' is the Imbens and Kalyanaraman (2012) method. 'CCT corrected p-value' is the bias-corrected p-values using the method in Calonico et al. (2014b)

Table 1.8: Labor Market Benefits

Change in Yearly Labor Market Benefits for			
(1) Young, Induced into getting more Skill			
With GE	Without GE	% Change	Fraction of cohort
0.121	0.157	-23.3%	0.17
(2) Always Skilled (Young)			
With GE	Without GE	% Change	Fraction of cohort
-0.037	0	-	0.39
(3) Always Unskilled (Young)			
With GE	Without GE	% Change	Fraction of cohort
0.014	0	-	0.44

Transfer in Yearly Benefits from Skilled to Unskilled			
Among Old with GE	Among Old without GE	Among Young with GE	Among Young without GE
0.004	0	0.051	0

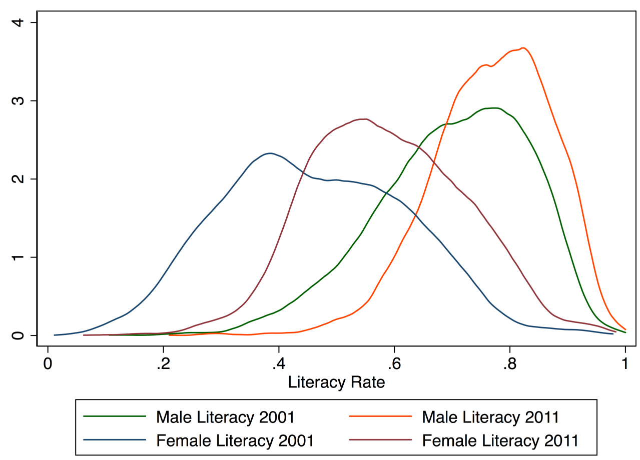
Change in Lifetime Welfare for Induced Students			
Costs	Benefits	Net	% Change (due to GE)
5.153	6.596	1.443	-23.3%

Welfare numbers are in monetary log-points. GE - indicates general equilibrium effects.
'Change in Benefits' shown for the sub-population that was young and changed their years of education to acquire skill. This is split up by 'With GE' effects, and a possible counterfactual of what would happen to their welfare in the absence of GE effects ('Without GE'). '% Change' is defined as change in welfare with the 'Without GE' as the base.
'Induced into getting more Skill' indicate the population that switched from unskilled to skilled only because of the policy. 'Always Skilled' indicate the population that would have acquired skill even in the absence of the policy. 'Always Unskilled' indicate the fraction of the population who would not have acquired skill even in the presence of the policy.
'Fraction switchers' is estimated (across RD cutoff) difference in sub-populations that acquired a higher level of education.
Yearly welfare calculations assume an interest rate of 2.37% and a gap of ten years between the costs of education and the labor market returns.
Real Interest Rates from the World Bank. The World Bank uses the lending rate and adjusts it for inflation using the GDP deflator. For the period 2010-13, the average real interest rate was 2.37%.
'Change in Lifetime Welfare for Induced Students' : Costs include (a) opportunity cost of foregone earnings for unskilled work, and (b) tuition costs for students in DPEP districts near the cutoff. Costs are calculated in 2004 (NSS 61st round).
'Change in Lifetime Welfare for Induced Students' : Benefits include present discounted value of lifetime earnings stream assuming a real interest rate of 2.37%.

1.11 Additional Tables and Figures

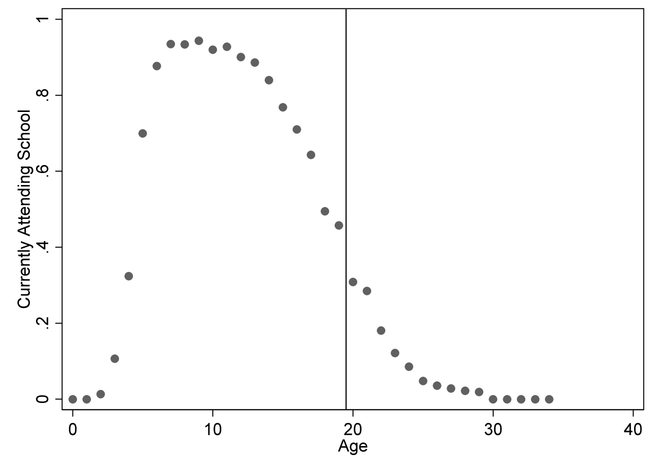
Figure 1.11: Literacy and Enrollment

(a) Literacy Rates Across Census Years



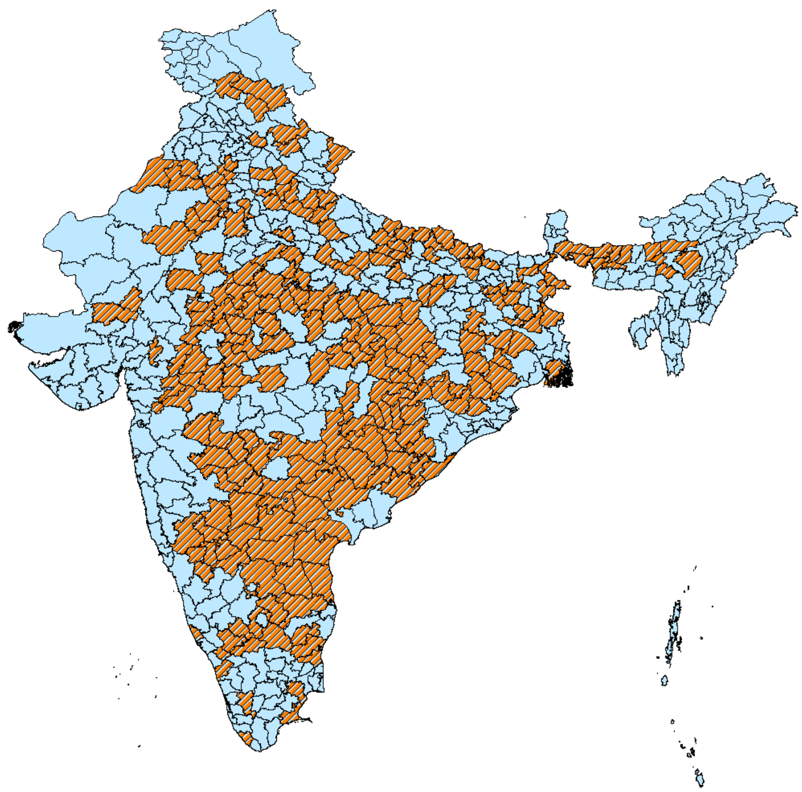
Source: Census of India 2001 and 2011.
Distributions calculated over sub-districts.

(b) Enrollment Rates by Age



National Sample Survey 2009. The largest drop in school enrollment occurs between the ages of 19 and 20 - representing a 15 percentage point fall.

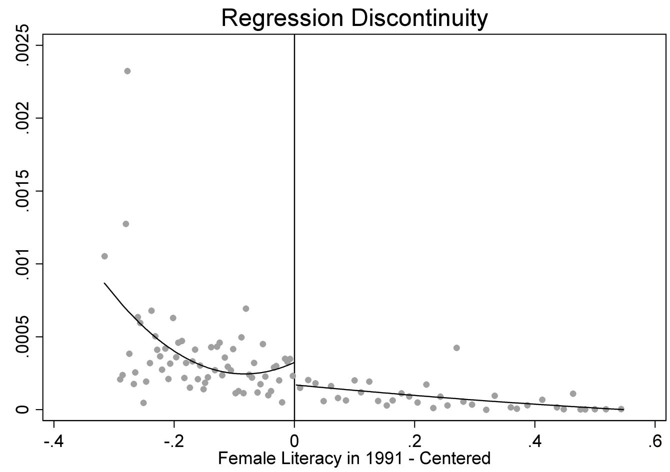
Figure 1.12: Map of DPEP Districts



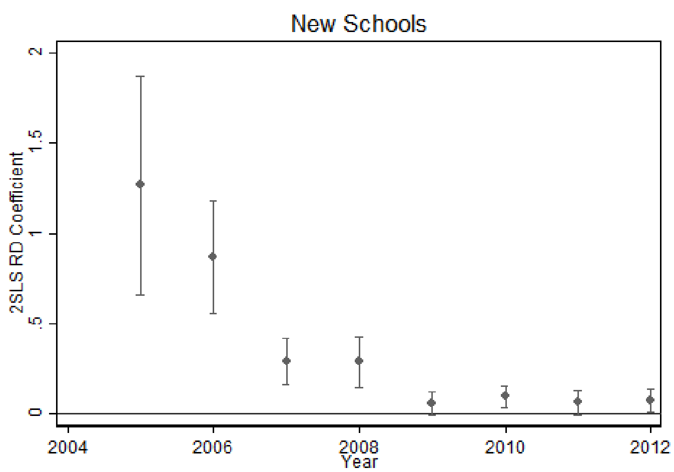
Orange and shaded districts received DPEP, whereas blue-unshaded districts did not.

Figure 1.13: Schools Built Post 1993 - Bandwidth Selection

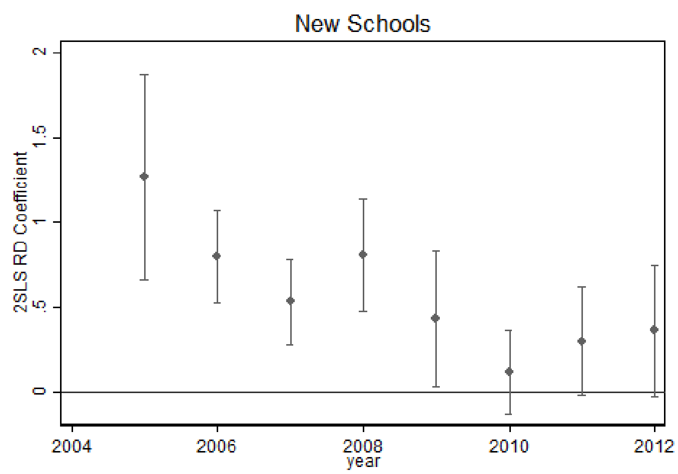
Total Schools (per cap) Built Post 1993



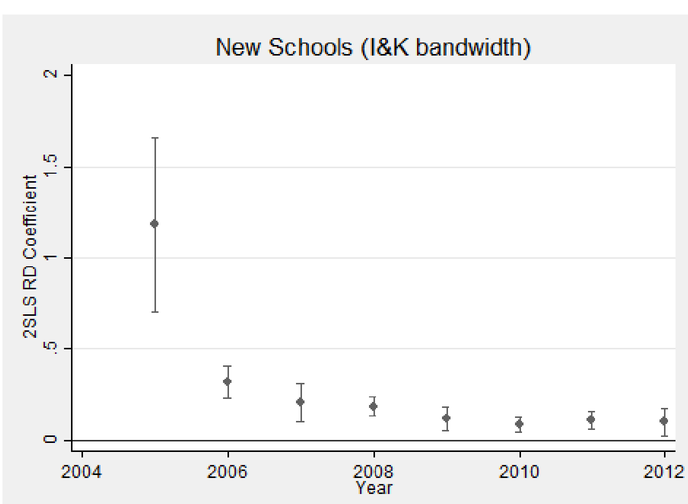
Total Government Schools (per cap) post 1993



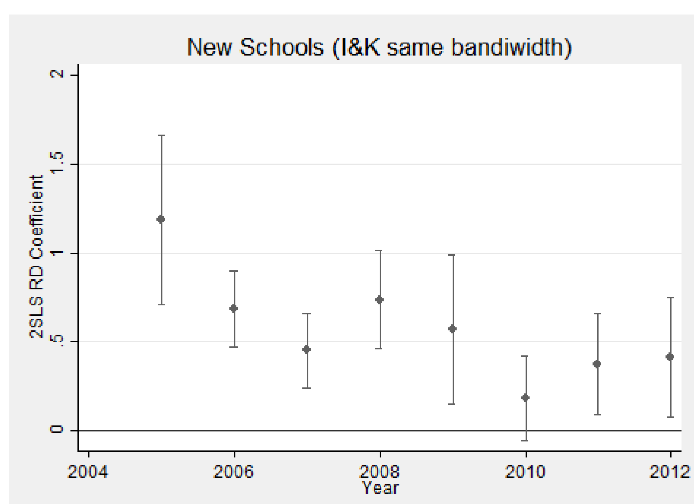
Fraction New - CCT



CCT Same Bandwidth for All Years



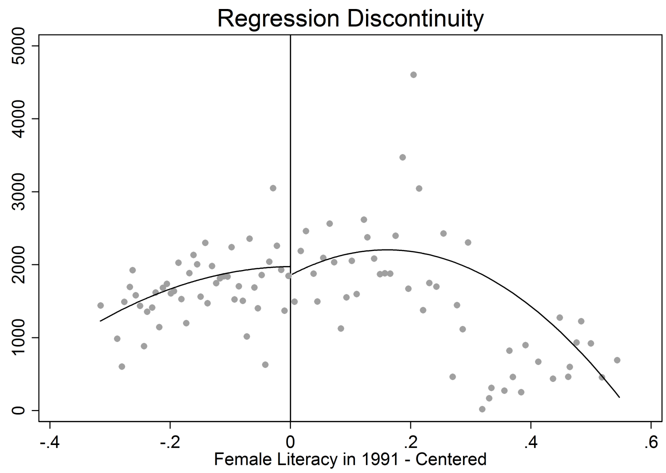
Fraction New - I&K Bandwidth



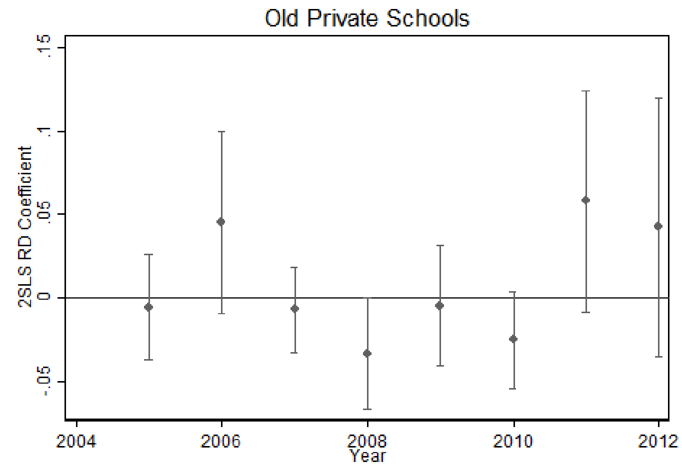
I&K Same Bandwidth for All Years

Source: DISE (District Information System for Education) data. CCT is Calonico et al. (2014b), whereas I&K is Imbens and Kalyanaraman (2012). 'per cap' figures normalized by total population in district. 'Same Bandwidth for All Years' is where the estimation is restricted to have the same bandwidth as it is in the first year of the data.

Figure 1.14: No Discontinuity in Number and Fraction of Old Schools

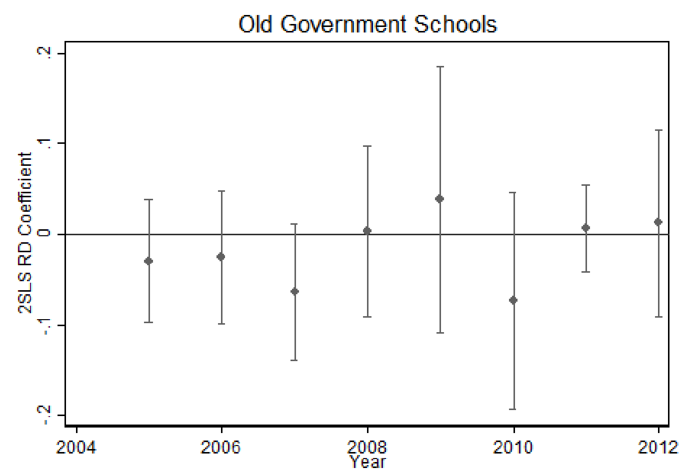


Total Number of Old Schools (built pre-1993)



Over Time: Private Schools (1973-93) as a Fraction of all Old Pvt Schools (pre-1993)

Total Number of Old Gov Schools (1993)

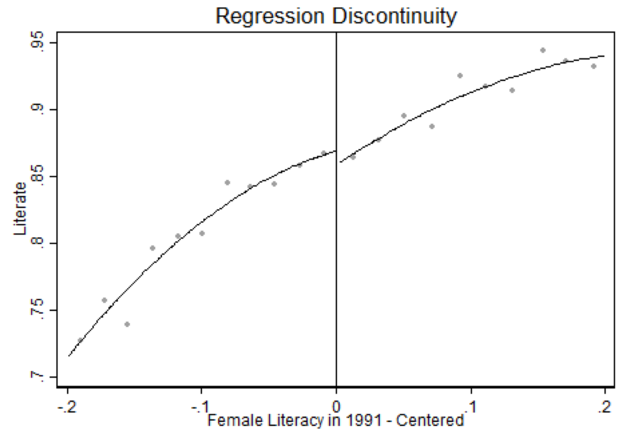


Over Time: Government Schools (1973-93) as a Fraction of all Old Gov Schools (pre-1993)

Source: DISE (District Information System for Education) data. RD graphs (Regression Function Fit) use the 2005 data. RD graph optimal binning and 2SLS RD coefficients calculated using Calonico et al. (2014b) procedure.

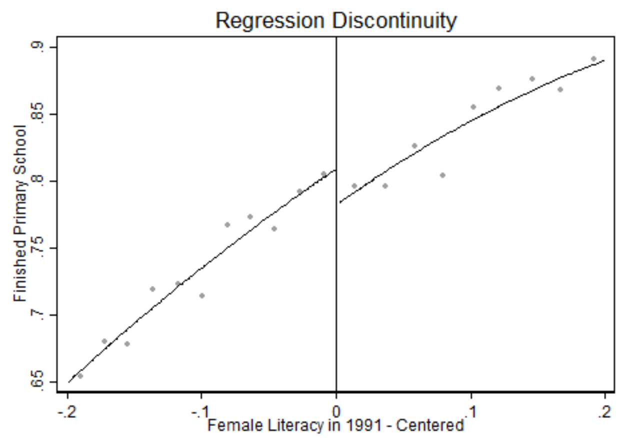
Figure 1.15: RD Figures - Levels of Education - Full Sample

Literate - Old



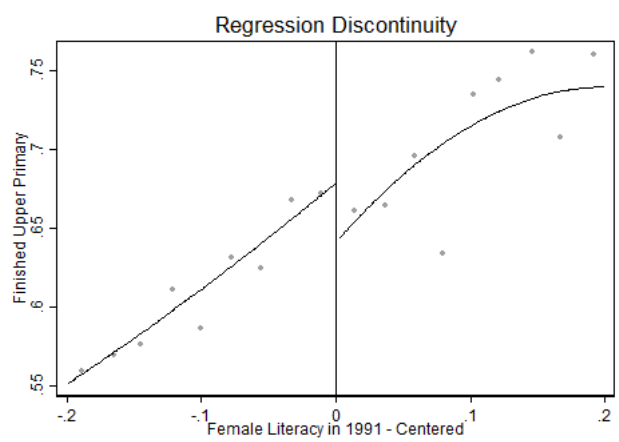
Literate - Young

Finished Primary School - Old



Finished Primary - Young

Years of Education - Old

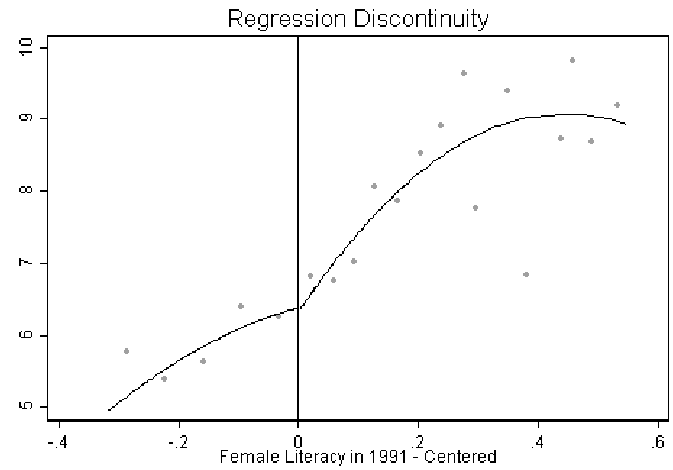


Finished Upper Primary - Young

National Sample Survey 2009 for all persons (whether or not earnings reported). Figures made using Calonico et al. (2014b) method of using regression curves to approximate the conditional means on either side of the cutoffs and the equally spaced sample means, and optimal number of bins.

Figure 1.16: RD figures for DISE districts

Years of Education - Younger

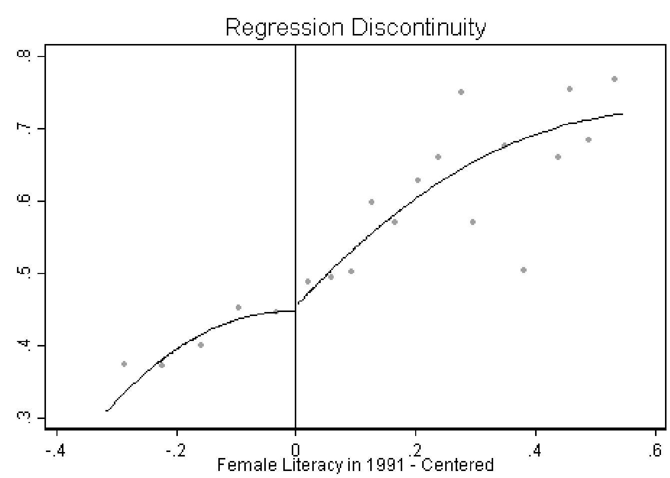


Years of Education - Older

Finished Primary - Younger

Finished Primary - Older

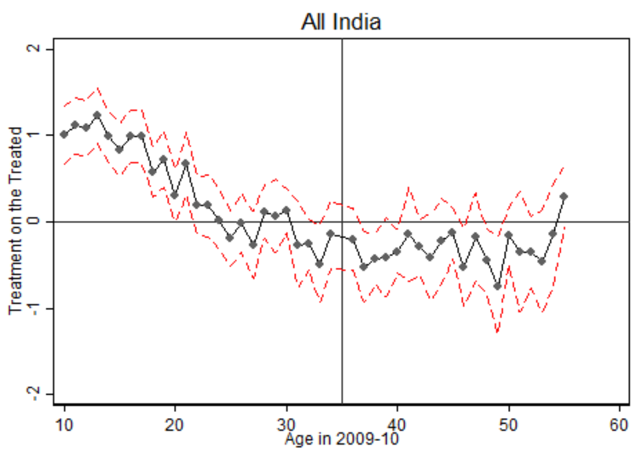
Finished Upper Primary - Younger



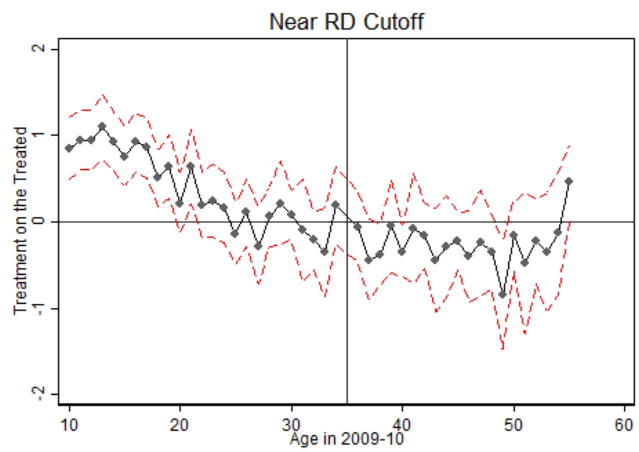
Finished Upper Primary - Older

National Sample Survey 2009. DISE districts include districts that have school-level DISE data. RD graph optimal binning and 2SLS RD coefficients calculated using Calonico et al. (2014b) procedure.

Figure 1.17: Difference-in-Differences: Years of Education



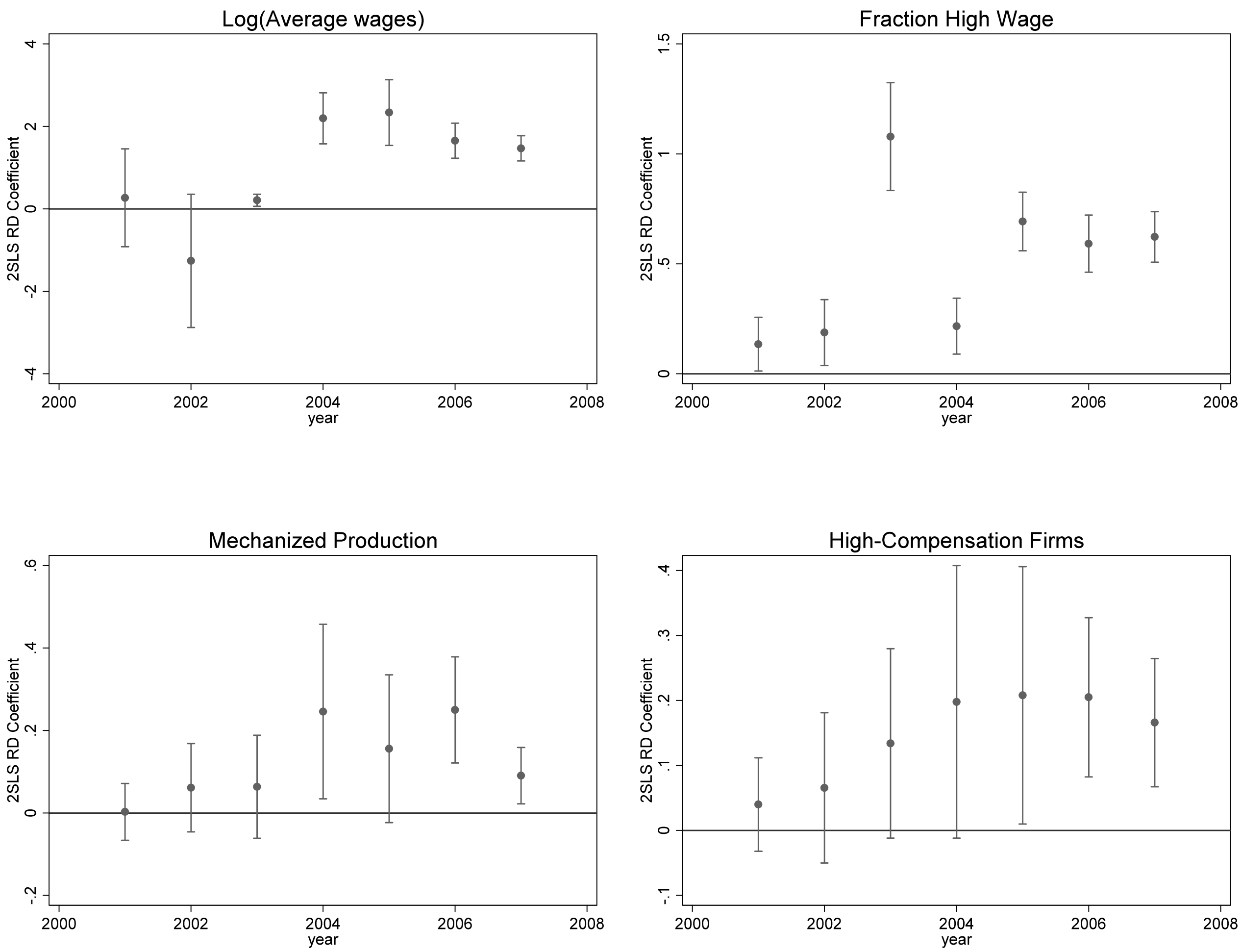
Full Sample



Short Bandwidth

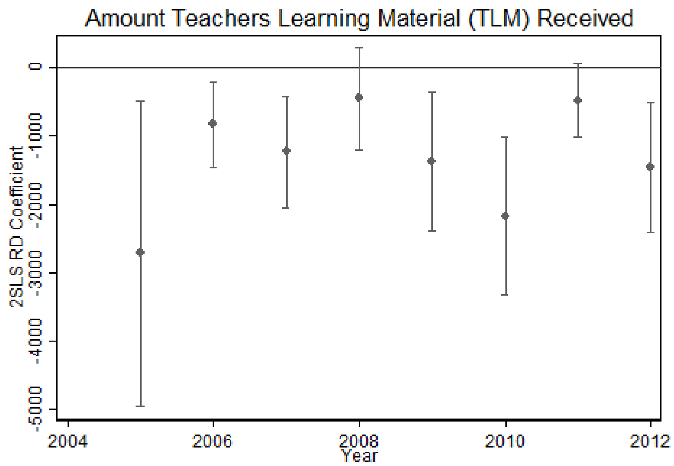
Coefficients of regression that includes age fixed effects and district fixed effects. Difference-in-Differences coefficient based on age and DPEP status. 'Short Bandwidth' restricts to sample near RD cutoff.

Figure 1.18: Firm Movement

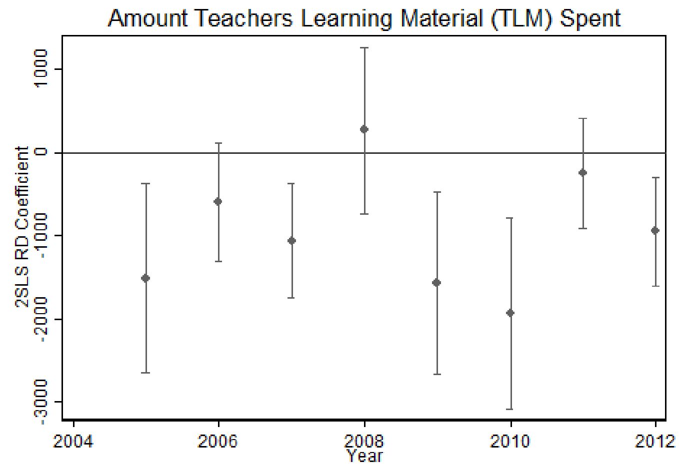


Source: Annual Survey of Industries (2001 to 2007). Firm level data. Wages and compensation calculated at the firm-level. 2SLS RD coefficients calculated using Calonico et al. (2014b) procedure. 'High-wage' or 'high-compensation' defined as being above median wages for the entire country.

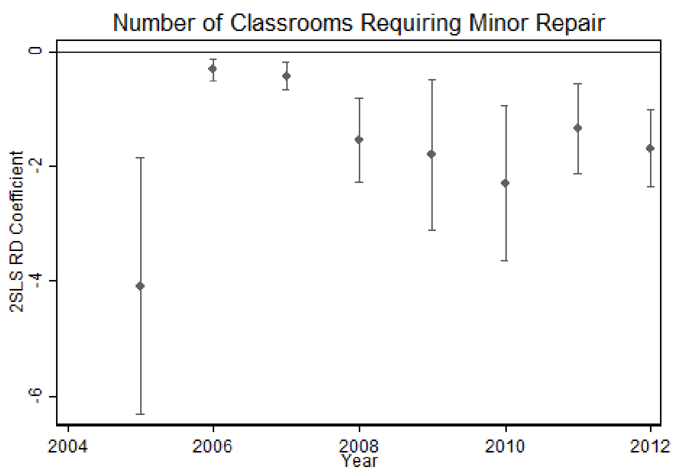
Figure 1.19: Other Funds Spent, and Condition of Rooms



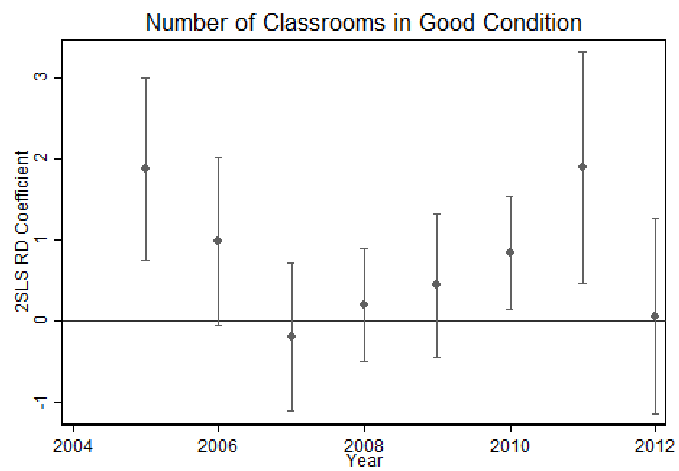
RD Coefficient Over Time: TLM grants Received



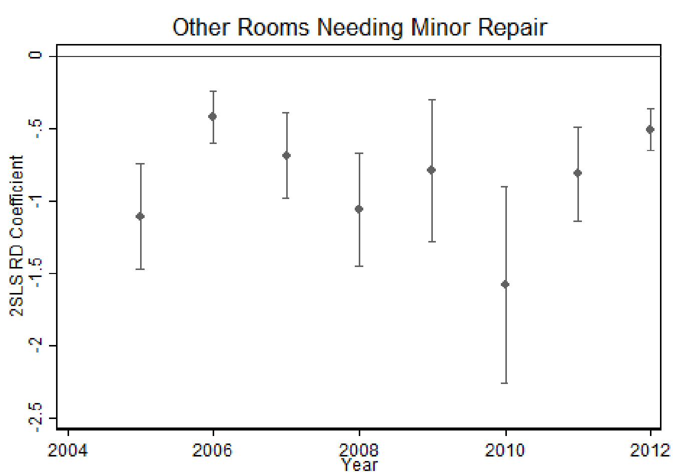
RD Coefficient Over Time: TLM grants Spent



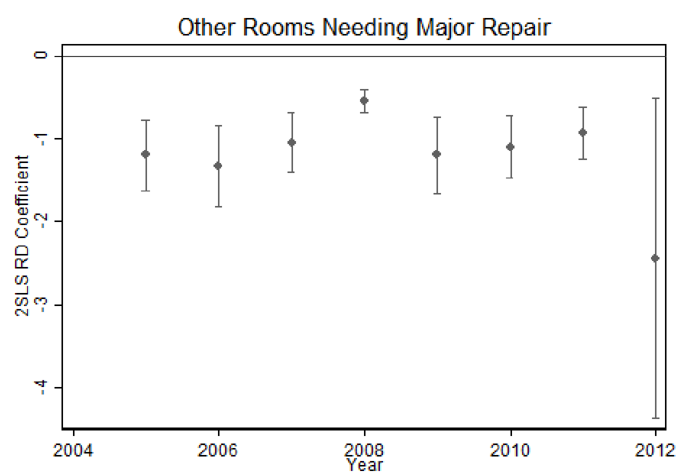
RD Coefficients: Classrooms Need Minor Repair



RD Coefficients: Classrooms in Good Condition



RD Coefficients: Other Rooms Need Minor Repair

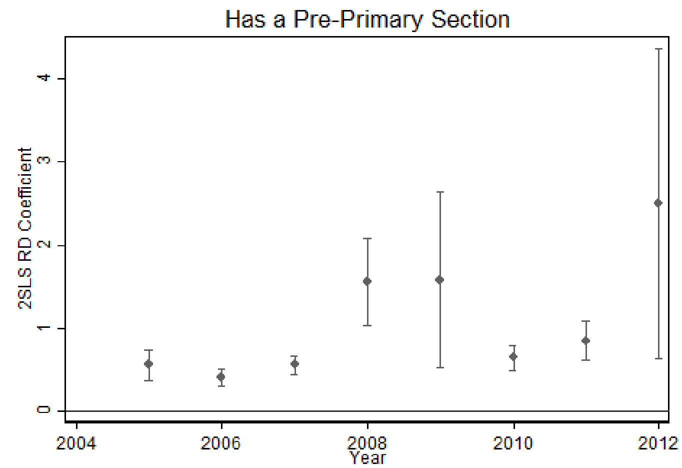


RD Coefficients: Other Rooms Need Major Repair

All schools, regardless of their which district they are in, are eligible to receive the Teacher Learning Materials (TLM) grant. Discontinuity for TLM grants spent, and for ‘Classrooms need Major Repair’ can be found in Figure 1.10.

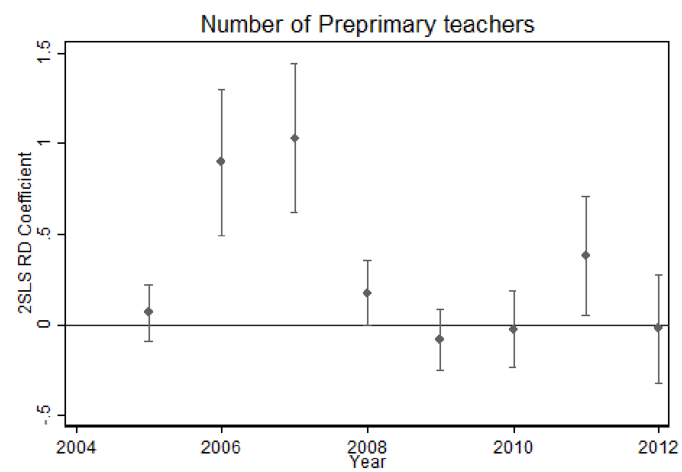
Figure 1.20: Pre-Primary Sections

Schools with Pre Primary Sections



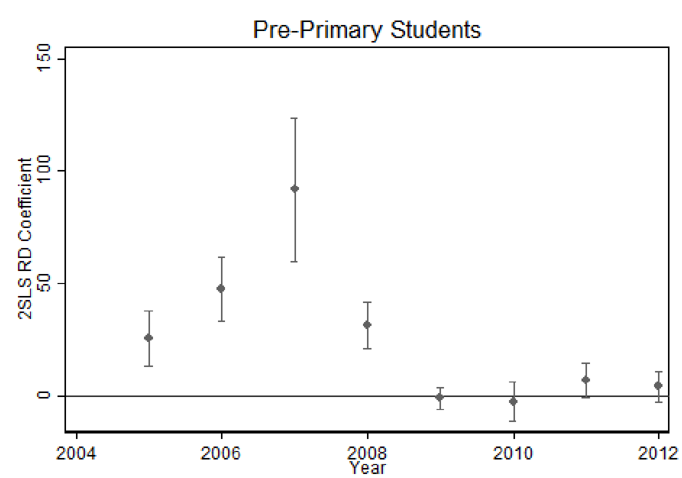
Coefficient Over Time: Pre-Primary Schools

Number of Pre Primary Teachers



Coefficient Over Time: Pre-Primary Teachers

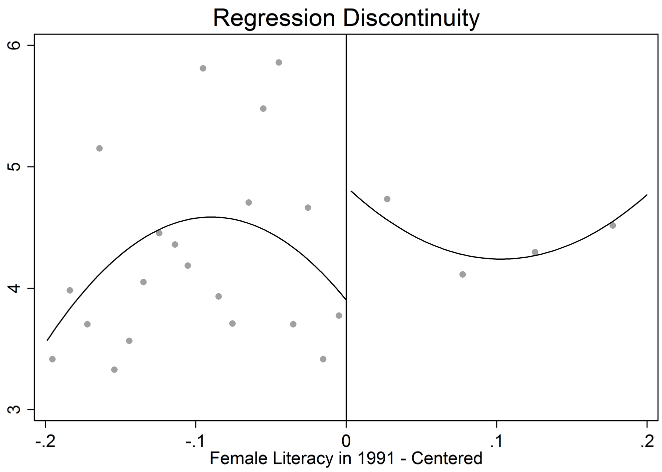
Number of Pre Primary Students



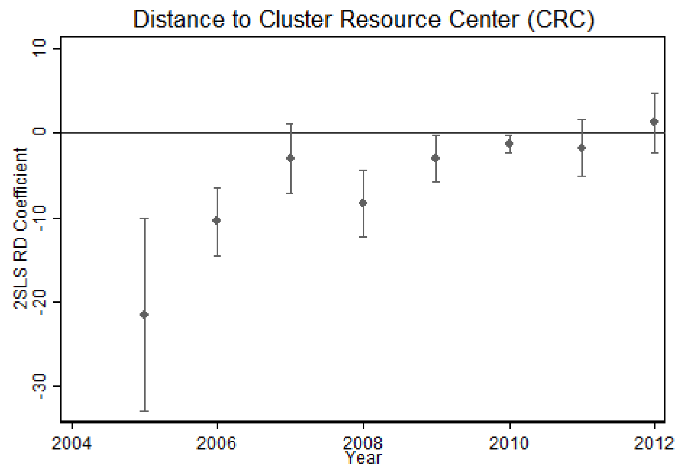
Coefficient Over Time: Pre-Primary Students

Source: DISE (District Information System for Education) data. RD graphs in the left-panel use the 2005 data. RD graph optimal binning and 2SLS RD coefficients calculated using Calonico et al. (2014b) procedure.

Figure 1.21: Academic Inspections and Regional Resource Centers

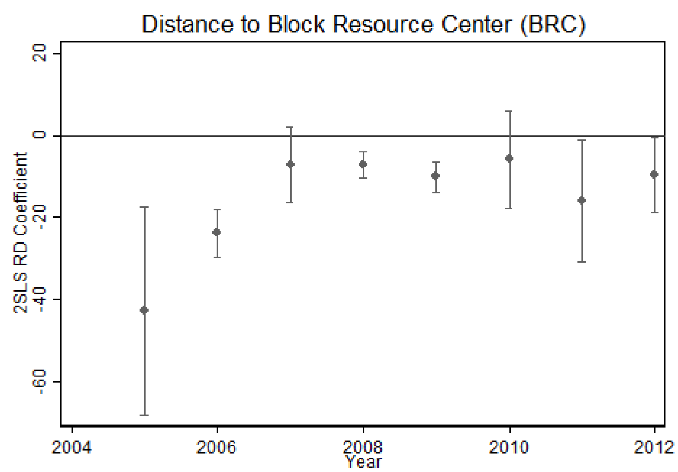


Distance to CRC (2005)

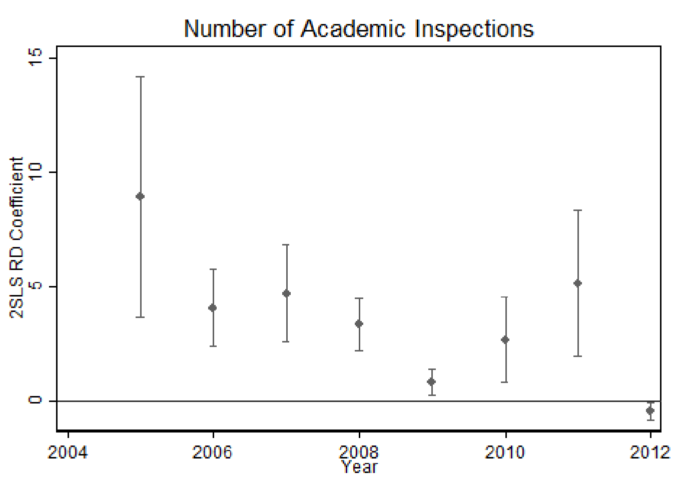


Coefficient Over Time: Distance to CRC

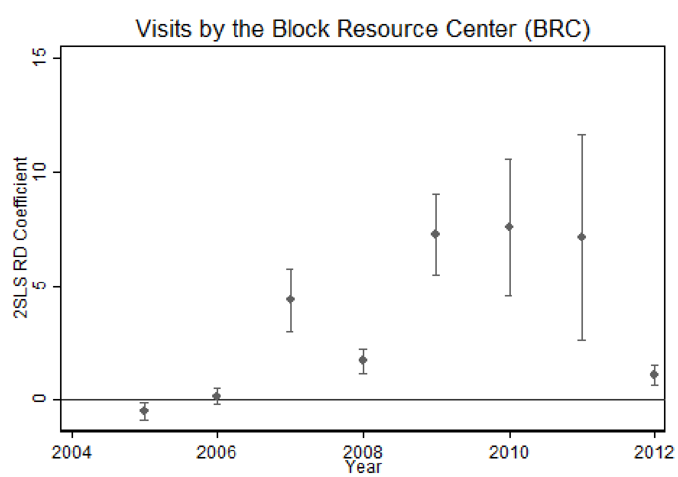
Distance to BRC (2005)



Coefficient Over Time: Distance to BRC



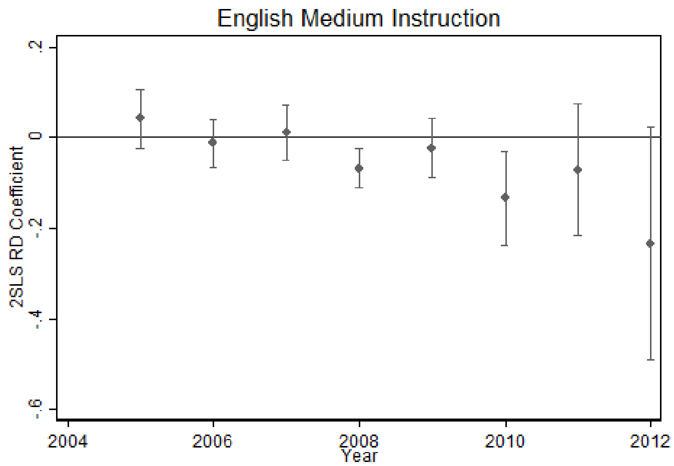
Number of Academic Inspections



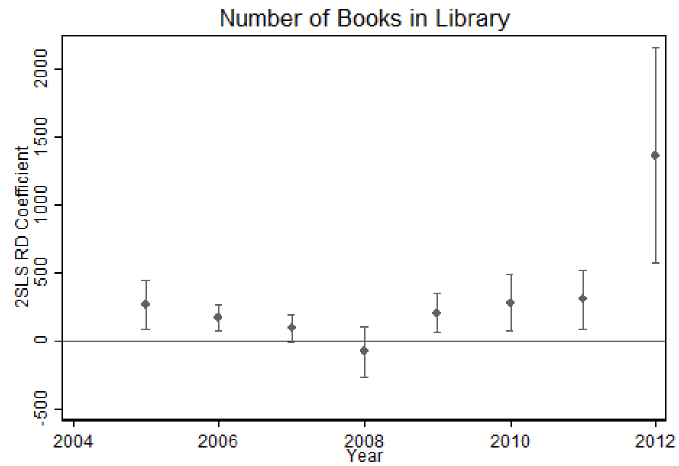
Visits by BRC Official

Source: DISE (District Information System for Education) data. Cluster Resource Centers (CRCs) and Block Resource Centers (BRCs) provide facilities and training to teachers. RD graphs in the left-panel use the 2005 data. RD graph optimal binning and 2SLS RD coefficients calculated using Calonico et al. (2014b) procedure.

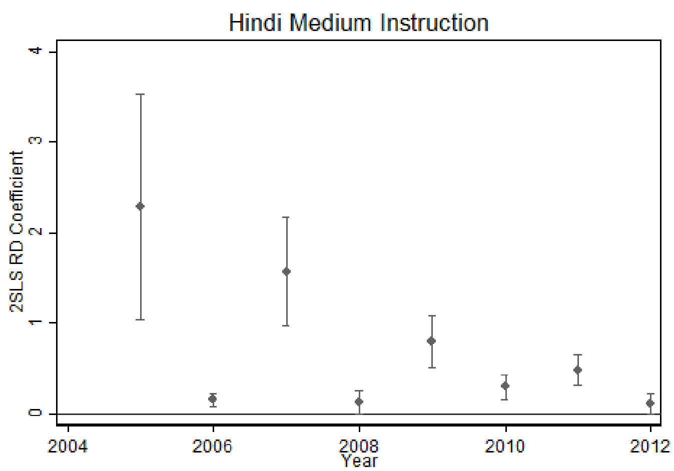
Figure 1.22: Medium of Instruction and Other Infrastructure



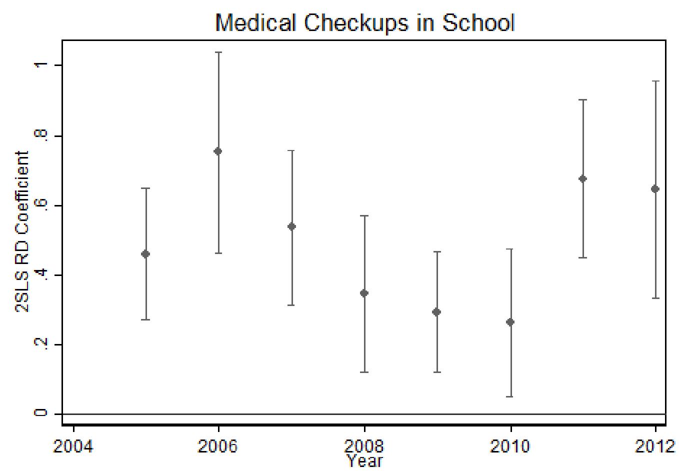
Coefficient Over Time: English Medium



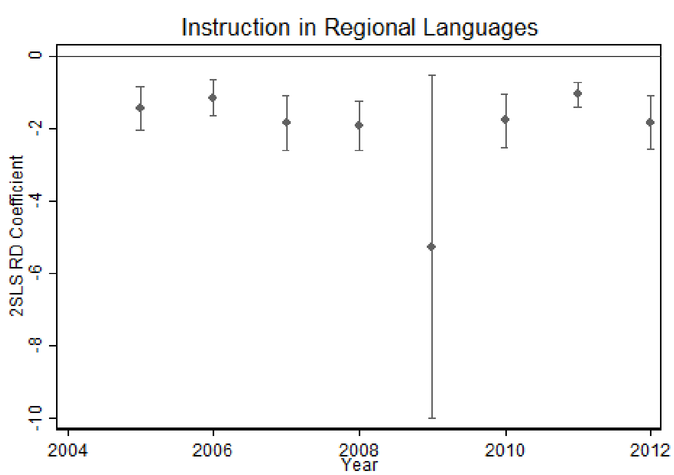
Coefficient Over Time: Library Books



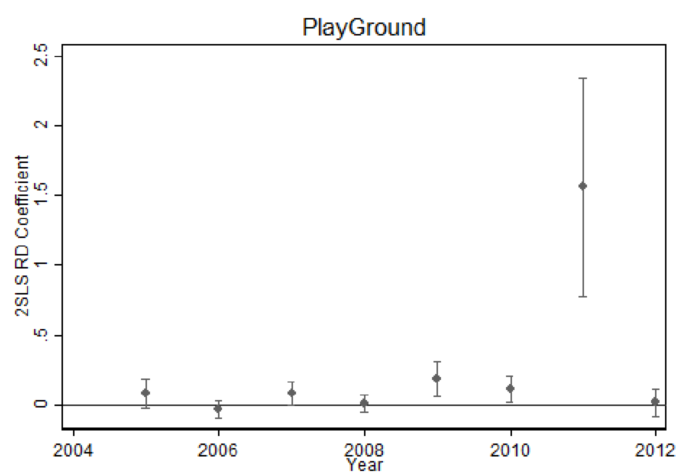
Coefficient Over Time: Hindi Medium



Medical Checkups



Coefficient Over Time: Regional Language

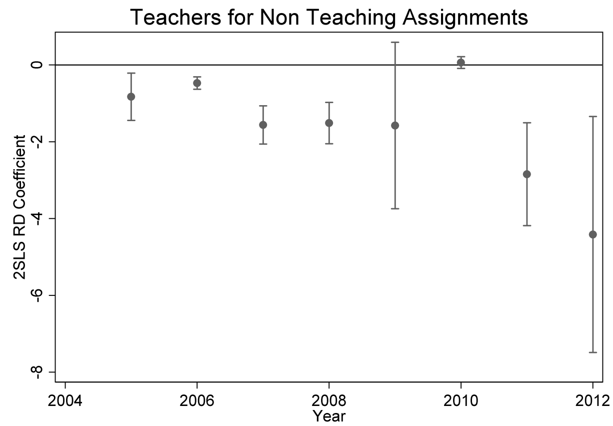


Coefficient Over Time: Playground

Source: DISE (District Information System for Education) data. RD graphs in the left-panel use the 2005 data. RD graph optimal binning and 2SLS RD coefficients calculated using Calonico et al. (2014b) procedure. Other infrastructure related graphs can be found in Figures 1.9, 1.20, 1.21 and 1.19.

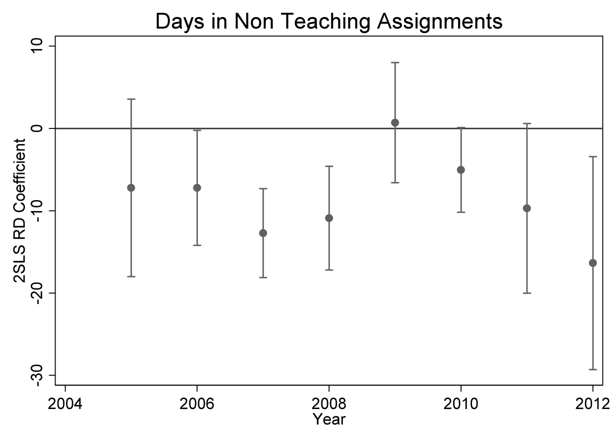
Figure 1.23: Involvement in Non Teaching Assignments

Teachers (per school) in Non-Teaching Assignments



Teachers (per school) in Non-Teaching Assignments

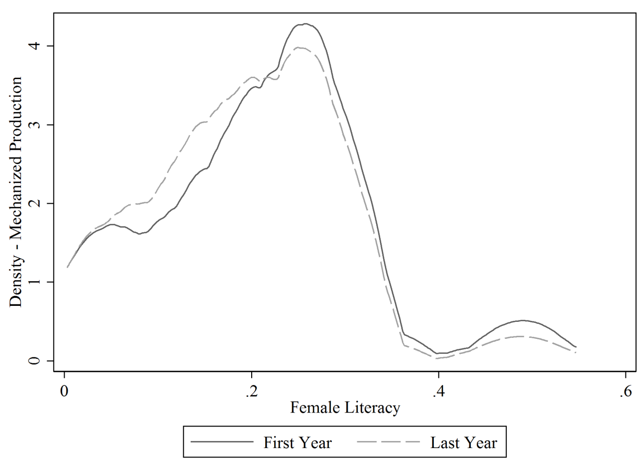
Days Involved in Non-Teaching Assignments



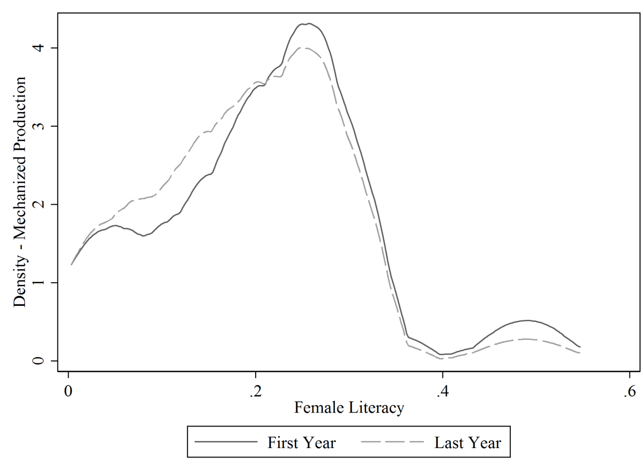
Days Involved in Non-Teaching Assignments

Source: DISE (District Information System for Education) data. RD graphs in the left-panel use the 2005 data. RD graph optimal binning and 2SLS RD coefficients calculated using Calonico et al. (2014b) procedure.

Figure 1.24: Density of Capital Intensive Firms Above Cutoff



Firms with Mechanized Production



High Compensation Firms

Source: Annual Survey of Industries (ASI) panel from 2001 (first year of data) and 2007 (last year of data).

Table 1.9: Education, Earnings and Returns By Age Groups

Years of Education - Younger	16 to 25	26 to 35	16 to 25	26 to 35
RD Estimate	2.751	1.161	2.379	1.179
	(0.768)***	(0.672)*	(0.559)***	(0.535)**
Observations	4,071	5,747	7,301	8,874
Fuzzy Conventional p-value	0.000340	0.0839	0	0.0277
Fuzzy CCT Corrected p-value	0.000170	0.108	0	0.0213
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Years of Education - Older	**36 to 45**	**46 to 55**	**36 to 45**	**46 to 55**
RD Estimate	-0.821	0.856	-0.450	0.649
	(0.787)	(1.008)	(0.684)	(0.850)
Observations	4,502	3,158	5,508	4,285
Fuzzy Conventional p-value	0.297	0.396	0.510	0.445
Fuzzy CCT Corrected p-value	0.180	0.257	0.171	0.198
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log(Earnings) - Younger	**16 to 25**	**26 to 35**	**16 to 25**	**26 to 35**
RD Estimate	0.403	0.136	0.481	0.265
	(0.134)***	(0.111)	(0.0973)***	(0.0884)***
Observations	4,072	5,747	7,302	8,874
Fuzzy Conventional p-value	0.00257	0.219	0	0.00270
Fuzzy CCT Corrected p-value	0.0109	0.844	0.00259	0.749
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log(Earnings) - Older	**36 to 45**	**46 to 55**	**36 to 45**	**46 to 55**
RD Estimate	-0.184	0.0350	-0.0585	0.192
	(0.134)	(0.182)	(0.117)	(0.157)
Observations	4,501	3,157	5,507	4,284
Fuzzy Conventional p-value	0.172	0.848	0.617	0.223
Fuzzy CCT Corrected p-value	0.0409	0.978	0.0697	0.432
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10, for all districts, and for persons that reported earnings. Bandwidths: Calonico et al. (2014b) method. Bias corrected p-value' is the bias-corrected p-values using the method in Calonico et al. (2014b). Earnings regressions are restricted to the same bandwidth as the years of education regressions.

Table 1.10: Earnings Reported, Migration, Paid Monthly, and Unemployment

P(Earnings Being Reported)	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	-0.0147	-0.0208	-0.0135	-0.0201
	(0.0209)	(0.0185)	(0.0223)	(0.0190)
Observations	37,201	42,316	32,742	39,823
Fuzzy Conventional p-value	0.481	0.261	0.546	0.289
Fuzzy CCT Corrected p-value	0.376	0.299	0.566	0.749
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Number of Migrants in District	**Total Migrants**		**Households Migrated**	
RD Estimate	10.93	4.230	-7.671	-1.863
	(38.95)	(36.95)	(4.590)*	(3.474)
Observations	153	277	175	523
Fuzzy Conventional p-value	0.779	0.909	0.0947	0.592
Fuzzy CCT Corrected p-value	0.786	0.853	0.0493	0
Bandwidth selection procedure	CCT	I and K	CCT	I and K
Paid monthly (non-daily)	**Below 35 years**	**Above 35 years**	**Below 35 years**	**Above 35 years**
RD Estimate	0.244	0.0450	0.239	0.0616
	(0.0581)***	(0.0526)	(0.0491)***	(0.0437)
Observations	7,962	7,680	10,395	9,869
Fuzzy Conventional p-value	0	0.393	0	0.159
Fuzzy CCT Corrected p-value	0	0.375	0	0.403
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Fraction Unemployed	**Below 35 years**	**Above 35 years**	**Below 35 years**	**Above 35 years**
RD Estimate	-0.0291	-0.00857	-0.0354	-0.00839
	(0.00527)***	(0.00379)**	(0.00616)***	(0.00343)**
Observations	82,936	38,060	62,393	50,887
Fuzzy Conventional p-value	0	0.0237	0	0.0143
Fuzzy CCT Corrected p-value	0	0.0105	0	0.00256
Bandwidth selection procedure	CCT	CCT	I and K	I and K

‘Number of Migrants in the District’ uses the small-sample National Sample Survey 2007-8 (64th Round) that asks questions on migration. ‘Household Migrated’ is indicator for whether the household every migrated for any reason. ‘Total Migrants’ counts the number of people who may have ever left the village for any reason - the most common reasons are marriage (54%). Less than 30% of migration is for work-related reasons.
The sample of ‘Below 35 years’ are of school going age during the policy, whereas those ‘Above 35’ are too old to change their schooling in response to the policy.
The other panels use National Sample Survey 2009-10. ‘P(Earnings Reported)’ is probability that earnings are reported - regresses indicator of whether earnings data is non-missing. ‘Paid-monthly’ is an indicator for whether the person receives earnings at a monthly (as opposed to daily) frequency. ‘Unemployed’ includes those who ‘sought-work’, those who ‘did not seek but were available for work’, did not work due to ‘sickness’ or ‘other reasons.’
Bandwidths: ‘CCT’ is the Calonico et al. (2014b) method. ‘I and K’ is the Imbens and Kalyanaraman (2012) method. ‘CCT corrected p-value’ is the bias-corrected p-values using the method in Calonico et al. (2014b).

Table 1.11: Education and Earnings - Men

Full sample Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.911	0.400	0.685	0.399
	(0.345)***	(0.285)	(0.245)***	(0.285)
Observations	16,197	29,622	34,248	29,622
Fuzzy Conventional p-value	0.00827	0.161	0.00521	0.161
Fuzzy CCT Corrected p-value	0.00285	0.183	0.000255	0.0711
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Reported Earnings Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	1.641	0.121	1.623	0.454
	(0.546)***	(0.615)	(0.501)***	(0.495)
Observations	8,047	6,767	9,638	12,517
Fuzzy Conventional p-value	0.00265	0.845	0.00119	0.359
Fuzzy CCT Corrected p-value	0.00485	0.992	0.00230	0.554
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Reported Earnings Finished Upper-Primary School	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.166	-0.00540	0.171	0.0465
	(0.0615)***	(0.0533)	(0.0509)***	(0.0412)
Observations	6,947	6,589	9,841	13,236
Fuzzy Conventional p-value	0.00697	0.919	0.000788	0.259
Fuzzy CCT Corrected p-value	0.00419	0.758	0.00520	0.661
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Reported Earnings Log Earnings	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.356	-0.0691	0.366	0.139
	(0.0921)***	(0.104)	(0.0836)***	(0.0825)*
Observations	8,047	6,766	9,638	12,516
Fuzzy Conventional p-value	0.000110	0.506	1.19e-05	0.0927
Fuzzy CCT Corrected p-value	0.00201	0.172	0.000377	6.60e-06
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10 for people between 16 and 75 years of age. Sample of males.
The sample of ‘Below 35 years’ are of school going age during the policy, whereas those ‘Above 35’ are too old to change their schooling in response to the policy.
Bandwidths: ‘CCT’ is the Calonico et al. (2014b) method. ‘I and K’ is the Imbens and Kalyanaraman (2012) method. ‘CCT corrected p-value’ is the bias-corrected p-values using the method in Calonico et al. (2014b).

Table 1.12: Education and Earnings - Women

Full sample Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.204	-0.0556	0.146	-0.0477
	(0.344)	(0.300)	(0.352)	(0.283)
Observations	17,244	16,834	16,486	19,809
Fuzzy Conventional p-value	0.553	0.853	0.678	0.866
Fuzzy CCT Corrected p-value	0.864	0.840	0.953	0.676
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Reported Earnings Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	1.616	-0.131	1.489	-0.159
	(1.099)	(0.967)	(0.904)*	(1.011)
Observations	2,213	2,128	2,945	2,026
Fuzzy Conventional p-value	0.141	0.892	0.0994	0.875
Fuzzy CCT Corrected p-value	0.127	0.736	0.0634	0.868
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Reported Earnings Finished Upper-Primary School	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.161	-0.0593	0.156	-0.0605
	(0.0818)**	(0.0757)	(0.0894)*	(0.0821)
Observations	2,620	2,157	2,250	1,998
Fuzzy Conventional p-value	0.0493	0.434	0.0801	0.461
Fuzzy CCT Corrected p-value	0.0486	0.365	0.0246	0.433
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Reported Earnings Log Earnings	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	-0.0910	-0.119	0.0684	-0.140
	(0.162)	(0.180)	(0.136)	(0.188)
Observations	2,213	2,126	2,945	2,024
Fuzzy Conventional p-value	0.575	0.509	0.615	0.457
Fuzzy CCT Corrected p-value	0.0761	0.187	0.0651	0.181
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10 for people between 16 and 75 years of age. Sample of females.
The sample of ‘Below 35 years’ are of school going age during the policy, whereas those ‘Above 35’ are too old to change their schooling in response to the policy.
Bandwidths: ‘CCT’ is the Calonico et al. (2014b) method. ‘I and K’ is the Imbens and Kalyanaraman (2012) method. ‘CCT corrected p-value’ is the bias-corrected p-values using the method in Calonico et al. (2014b).

Table 1.13: Difference-in-Differences Table

Full Sample Years of Education	Non DPEP	DPEP	Difference
Young	8.742	7.634	-1.108
	0.098	0.105	0.143
Old	6.255	4.758	-1.497
	0.118	0.096	0.152
Difference	2.487	2.876	0.389***
	0.071	0.074	0.102
Reported Earnings Years of Education	Non DPEP	DPEP	Difference
Young	8.57	7.20	-1.37
	0.14	0.15	0.20
Old	7.91	6.08	-1.83
	0.15	0.15	0.21
Difference	0.66	1.12	0.458**
	0.13	0.13	0.18
Log Earnings	Non DPEP	DPEP	Difference
Young	6.759	6.521	-0.238
	0.031	0.026	0.041
Old	7.102	6.800	-0.303
	0.031	0.026	0.040
Difference	-0.344	-0.279	0.065**
	0.023	0.021	0.031

National Sample Survey 2009-10 for people between 16 and 75 years of age.
The two dimensions for the Difference-in-Differences are district (received policy vs did not receive policy) and age (young enough to change schooling).
Table reports means for each sub-group and standard errors calculated at the district level below the means.

Table 1.14: District-Age Cells

Literate	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.0833	-0.0132	0.0808	-0.0139
	(0.0381)**	(0.0526)	(0.0414)*	(0.0567)
Observations	3,983	3,064	2,839	2,736
Fuzzy Conventional p-value	0.0289	0.802	0.0510	0.806
Fuzzy CCT Corrected p-value	0.0157	0.568	0.0944	0.718
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Primary School	**Below 35 years**	**Above 35 years**	**Below 35 years**	**Above 35 years**
RD Estimate	0.103	-0.0238	0.104	-0.0275
	(0.0453)**	(0.0557)	(0.0464)**	(0.0571)
Observations	3,892	3,064	3,432	2,899
Fuzzy Conventional p-value	0.0224	0.669	0.0249	0.630
Fuzzy CCT Corrected p-value	0.0224	0.464	0.922	0.798
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper-Primary School	**Below 35 years**	**Above 35 years**	**Below 35 years**	**Above 35 years**
RD Estimate	0.142	-0.0358	0.158	-0.0363
	(0.0557)**	(0.0562)	(0.0527)***	(0.0566)
Observations	3,908	3,080	4,798	3,057
Fuzzy Conventional p-value	0.0109	0.524	0.00278	0.522
Fuzzy CCT Corrected p-value	0.0146	0.464	0.677	0.663
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Years of Education	**Below 35 years**	**Above 35 years**	**Below 35 years**	**Above 35 years**
RD Estimate	0.975	-0.415	1.153	-0.430
	(0.598)	(0.673)	(0.461)**	(0.670)
Observations	3,526	3,182	6,470	3,296
Fuzzy Conventional p-value	0.103	0.538	0.0123	0.521
Fuzzy CCT Corrected p-value	0.178	0.537	0	0.831
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log Earnings	**Below 35 years**	**Above 35 years**	**Below 35 years**	**Above 35 years**
RD Estimate	0.0922	-0.0842	0.289	-0.0738
	(0.110)	(0.138)	(0.0860)***	(0.137)
Observations	3,526	3,182	6,470	3,296
Fuzzy Conventional p-value	0.400	0.540	0.000768	0.590
Fuzzy CCT Corrected p-value	0.822	0.257	0	0.344
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10. Data collapsed to the district-age cell level. Sample of persons that reported earnings, ages between 16 and 75 years.
The sample of 'Below 35 years' are of school going age during the policy, whereas those 'Above 35' are too old to change their schooling in response to the policy.
Bandwidths: 'CCT' is the Calonico et al. (2014b) method. 'I and K' is the Imbens and Kalyanaraman (2012) method. 'CCT corrected p-value' is the bias-corrected p-values using the method in Calonico et al. (2014b).

Table 1.15: Robustness: In-Progress RD Methods for Bandwidths and Standard Errors

Panel A: Bartalotti and Brummet (2015) Standard Errors				
Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	1.654	-0.337	1.569	-0.0985
	(0.742)**	(0.877)	(0.650)**	(0.836)
Bandwidth	CCT	CCT	I and K	I and K
Finished Primary School	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.121	-0.0580	0.139	-0.0266
	(0.0615)**	(0.0793)	(0.0568)**	(0.0669)
Bandwidth	CCT	CCT	I and K	I and K
Finished Upper-Primary	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.167	-0.0507	0.170	-0.0291
	(0.0766)**	(0.0659)	(0.0730)**	(0.0610)
Bandwidth	CCT	CCT	I and K	I and K

Panel B: Two-sided Bandwidth with district-age group Standard Errors				
	Years of Education		Finished Primary	
	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	1.188	-0.0516	0.115	-0.0350
	(0.632)*	(0.706)	(0.0526)**	(0.0605)
	Finished Upper Primary		Finished Secondary	
	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.174	-0.0512	0.169	-0.0182
	(0.0631)***	(0.0518)	(0.0797)**	(0.0659)

National Sample Survey 2009-10. Sample of persons that reported earnings, ages between 16 and 75 years.
The sample of ‘Below 35 years’ are of school going age during the policy, whereas those ‘Above 35’ are too old to change their schooling in response to the policy.
Bandwidths: ‘CCT’ is the Calonico et al. (2014b) method. ‘I and K’ is the Imbens and Kalyanaraman (2012) method. ‘CCT corrected p-value’ is the bias-corrected p-values using the method in Calonico et al. (2014b)
Panel A: Uses the Bartalotti and Brummet (2015) method to compute standard errors at the district-age group level. The optimal bandwidths are chosen using the Calonico et al. (2014b) and Imbens and Kalyanaraman (2012) methods. I thank
Panel B: Uses an in-progress method developed by some authors of the Calonico et al. (2014b) paper that allows for a separate optimal bandwidth on either side of the cutoff and allows for computing standard errors at the district-age group level.

Table 1.16: Parametric RD - Short Bandwidth

Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.999***	0.769	0.911**	0.782
	(0.387)	(0.487)	(0.383)	(0.488)
Observations	128,799	124,077	128,799	124,077
R-squared	0.039	0.022	0.030	0.024
Control Function	Linear	Linear	Quadratic	Quadratic
Reported Earnings Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	1.406***	1.194	1.416***	1.206
	(0.498)	(0.892)	(0.493)	(0.892)
Observations	26,898	29,343	26,898	29,343
R-squared	0.026	0	0.028	0
Control Function	Linear	Linear	Quadratic	Quadratic
Reported Earnings Finished Upper Primary	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.107**	0.0912	0.108**	0.0916
	(0.0458)	(0.0633)	(0.0455)	(0.0632)
Observations	26,899	29,346	26,899	29,346
R-squared	0.022	0	0.023	-0.006
Control Function	Linear	Linear	Quadratic	Quadratic
Log Wage & Salary Earnings	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.470***	0.358**	0.544***	0.413**
	(0.138)	(0.174)	(0.126)	(0.170)
Observations	26,894	29,342	26,894	29,342
R-squared	0	0	0	0
Control Function	Linear	Linear	Quadratic	Quadratic

National Sample Survey 2009-10. Parametric RDs using local linear and quadratic functions. Bandwidth restricted to 0.3 on either side of the cutoff. Sample of persons between 16 and 75 years.
The sample of ‘Below 35 years’ are of school going age during the policy, whereas those ‘Above 35’ are too old to change their schooling in response to the policy.

Table 1.17: Parametric RD - Longer Bandwidth

Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.665*	0.632	0.781**	0.662
	(0.345)	(0.401)	(0.352)	(0.427)
Observations	133,669	129,192	133,669	129,192
R-squared	0.043	0.035	0.035	0.030
Control Function	Linear	Linear	Quadratic	Quadratic

Reported Earnings Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	1.035**	0.913	1.097**	0.992
	(0.416)	(0.716)	(0.449)	(0.789)
Observations	28,290	30,836	28,290	30,836
R-squared	0.044	0.011	0.032	-0.004
Control Function	Linear	Linear	Quadratic	Quadratic

Reported Earnings Finished Upper Primary	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.0832**	0.0711	0.0865**	0.0766
	(0.0398)	(0.0511)	(0.0418)	(0.0561)
Observations	28,291	30,839	28,291	30,839
R-squared	0.036	0.011	0.028	-0.002
Control Function	Linear	Linear	Quadratic	Quadratic

Log Wage & Salary Earnings	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.427***	0.332**	0.399***	0.305**
	(0.117)	(0.145)	(0.113)	(0.147)
Observations	28,285	30,835	28,285	30,835
R-squared	0	0	0	0
Control Function	Linear	Linear	Quadratic	Quadratic

National Sample Survey 2009-10. Parametric RDs using local linear and quadratic functions. Bandwidth restricted to 0.4 on either side of the cutoff. Sample of persons between 16 and 75 years.
The sample of 'Below 35 years' are of school going age during the policy, whereas those 'Above 35' are too old to change their schooling in response to the policy.

Table 1.18: Returns to Education using Two-Staged Least Squares

First-Stage Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	1.654	-0.381	1.569	-0.199
	(0.491)***	(0.590)	(0.417)***	(0.553)
Observations	10,175	7,997	14,277	8,630
Fuzzy Conventional p-value	0.000753	0.519	0.000168	0.719
Fuzzy CCT Corrected p-value	0.00142	0.469	0	0.217
Bandwidth selection procedure	CCT	CCT	I and K	I and K

2 SLS Log(Earnings)	Below 35 years	Above 35 years	Below 35 years	Above 35 years
Years of Education	0.155	0.567	0.208	0.744
	(0.0465)***	(0.699)	(0.0494)***	(1.706)
Observations	10,175	7,994	14,277	8,627
Fuzzy Conventional p-value	0.000856	0.417	0	0.663
Fuzzy CCT Corrected p-value	0.0394	0.269	0.569	0.813
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10 for persons between 16 and 75 years of age. '2SLS' regressions treats the first stage as 'change in years of education' as opposed to probability of receiving DPEP funds.
The sample of 'Below 35 years' are of school going age during the policy, whereas those 'Above 35' are too old to change their schooling in response to the policy.
Bandwidths: 'CCT' is the Calonico et al. (2014b) method. 'I and K' is the Imbens and Kalyanaraman (2012) method. 'CCT corrected p-value' is the bias-corrected p-values using the method in Calonico et al. (2014b)

Table 1.19: Robustness: Widening Age Restrictions - Full Sample

Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.605	0.209	0.600	0.287
	(0.166)***	(0.237)	(0.176)***	(0.225)
Observations	74,342	35,064	63,388	39,456
Fuzzy Conventional p-value	0.000266	0.378	0.000641	0.202
Fuzzy CCT Corrected p-value	0	0.262	0	0.165
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Finished Primary School	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.0288	0.0218	0.0429	0.0243
	(0.0176)	(0.0222)	(0.0147)***	(0.0216)
Observations	42,713	37,199	66,472	39,839
Fuzzy Conventional p-value	0.102	0.327	0.00346	0.259
Fuzzy CCT Corrected p-value	0.219	0.457	0.0599	0.456
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Finished Upper Primary School	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.0729	0.0276	0.0765	0.0299
	(0.0216)***	(0.0216)	(0.0178)***	(0.0185)
Observations	42,713	36,145	70,270	57,738
Fuzzy Conventional p-value	0.000754	0.201	1.81e-05	0.106
Fuzzy CCT Corrected p-value	0.00188	0.277	8.64e-09	0.344
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10 for sample of persons aged 15 to 100 years of age.
The sample of 'Below 35 years' are of school going age during the policy, whereas those 'Above 35' are too old to change their schooling in response to the policy.
Bandwidths: 'CCT' is the Calonico et al. (2014b) method. 'I and K' is the Imbens and Kalyanaraman (2012) method. 'CCT corrected p-value' is the bias-corrected p-values using the method in Calonico et al. (2014b)

Table 1.20: Robustness: Widening Age Restrictions - For Reported Earnings

Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	1.733	-0.374	1.674	-0.113
	(0.487)***	(0.589)	(0.438)***	(0.534)
Observations	10,559	8,002	12,866	9,057
Fuzzy Conventional p-value	0.000372	0.525	0.000132	0.832
Fuzzy CCT Corrected p-value	0.000668	0.477	0.000359	0.186
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Primary School	**Below 35 years**	**Above 35 years**	**Below 35 years**	**Above 35 years**
RD Estimate	0.136	-0.0607	0.149	-0.0272
	(0.0443)***	(0.0488)	(0.0422)***	(0.0403)
Observations	9,822	8,002	10,560	11,033
Fuzzy Conventional p-value	0.00210	0.214	0.000412	0.500
Fuzzy CCT Corrected p-value	0.00178	0.121	0.000143	0.0543
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Upper Primary	**Below 35 years**	**Above 35 years**	**Below 35 years**	**Above 35 years**
RD Estimate	0.173	-0.0549	0.172	-0.0334
	(0.0514)***	(0.0506)	(0.0503)***	(0.0381)
Observations	9,662	7,734	10,117	13,441
Fuzzy Conventional p-value	0.000773	0.278	0.000637	0.380
Fuzzy CCT Corrected p-value	0.000549	0.236	0.000545	0.00170
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log(Earnings)	**Below 35 years**	**Above 35 years**	**Below 35 years**	**Above 35 years**
RD Estimate	0.272	-0.218	0.298	-0.112
	(0.0834)***	(0.105)**	(0.0747)***	(0.0956)
Observations	10,560	7,999	12,867	9,054
Fuzzy Conventional p-value	0.00109	0.0380	6.57e-05	0.242
Fuzzy CCT Corrected p-value	0.0508	0.00217	0.839	4.39e-06
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10 for sample of persons aged 15 to 100 years of age that reported earnings.
The sample of ‘Below 35 years’ are of school going age during the policy, whereas those ‘Above 35’ are too old to change their schooling in response to the policy.
Bandwidths: ‘CCT’ is the Calonico et al. (2014b) method. ‘I and K’ is the Imbens and Kalyanaraman (2012) method. ‘CCT corrected p-value’ is the bias-corrected p-values using the method in Calonico et al. (2014b)

Table 1.21: Robustness: Restricting to DISE districts

Years of Education	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	0.978	-0.0310	0.770	-0.105
	(0.269)***	(0.202)	(0.216)***	(0.177)
Observations	21,099	34,331	31,727	46,462
Fuzzy Conventional p-value	0.000275	0.878	0.000356	0.552
Fuzzy CCT Correct p-value	0	0.888	0	0.308
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Primary	**Below 35 years**	**Above 35 years**	**Below 35 years**	**Above 35 years**
RD Estimate	0.0801	-0.00677	0.0647	-0.00571
	(0.0221)***	(0.0172)	(0.0168)***	(0.0186)
Observations	21,258	46,012	35,465	37,713
Fuzzy Conventional p-value	0.000280	0.694	0.000111	0.759
Fuzzy CCT Correct p-value	0	0.387	0	0.707
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper Primary	**Below 35 years**	**Above 35 years**	**Below 35 years**	**Above 35 years**
RD Estimate	0.144	0.00521	0.126	0.00343
	(0.0314)***	(0.0180)	(0.0250)***	(0.0175)
Observations	18,169	37,713	22,612	41,715
Fuzzy Conventional p-value	0	0.772	0	0.845
Fuzzy CCT Correct p-value	0	0.867	0	0.450
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10.
The sample of ‘Below 35 years’ are of school going age during the policy, whereas those ‘Above 35’ are too old to change their schooling in response to the policy.
Bandwidths: ‘CCT’ is the Calonico et al. (2014b) method. ‘I and K’ is the Imbens and Kalyanaraman (2012) method. ‘CCT corrected p-value’ is the bias-corrected p-values using the method in Calonico et al. (2014b)

Table 1.22: Difference-in-Differences (Full Model)

Full Sample	Years of Education	Literate	Finished Primary	Finished Upper Primary
Estimate	0.332***	0.0551***	0.0386***	0.0196***
	(0.0388)	(0.00311)	(0.00338)	(0.00363)
Observations	279,452	279,483	279,483	279,483
R-squared	0.176	0.189	0.193	0.170
Small Bandwidth	Years of Education	Literate	Finished Primary	Finished Upper Primary
Estimate	0.311***	0.0426***	0.0302***	0.0209**
	(0.106)	(0.00764)	(0.00834)	(0.00959)
Observations	144,248	144,261	144,261	144,261
R-squared	0.108	0.118	0.117	0.103
Reported Earnings	Years of Education	Literate	Finished Primary	Finished Upper Primary
Estimate	0.377**	0.0558***	0.0410***	0.0299**
	(0.155)	(0.0111)	(0.0119)	(0.0150)
Observations	66,093	66,098	66,098	66,098
R-squared	0.157	0.166	0.164	0.139
	Log(Earnings)		2SLS Returns	
Estimate	0.0596**		0.159***	
	(0.0251)		(0.0473)	
Observations	66,086		66,081	
R-squared	0.241		0.393	
	Log (Earnings) Skilled	Log (Earnings) Unskilled	Additional GE on young	
Estimate	-0.0611**	0.0183	-0.0794**	
	(0.0283)	(0.0213)	(0.0354)	
Observations	37,748	28,338		
R-squared	0.311	0.225		

National Sample Survey 2009-10 – 17 to 75 year olds. Regressions include district and cohort fixed effects. Diff-in-diff coefficient on interaction between being below 35 and in DPEP district. Robust standard errors at the district level.
'Small Bandwidth' restricts the sample in two ways: (1) restricts ages to be $+/-$ 15 years of the 35 year cutoff, (2) restricts districts to have female literacy $\in (-0.2, 0.2)$. '2SLS Returns' estimates two-staged least squares returns where the first stage dependent variable is the years of education, and the second stage dependent variable is log-earnings. 'Additional GE on young' estimates the GE effect that only affects the skill-premium of the young (note: this excludes the average change in wages due to changes in output, and the portion of the change in the skill premium experienced by all-cohorts).

Table 1.23: Test Scores

Panel A: Reading Scores 2008	Read Letter	Read Word	Reading Level 1
RD Estimate	0.00411	-0.0158	-0.0147
	(0.0107)	(0.0118)	(0.0120)
Bandwidth	CCT	CCT	CCT
Panel B: Math Scores 2008	**Numbers 1-9**	**Numbers 10-99**	**Subtraction**
RD Estimate	0.0531	0.0197	0.0196
	(0.0116)***	(0.0136)	(0.0137)
Bandwidth	CCT	CCT	CCT
Panel C: Reading Scores 2012	**Read Letter**	**Read Word**	**Reading Level 1**
RD Estimate	-0.0143	0.0164	0.0216
	(0.0148)	(0.0141)	(0.0145)
Bandwidth	CCT	CCT	CCT
Panel D: Math Scores 2012	**Numbers 1-9**	**Numbers 10-99**	**Subtraction**
RD Estimate	0.0514	-0.0277	0.0351
	(0.0156)***	(0.0184)	(0.0183)*
Bandwidth	CCT	CCT	CCT

Source: Annual Status of Education Report (ASER) Data – years 2008 and 2012 – for children (aged 3 through 15) still in school.
Bandwidths: 'CCT' is the Calonico et al. (2014b) method.
Variables: 'Read Letter' is if the child can recognize the letter. 'Read Word' is if the child can read the word. 'Read Level 1' if the child has achieved reading level 1. 'Numbers 1-9' if the child can identify the digits between 1 and 9. 'Numbers 10-99' can identify 10 through 99. 'Subtraction' can perform simple subtractions.

Table 1.24: District GDP

Log(District GDP) 2000-06		
RD Estimate	0.137	0.190
	(0.132)	(0.126)
Observations	664	838
Fuzzy Conventional p-value	0.303	0.132
Fuzzy CCT Corrected p-value	0.316	0.141
Bandwidth selection procedure	CCT	I and K
District GDP (Rupees) 2000-6		
RD Estimate	5,346	3,711
	(2,874)*	(3,142)
Observations	1,109	650
Fuzzy Conventional p-value	0.0629	0.237
Fuzzy CCT Corrected p-value	0.0181	0.236
Mean dependent variable	17471.8	17471.8
Bandwidth selection procedure	CCT	I and K

District Domestic Product Sources: Department of Statistics and Programme Implementation, Government of West Bengal; Planning Commission; Directorate of Economics and Statistics Government of Uttar Pradesh; Department of Economics and Statistics Government of Tamil Nadu; Directorate of Economics and Statistics Government of Rajasthan; Department of Planning Government of Punjab; Planning and Coordination Government of Odisha; Directorate of Economics and Statistics Government of Maharashtra; Directorate of Economics and Statistics Government of Kerala; Planning Programme Monitoring and Statistics Department Government of Karnataka; Directorate of Economics and Statistics Government of Bihar; Directorate of Economics and Statistics Government of Assam; Andhra Pradesh State Portal.
Bandwidths: 'CCT' is the Calonico et al. (2014b) method. 'I and K' is the Imbens and Kalyanaraman (2012) method. 'CCT corrected p-value' is the bias-corrected p-values using the method in Calonico et al. (2014b)

1.12 Derivations in the Model

1.12.1 Education Sector

1.12.1.1 Supply of Public and Private Schools

Public schools want to maximize the overall access to education A_d for the students in the entire district d. The district d receives R_d from the government, and spends p_m for each input x_m into the schooling production functions. The vector of inputs at the district level $\mathbf{x_m}$ can consist of new schools, better qualified teachers, better infrastructure, more resource-centers, etc.

$$\max_{\mathbf{x_m}} A_d(\mathbf{x_m}) \tag{1.32}$$

$$s.t. \sum_{m=1}^{M} p_m x_m \leq R_d \,, \tag{1.33}$$

where $\frac{\partial A}{\partial x_m} > 0$, $\frac{\partial^2 A}{\partial x_m \partial x_m} < 0$, $\frac{\partial^2 A}{\partial x_m \partial x_n} > 0$. From the first order conditions, it is easy to derive the optimal amount of inputs of type m: $x^*_{md}(R_d, \mathbf{p_m})$, where $\frac{\partial x^*_m}{\partial R_d} \geq 0$ and $\frac{\partial x^*_m}{\partial p_m} \leq 0$. An increase in government funding R_d, thus increases the amounts of inputs into the schooling-access production function, and increases the overall access to education for the students in the district A_d.

For example, one functional form that is consistent with the setup is a simple Cobb-Douglas function:

$$A(\mathbf{x_m}) = \prod_{\mathbf{m}} \mathbf{x_m}^{\alpha_\mathbf{m}} \,, \tag{1.34}$$

where $0 < \alpha_m < 1$ and $\sum_m \alpha_m = 1$.

The optimal amount of inputs of type m are therefore $x^*_m = R_d \frac{\alpha_m}{p_m}$, and the overall access to education is given by:

$$A_d(R_d, \mathbf{p_m}) = \mathbf{R_d} \prod_{\mathbf{m}} \left(\frac{\alpha_\mathbf{m}}{\mathbf{p_m}} \right)^{\alpha_\mathbf{m}} \tag{1.35}$$

An increase in government funding increases the overall access to education in a proportional manner under the Cobb-Douglas form.

Private schools, however, are profit maximizers with heterogeneous costs:

$$\max_{X_j} \; p_d\bar{\theta}_d X_j - Z(X_j) \;, \tag{1.36}$$

where the costs are $Z(X_j) = z_{1j}X_j + \frac{1}{2}z_{2d}X_j^2$. The supply-curve of schooling for school j is therefore:

$$Q_{jd} = \bar{\theta}_d X_j^* = \bar{\theta}_d \frac{p_d\bar{\theta}_d - z_{1j}}{z_{2d}} \;, \tag{1.37}$$

Since there is free entry of private schools into these regions, schools will enter until $\pi_{jd} = 0$. The marginal school, therefore will have a cost-parameter $\tilde{z_{1d}} = \bar{\theta}_d p_d$. If costs are drawn from a distribution $F(z_{1j})$, then the fraction of schools that enter the region is given by: $F\left(\bar{\theta}_d p_d\right)$.

The overall supply of private schooling is therefore:

$$S^{sy}_{pvt,d} = \int_0^{p_d\bar{\theta}_d} \bar{\theta}_d \frac{p_d\bar{\theta}_d - z_{1j}}{z_{2d}} f(\tilde{z}_1) dz_{1j} = \frac{\bar{\theta}_d}{z_{2d}} [p_d\bar{\theta}_d - \mathbb{E}_d(z_{1j}|z_{1j} < p_d\bar{\theta}_d)] \;, \tag{1.38}$$

where $f(\tilde{z}_1)$ is the conditional distribution of private school costs of entrants.

The aggregate profits of private schools, Π, will also be affected by changes in prices and average productivity, where the aggregate profits are:

$$\Pi = \int_0^{\bar{\theta}_d p_d} \frac{(p_d\bar{\theta}_d - z_{1j})^2}{z} dF(z_{1j}) \tag{1.39}$$

1.12.1.2 Education Market Equilibrium and Changes in Policy

The demand for schooling is determined by the household decisions, where $s^*_{id} = \frac{\beta_d - \bar{r_d} - \eta_i}{\Gamma}$. Given a distribution for $\eta_i \sim H(\eta)$, the overall demand for schooling in district d comes from households:

$$S_d^{Dd} = \int \frac{\beta_d + \Psi A_d - p_d - \eta_i}{\Gamma} dH(\eta) = \frac{\beta_d + \Psi A_d - p_d - \bar{\eta_d}}{\Gamma}, \tag{1.40}$$

where $\bar{\eta_d} = \mathbb{E}[\eta_i | i \in d]$. The overall supply of schooling comes from both public and private schools:[81]

$$S_d^{Sy} = \frac{\bar{\theta}_d}{z_{2d}} [p_d\bar{\theta}_d - \mathbb{E}_d(z_{1j}|z_{1j} < p_d\bar{\theta}_d)] + A_d \tag{1.41}$$

[81] Alternatively, the public-school "supply" can be separated from the notion of access A_d. For example, the supply of public schools, specifically, could be $x^*_{school} = R_d \frac{\alpha_{school}}{p_{school}}$. Doing this, would not change the model's predictions.

Here, it is clear that the supply of public-schools doesn't depend on the fees, since many do not charge fees, and profit-maximization is not the motive of public school provisioning. Together, equations (1.40) and (1.41) determine the equilibrium price and quantities of schooling in the district. Depending on the distribution of z_{1j}, a closed-form solution may be found. For example, if the conditional distribution of private school costs is uniform $\tilde{f(a)} \sim U[0, p_d\bar{\theta_d}]$, then the equilibrium price and quantity is:[82]

$$p_d^* = \frac{\beta_d + (\Psi - \Gamma)A_d - \bar{\eta_d}}{\Gamma\left(\frac{\bar{\theta}_d^2}{z_{2d}}\right) + 1} \quad and \quad S_d^* = \frac{\bar{\theta}_d^2\left(\beta_d + \Psi A_d\right) + z_{2d}A_d}{\Gamma\bar{\theta}_d^2 + z_{2d}} - \frac{\bar{\eta_d}}{\Gamma} \tag{1.42}$$

Improving access to schooling, by building newer schools or upgrading its infrastructure will reduce the marginal costs of schooling (Behrman et al., 1996; Birdsall, 1985). For example, under the Cobb-Douglas public-schooling production function, one can see that the fall in the marginal costs of schooling are directly in proportion to the increase in revenues from the government.

$$r_{id} = -R_d\Psi\prod_m \frac{\alpha_m}{p_m}^{\alpha_m} + p_d^*(R_d) + \eta_i \tag{1.43}$$

One can define $D = 1$ for districts that received government funds. Then the optimal years of schooling becomes:

$$S_d^* = \phi_1\beta_d + \phi_2 R_d - \frac{\eta_d}{\Gamma} \ , \tag{1.44}$$

where $\phi_1 \equiv \left(\frac{\bar{\theta}_d^2}{\Gamma\bar{\theta}_d^2 + z_{2d}}\right)$ and $\phi_2 \equiv \left(\frac{\left(z_{2d} + \Psi\bar{\theta}_d^2\right)\left(\prod_m \frac{\alpha_m}{p_m}^{\alpha_m}\right)}{\Gamma\bar{\theta}_d^2 + z_{2d}}\right)$. In equation (1.44) the equilibrium amount of schooling is affected by the expansion of public schooling.

1.12.2 Elasticity of Capital

So far the model assumes (a) that capital is perfectly supplied at the rate R^*, and (b) is not skill-biased. If however, capital was fixed at a value $\bar{K_d}$ in a district, it would not change the qualitative predictions of the model, nor the parameters estimated. The average earnings for a worker with age a and skill s in district d would be:

[82] If the supply of public schools was instead modeled as x_{school}^*, then the equilibrium quantity would be $S_d^* = \frac{\bar{\theta}_d^2(\beta_d + \Psi A_d) + z_{2d}\left(R_d\frac{\alpha_{school}}{p_{school}}\right)}{\Gamma\theta_d^2 + z_{2d}} - \frac{\bar{\eta_d}}{\Gamma}$. This would produce the same qualitative results going forward.

$$\log w_{asd} = \log\left(\frac{\partial Y_d}{\partial \ell_{asd}}\right) = \log\theta_{sd} + \log\psi_a + \left(\left(\frac{1}{\sigma_E}-1\right)\left(\frac{1}{\varrho}\right)\log Y_d - \left(\frac{1}{\sigma_E}-1\right)\left(\frac{1-\varrho}{\varrho}\right)\log \bar{K}_d\right) + \tag{1.45}$$

$$\left(\frac{1}{\sigma_A}-\frac{1}{\sigma_E}\right)\log L_{sd} - \frac{1}{\sigma_A}\log \ell_{asd}\ ,$$

Here the term $\left(\left(\frac{1}{\sigma_E}-1\right)\left(\frac{1}{\varrho}\right)\log Y_d - \left(\frac{1}{\sigma_E}-1\right)\left(\frac{1-\varrho}{\varrho}\right)\log \bar{K}_d\right)$ is common across cohorts and skill levels. Along with Y_d, it gets differenced out in the derivation.

1.12.3 Skill Biased Capital

In Model subsection 1.3.1 I introduce skill biased capital as affecting the productivity parameter θ_{sd}. Below, I explicitly model skill biased capital to show how flexible forms of introducing it do not influence the estimation strategy or results. In the following set up, the noticeable changes are where Equation (1.3) has been modified into Equation (1.48), which includes an elasticity of substitution between labor ℓ_{sd} and skill biased capital k_{sd} represented by σ_s:

$$Y_d = L_d^{\varrho} K_d^{(1-\varrho)} \tag{1.46}$$

$$L_d = \left(\sum_s \theta_{sd} L_{sd}^{\frac{\sigma_E-1}{\sigma_E}}\right)^{\frac{\sigma_E}{\sigma_E-1}} \tag{1.47}$$

$$L_{sd} = \left(\Lambda_s k_{sd}^{\frac{\sigma_s-1}{\sigma_s}} + (1-\Lambda_s)\ell_{sd}^{\frac{\sigma_s-1}{\sigma_s}}\right)^{\frac{\sigma_s}{\sigma_s-1}} \tag{1.48}$$

$$\ell_{sd} = \left(\sum_a \psi_a \ell_{asd}^{\frac{\sigma_A-1}{\sigma_A}}\right)^{\frac{\sigma_A}{\sigma_A-1}} \tag{1.49}$$

Given this new set up, earnings can be represented by Equation (1.50), instead of Equation (1.4):

$$\log w_{asd} = \log\widetilde{\varrho} + \log\psi_a + \frac{1}{\sigma_E}\log Y_d + \left(\frac{1}{\sigma_s}-\frac{1}{\sigma_E}\right)\log L_{sd} + \left(\frac{1}{\sigma_A}-\frac{1}{\sigma_s}\right)\log \ell_{sd} - \frac{1}{\sigma_A}\log \ell_{asd} \tag{1.50}$$

This new set up does not change the estimation or the interpretation of the estimates. In the following equation, that replaces Equation (1.24) to estimate the GE effects on all workers, the skill-biased capital term is captured by the term L_{sd}:

$$log\frac{w_{so,D=1}}{w_{so,D=0}}-log\frac{w_{uo,D=1}}{w_{uo,D=0}}=\left(\frac{1}{\sigma_s}-\frac{1}{\sigma_E}\right)\left[log\frac{L_{s,D=1}}{L_{u,D=1}}-log\frac{L_{s,D=0}}{L_{u,D=0}}\right]+\left(\frac{1}{\sigma_A}-\frac{1}{\sigma_s}\right)\left[log\frac{\ell_{s,D=1}}{\ell_{u,D=1}}-log\frac{\ell_{s,D=0}}{\ell_{u,D=0}}\right] \tag{1.51}$$

1.12.4 Deriving Equations (1.26) and (1.27)

In Equations (1.26) and (1.27) I derive how to estimate the two different returns to education $\beta_{as,D=1}$ and $\beta_{as,D=0}$, in terms of earnings for the younger cohorts. First to derive $\beta_{as,D=0}$, we use the fact that the average earnings is a weighted average of skilled and unskilled workers:

$$log\frac{w_{y,D=1}}{w_{y,D=0}}=(\ell_{sy,D=1}\log w_{sy,D=1}+\ell_{uy,D=1}\log w_{uy,D=1})-(\ell_{sy,D=0}\log w_{sy,D=0}+\ell_{uy,D=0}\log w_{uy,D=0})$$

$$=\ell_{sy,D=1}(\log w_{sy,D=1}-\log w_{sy,D=0})+(\ell_{sy,D=1}-\ell_{sy,D=0})\log w_{sy,D=0}+$$
$$\ell_{uy,D=1}(\log w_{uy,D=1}-\log w_{uy,D=0})+(\ell_{uy,D=1}-\ell_{uy,D=0})\log w_{uy,D=0}$$

$$=\ell_{sy,D=1}log\frac{w_{sy,D=1}}{w_{sy,D=0}}+\ell_{uy,D=1}log\frac{w_{uy,D=1}}{w_{uy,D=0}}+$$
$$(\ell_{uy,D=1}-\ell_{uy,D=0})\log w_{uy,D=0}+(\ell_{sy,D=1}-\ell_{sy,D=0})\log w_{sy,D=0}$$

$$=\ell_{sy,D=1}log\frac{w_{sy,D=1}}{w_{sy,D=0}}+\ell_{uy,D=1}log\frac{w_{uy,D=1}}{w_{uy,D=0}}+\Delta\ell_{sy}\underbrace{log\frac{w_{sy,D=0}}{w_{uy,D=0}}}_{\beta_{as,D=0}} \tag{1.52}$$

Similarly, I derive $\beta_{as,D=1}$ in terms of observable wage discontinuities that I can estimate:

$$log\frac{w_{y,D=1}}{w_{y,D=0}} = (\ell_{sy,D=1}\log w_{sy,D=1} + \ell_{uy,D=1}\log w_{uy,D=1}) - (\ell_{sy,D=0}\log w_{sy,D=0} + \ell_{uy,D=0}\log w_{uy,D=0})$$

$$= \ell_{sy,D=0}(\log w_{sy,D=1} - \log w_{sy,D=0}) + (\ell_{sy,D=1} - \ell_{sy,D=0})\log w_{sy,D=1} +$$
$$\ell_{uy,D=0}(\log w_{uy,D=1} - \log w_{uy,D=0}) + (\ell_{uy,D=1} - \ell_{uy,D=0})\log w_{uy,D=1}$$

$$= \ell_{sy,D=0}log\frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=0}log\frac{w_{uy,D=1}}{w_{uy,D=0}} +$$
$$(\ell_{uy,D=1} - \ell_{uy,D=0})\log w_{uy,D=1} + (\ell_{sy,D=1} - \ell_{sy,D=0})\log w_{sy,D=1}$$

$$= \ell_{sy,D=0}log\frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=0}log\frac{w_{uy,D=1}}{w_{uy,D=0}} + \Delta\ell_{sy}\underbrace{log\frac{w_{sy,D=1}}{w_{uy,D=1}}}_{\beta_{as,D=1}} \tag{1.53}$$

1.13 Data Appendix

DISE: Data for inputs into schools comes from the District Information System for Education (DISE), which was established to collect data at the school level in order to inform policy makers in the Indian government about the bottlenecks in the education sector. While a limited number of their variables are available freely at an aggregated level, the bulk of their interesting data is obtainable only at a school-by-school basis on their website. I therefore collected 10% of the data, stratified by year, on a school-by-school basis and compiled it for each school separately. DISE claims to cover all the schools in the country (about 1.45 million schools in 2014) each year between 2005 and 2014, and consists of detailed information on number of schools, when they were built, whether they are public or privately owned, number of teachers by levels of education, and various infrastructural features. The DISE data was initially meant to cover only in DPEP districts, but was expanded to cover the rest of the country in the early 2000s. The data is collected by head teachers, and verified by cluster resource coordinators and block educational officers. Cross verification is done by head teachers of one school for another, and by Department of Education officials. See table 1.1 for summary statistics for the year 2005.

Census data has a limited number of outcome variables, including literacy by gender and rural-urban status. The Census has detailed tables at all three of the administrative levels - states, districts and sub-district. A panel of sub-districts can be created using the 1991, 2001 and 2011 Census years, all of which include sub-district-level statistics. The panel is particularly challenging because of splits and merges in various districts, so I used detailed information on administrative areas to compile the panel. The 1991 Census determines the running variable for the RD, since the 1991 female literacy rate was used to determine which districts are eligible for DPEP funds. I calculate this female literacy rate in 1991 for females above 6 years old, and exactly replicate the numbers highlighted in the DPEP reports.

National Sample Survey (NSS): I use household surveys to study the impact on education, earnings, expenditures, migration and other labor market characteristics. The National Sample Survey (NSS) is a nationally representative survey used by many researchers studying India. It is the largest household survey in the country, and asks questions on weekly activities for up to five different occupations per person, and earnings during the week for each individual in the household. The NSS asks detailed questions about thirteen different levels of education, which I convert into years for some of the analysis. There is also a consumption module which asks detailed questions on expenditures on various goods, including education-related expenditures, with a 30 day recall period. The probability-weighted sample is constructed using a two-staged stratified sampling procedure with the first stage comprising of villages and block, and the second stage consisting of households. Households are selected systematically with equal probability, with a random start.

I use three different rounds of the NSS data. The 2004-5 "thick" round is the last large-sample round while the policy was still in place. This allows me to get at costs of education from the household side. The 2007-8 small-sample "thin" round asks detailed questions on migration, which I use to test the effect of this policy on migration decisions as well. The main dataset, however, is the 2009 round, which was used to study the longer term impacts of the DPEP policy. The 2009 round is the first large-sample round after the end of the DPEP program, and has the added advantage of allowing enough time for students affected by the policy to become a part of the labor market. Summary statistics for the 2009 NSS round are presented in Table 1.2. In my analysis, I restrict individuals to be between 17 and 75 years of age, and the results are robust to relaxing these constraints.

Annual Survey of Industries (ASI): To study the behavior of firms, I use the Annual Survey of Industries (ASI), which is a census of all manufacturing firms in the country that employ more than ten persons. This data is available at the establishment level, and has information on the type of products produced, wages paid, and number of employees among other things. One can then use this data to study whether changes in the skill level of the population can affect firm mobility and production decisions.

District Domestic Product (DDP) Data: DDP data is compiled from each state's statistical office and made into a panel. The series is for gross (rather than net) domestic product, and the base year is the year 2000. The various statistical offices are: Department of Statistics and Programme Implementation, Government of West Bengal; Planning Commission; Directorate of Economics and Statistics Government of Uttar Pradesh; Department of Economics and Statistics Government of Tamil Nadu; Directorate of Economics and Statistics Government of Rajasthan; Department of Planning Government of Punjab; Planning and Coordination Government of Odisha; Directorate of Economics and Statistics Government of Maharashtra; Directorate of Economics and Statistics Government of Kerala; Planning Programme Monitoring and Statistics Department Government of Karnataka; Directorate of Economics and Statistics

Government of Bihar; Directorate of Economics and Statistics Government of Assam; Andhra Pradesh State Portal.

Creating the Panel: Due to splits and merges, and other changes in district boundaries, creating a consistent dataset is a non-trivial task. Only 41% of districts were unaffected by changes in district boundaries between 1991 and 2009. Of the 607 districts in the 2009 NSS household survey data, 571 were successfully merged with the 1991 Census (to obtain the running variable) and the list of DPEP districts. This merging was done based on administrative Census reports and shapefiles using Arc-GIS. Of these, 551 were merged with the manufacturing industries ASI data (the other twenty districts had no manufacturing firms). The school-level DISE dataset only covers 408 of these districts since the schools were surveyed only in the larger states. The household-level results will therefore be shown for both the entire dataset and the sub-sample of DISE districts only as well.

CHAPTER 2

Incentivizing Standards or Standardizing Incentives? Affirmative Action in India

I study the impacts of affirmative action policies on schooling incentives in India. My results indicate that when the probability of getting into college or of getting a job is increased through affirmative action, minority group students are incentivized to stay in school longer given the prospects of directly benefiting from such policies in the near future. This approach is unique, in that it focuses on the incentives affirmative action gives to those who are not yet eligible for the policy *per se*. I create a comprehensive primary dataset using state commission reports which allows for a regression discontinuity and difference-in-discontinuities analysis. These results are supported at the national level by using a difference-in-differences approach, and utilizing variation in state-level policies. Together these estimators consistently show that affirmative action policies incentivize about 1.38 more years of education for the average minority group student, and 2.2 more years of education for a student from a marginal minority sub-group. Given the debate about the effectiveness of such policies, my research shows that it is particularly important for both researchers and policy makers to take into account these incentive effects when evaluating affirmative action programs.

2.1 Introduction

Affirmative action is a contentious issue for policy-makers and academics in many countries across the world, including the US, India, Sri Lanka, Malaysia, Nigeria and Brazil. The research and policy debates center around issues of college-mismatch (Arcidiacono et al., 2011), direct effects on college enrollment or test-scores of beneficiaries (Bagde et al., 2016), and the consequent effects on non-minority groups (Bertrand et al., 2010). However, there is little known about the impacts on human capital investment incentives for potential *future* beneficiaries. Since affirmative action changes the returns to education, this can lead to large indirect effects on skill-acquisition. In this paper, I study the causal impact of affirmative action policies on educational attainment in India, focusing on policies that make it easier for minority-groups to get into college or get a government job. I find that by raising the future expected returns to education, such policies incentivize minority-group students to stay in school for longer.

I take three distinct empirical approaches in this analysis. First, I study a nationwide law change that reserved federal government jobs for certain lower-caste candidates. These jobs required specific educational qualifications, thereby raising the returns to certain educational levels. By comparing eligible to non-eligible castes, and student age-cohorts that were young enough to change their schooling decisions to those who were too old, I determine the change in educational attainment for the average low-caste student. I find that, on average, such students attain 1.38 more years of education. These effects are absent among non-eligible minority groups, non-eligible candidates within the eligible minority groups, and low-income students from the upper caste.

The average effects, however, say little about how this impact would change as we increase the 'intensity of reservations,' which I define to be the fraction of seats reserved relative to the fraction of the population that is from the minority group. So for my second approach, I use affirmative action laws for college-admission and government jobs at the state level, rather than the federal level. I create an original dataset based on historical laws passed in each state in the country by petitioning the Government for archived commission reports. I then exploit three sources of variation – the timing of these laws, the minority groups eligible, and this intensity of reservations – to determine how changing the intensity of these programs affects educational attainment. Certain states reserve a larger fraction of their seats than other states. By comparing low-intensity states to high-intensity ones, I show that the relationship between the change in educational attainment and relative fraction of seats reserved is concave. This suggests that extremely intensive affirmative action programs may detrimentally lower the educational attainment of minority groups.

These results, however, do not address the effects of expanding these benefits to an additional marginal sub-group. India has numerous sub-castes, some of which are eligible for affirmative action benefits, and others that are not. In my third approach, I compare sub-castes that just received the programs to sub-castes that just lost out, to causally identify this parameter. One of the states in India conducted a large socio-economic survey and ranked the sub-castes on an 'index of backwardness' in order to determine which sub-castes should be eligible for these programs. Any sub-caste that had a score greater than half of the total value of the index was eligible. Using a regression discontinuity (RD) design I compare sub-castes on either side of the cutoff. Since the educational attainment of older members of these castes should not be affected by the introduction of this policy, I further use them as a control group in a difference-in-discontinuities approach. I find that, on average, a student from the marginal sub-caste in that state attains 2.2 more years of education. This suggests that there are plausibly large positive unintended effects of expanding the coverage of these programs to other marginal minority sub-groups.

While the Regression Discontinuity (RD) and Difference-in-Discontinuities approach exploit certain policy features in a particular Indian state, the cross-state intensity and the national-level Difference-in-Differences approaches are representative of the entire country. All estimators consistently point towards an increase in educational attainment for the targeted minority group in response to reservations in higher education and government jobs. More importantly, the different estimators determine different parameters that would be useful for any welfare analysis of these policies. While the Difference-in-Differences approach estimates the average impact, the RD determines the impacts on the marginal sub-caste. The state-level variation allows me to determine how these impacts vary with differing intensities of the policies. All three parameters are crucial for any meaningful discussion of the costs and benefits of these policies.

This paper is unique in its approach, in that it is among the first papers to empirically study the causal impacts on incentives before the benefits of the policy actually kick-in. Unlike the other papers in the India context, I use nationally representative data to look at the impacts on the entire country, rather than on a subset of engineering colleges. While most research focuses on college-admissions I also look at labor market affirmative action policies, and study the impacts on the extensive margin of drop-outs rather than the intensive margin of test-scores. Last, I compile an original data set of state-level laws, and exploit a state's law to perform a regression discontinuity and a difference-in-discontinuities analysis to identify the causal impact of the policies.

In the rest of this section I discuss the possible effects of affirmative action programs by couching it in the relevant theoretical and empirical literature on the returns to education and affirmative action policies. In Section 2.2 I discuss the context of caste and class in India, and the underlying legal and historical foundation behind these policies. After which, in Section 2.3 I discuss the data and provide some descriptive evidence of the trends over time for different socio-economic groups. In Section 2.5, I set-up a dynamic optimization model with testable implications, and predictions on what may confound the empirical analysis. The main focus of this paper is Section 2.6. It discusses the various empirical strategies used and their corresponding results. The last section concludes, and discusses policy implications.

2.1.1 Returns to Education and Human Capital Investments

Government education policy in low-income countries is usually associated with lowering the costs of education rather changing the returns (King and Orazem, 2008). Numerous examples of schooling expansionary programs that reduce both the monetary and non-monetary costs for students can be found across the developing world.[1] The government policies I study in this paper, are therefore unique in that they change the future returns, rather than the current costs of schooling. Since the benefits are in the future, programs that affect returns are likely to have different impacts than an immediate tangible fall in the costs of schooling. Furthermore, while costs are easy to perceive, information on returns and future opportunities may be poor in low-income settings. This lack of information may skew the demand for schooling (Dinkelman and Martinez, 2014; Jensen, 2010).

Nonetheless, outside of government programs, a large literature exists on human capital investments in response to changes in the returns to education. This is true not just of the US (Freeman, 1976; Griliches, 1997; Kane, 1994; Ryoo and Rosen, 2004), but also in the developing world, and especially the Indian context. On the one hand, this research finds evidence of increases in schooling with increasing returns. One of the earliest papers in the Indian context by Foster and Rosenzweig (1996) shows how the Green Revolution led to an increase in primary schooling arguably because of higher returns to education. Similarly, Kochar (2004) finds that households increase schooling in response to higher returns in the nearest urban labor market, and Jensen (2012) shows that better jobs for women in the IT sector of Delhi increases schooling for girls. At higher education levels, Khanna and Morales (2015) studies how an increase in returns to IT sector jobs in the US and India increases enrollment in engineering schools in the 1990s.

On the other hand, increasing the returns to education may have adverse effects in such low-income settings. Jensen and Miller (2015) show that strategic incentives amongst rural Indian households can actually lower educational attainment in response to higher returns to education. They argue that parents that want a child to remain at home and look after them, curb their child's migration opportunities by lowering their educational investments. Similarly de Brauw and Giles (2008) find that school attainment falls in rural China in response to better migration opportunities, because of higher opportunity costs of schooling. It is, therefore, unclear what the expected impacts of increased returns are in the developing country context.

[1] Some examples can be found in Indonesia (Duflo, 2001), Burkina Faso (Kazianga et al., 2013), Zimbabwe (Aguero and Bharadwaj, 2014), Nigeria (Osili and Long, 2008), Sierra Leone (Cannonier and Mocan, 2012), Uganda (Deininger, 2003), Zambia (Ashraf et al., 2015), Kenya (Bold et al., 2013b), Tanzania (Sifuna, 2007), West Bank & Gaza (Angrist, 1995), and India (Afridi, 2010; Chin, 2005; Khanna, 2016).

2.1.2 Affirmative Action in the US and India: Theory and Evidence

Theoretical work on affirmative action discusses different types of behavioral responses. The Coate and Loury (1993) employer-learning model shows that under certain assumptions such policies can indeed encourage effort, and over time the policies could lead to a 'benign equilibrium' where employers' negative stereotypes about the minority group are eradicated. However, under other assumptions it could lead to a 'patronizing equilibrium' where the negative stereotypes persist, potentially discouraging human capital accumulation. Furthermore, if employers devalue the credentials of any minority group candidate because of the affirmative action policies, it can disincentivize members of the minority group from obtaining education.

Similarly, in a signaling model, affirmative action may discourage investments for low-ability minority group students. Suppose that in the absence of affirmative action, even the high-ability low-caste students would not get into college and only finish high-school. Then the low-ability low-caste students would finish high-school as well, so as to take advantage of being 'pooled' with the high-ability students. If affirmative action allows the high-ability low caste students to attend college, then we get a separating equilibria where the low ability students would drop out a lot earlier.[2]

It is, therefore, crucial to understand not only who is affected but also how intense these programs are. In the Indian context, affirmative action programs are more salient and larger in magnitude than in most other countries. Reserving a large fraction of seats may allow low-ability low caste students to get into college and into public sector jobs, exacerbating employers' negative stereotypes. If such employers discriminate against future applicants it may discourage further human capital investments. On the other hand, large-scale reservations may also lead to a higher 'pooling equilibrium' whereby both the high and low ability students from the minority group get more education – the low-ability students taking advantage of being pooled with the high-ability students. It is, therefore, important to study not just the average impacts of the entire minority group, but also how increasing the intensity of reservations affects these average impacts, and how enlarging the definition of the minority group may affect outcomes for this additional sub-group. I explore all these parameters in my results.

What then should be the form of affirmative action? Fryer and Loury (2005) show that 'equal opportunity' is often not enough to close educational inequalities that arise from historical discrimination, and increasing the probability of getting into college may motivate students to graduate from school, overcoming 'effort pessimism.'[3] Hickman (2013) uses an auction theory based structural model and compares various forms of potential policies to show that race quotas in the US would induce more human capital investment by minorities, but could involve a larger welfare loss. The literature also mentions 'complacency' effects of such policies on incentives for schooling – for instance, smarter sections of the minority

[2]This is a modification of a result shown in Bedard (2001).

[3]Anthropological studies in the American context suggest that difficulties faced by minority groups in finding employment ('job ceiling' hypothesis) discourage them from attaining education (Ogbu, 2003).

group could put in less effort knowing that it is easier to get into college.[4] In the political and academic sphere these possible outcomes are the topic of contentious debate. Nonetheless, there is little empirical evidence to back up these claims.[5]

The evidence in the US context highlights large possible costs associated with such policies. Arcidiacono et al. (2014, 2011) study the impact of affirmative action policies on college fit and mismatch, and show that laws banning the use of racial preferences in California public colleges lead to better match quality and higher graduation rates in colleges. The form of preferential treatment is also important – Domina (2007) shows that the diversity programs enacted in Texas, after affirmative action was banned, boosted educational outcomes at the high-school level. Furthermore, eliminating the use of affirmative action in Texas and California state universities did not seem to adversely impact the SAT-sending behavior of highly qualifies minority group students (Card and Krueger, 2005). Lastly, if peers are seen to benefit from this policy, then a 'role model' effect may also have a positive impact on educational attainment. However, evidence in the American context shows little support for 'role model' hypothesis – it instead posits that benefiting minority students are less popular because they are accused of 'acting white' (Fryer and Torelli, 2010; Ogbu, 2003).[6]

In the Indian context, the literature has focused on the direct impacts on a sample of engineering colleges. Empirical work suggests that college-reservations are well targeted, improve the performance of the minority groups in question, and have "*strong positive economic effects*" (Bertrand et al., 2010). Bagde et al. (2016) also look at a sub-sample of engineering colleges in a particular Indian state, and argue that reservations have a "*significant and substantial positive effect both on college attendance and first-year academic achievement.*" On the other hand, Krishna and Robles (2015) look at a detailed data set from an engineering college in India and show that affirmative action policies lead to mismatch – minority students end up earning less than they would have if they picked less selective majors. While these papers study a group of engineering colleges and focus on outcomes at the collegiate level, my work looks at educational attainment at all levels of schooling (before the policy benefits kick in), and studies the country as a whole.[7]

2.2 Caste, Class and Reservations in India

In India, affirmative action policies are defined on the basis of caste or social class, and the policy interventions are much larger and more salient than in most other countries. The Constitution identifies

[4]Assuncao and Ferman (2015) show that affirmative action policies reduces test scores for the minority group in Brazil.

[5]Weisskopf (2004) provides a theoretical comparison of affirmative action policies in the US and India, and discusses the various expected effects of such policies, including the impacts on incentives to stay in school.

[6]Teacher-student pairings of the same race, however, have been found to have positive impacts, which may be evidence in support of a role-model effect (Dee, 2004).

[7]In a somewhat different vein, Rao (2016) studies a program that required Delhi public schools to admit students from poorer backgrounds, and found that this had large positive impacts on the prosocial behavior of richer peers.

certain castes as the most disadvantaged group and codifies them as the Scheduled Castes (SCs). It also enumerates certain aboriginal tribal groups, which are referred to as Scheduled Tribes (STs). Over time there has been an attempt to identify groups that are better-off than SCs and STs but less well to do than upper caste members of the different communities. These groups are known as Other Backward Classes (OBCs). There is almost no literature on affirmative action for OBCs, and I focus on this group in my paper.

Over time, the Indian government has instituted laws whereby a certain percentage of seats in colleges or government jobs are set aside for low-caste candidates. This 'reservation policy' primarily benefits SCs, STs and OBCs in various ways. The primary purpose of this law is to provide a level playing field for communities that have suffered from historical discrimination. The Constitution states that "*the State shall promote with special care the educational and economic interests of the weaker sections of society .... and shall protect them from social injustice and all forms of exploitation.*" This law allows states to autonomously reserve seats for different communities in state-run universities and in government jobs, producing useful variation from a researcher's point of view. There is some macro-level evidence highlighting the possible effects of these laws on SC-STs – Desai and Kulkarni (2008) show that educational inequalities have been falling over time for SCs and STs, that do benefit from reservations, but have not been declining for the Muslim community who are excluded from the current reservation policy.

In 1980, a Commission was established to determine what percentage of seats should be reserved in national universities and federal government jobs for OBCs (Mandal, 1980). The report recommended reserving 27% of the seats in national colleges and federal jobs for the OBCs that they identified. This was met with large protests from the urban upper-class public who argued that they were being discriminated against, and that the disadvantaged groups already had a 'level playing field' (Kohli, 2001). In 1993, the federal government implemented the first stage of the Mandal (1980) Commission recommendations by reserving 27% of government jobs for OBCs, and then in 2006 the reservations in colleges were implemented. The Indian Supreme Court excluded the more well-off members of the OBC community (known as the 'creamy layer') from taking advantage of these policies, and this is another source of variation that is exploited.[8]

These caste-based reservations at the central level exist alongside the state laws, which vary in intensity across states. In one empirical strategy, I focus on the state of Haryana, which ranked sub-castes on the basis of their socio-economic disadvantage, and classified the worst-off sub-castes as OBCs. This ranking allows me to obtain a Regression Discontinuity estimate of the impact of these policies. I, therefore, exploit variation across various dimensions: caste, age, region and eligibility in order to determine the impact of such policies. While state-level laws provide quotas in both educational institutions and government jobs, the federal law changes studied here will focus on OBC reservations in government

[8]In 1993, the Supreme Court upheld the implementation of reservations for OBCs in government jobs in the landmark case: *Indira Sawhney v. Union of India, 1993* and introduced the concept of the creamy layer.

jobs.[9]

Importantly, there are four categories of government jobs, all of which are eligible for quotas. The highest category (Groups A and B) require finishing high-school or having a college degree, and these comprise of 11.5% of all the jobs. These are mostly high-level civil servants. The next level (Group C) need candidates to finish either middle school or secondary school, and consist of 58% of the jobs. The last category (Group D) consist of the last 30.5% of the jobs, and require candidates to be either literate or complete primary school. Group C and D jobs include lower skilled jobs like revenue inspectors, assistants, clerks and drivers. Therefore, the incentive effects will not just be seen in graduating from high-school, but also in attaining certain levels of education that make candidates competitive for these jobs.

My paper is among the first to look at reservations in government jobs. Prakash (2010) shows that policies in the 1980s that reserved jobs for a different minority group – the Scheduled Castes (SCs) – had significant impacts on the probabilities of formal sector employment and wages for this specific minority group. Given that OBCs are slightly better off than SCs to start with, we may expect that there would be even more qualified OBCs to avail of these quotas. My work tests whether these future prospects of better employment and wages can actually induce students to reach the educational thresholds required for these jobs.[10]

2.3 Data

For this analysis, I compiled a number of data sources specifically for this analysis. These include a few different household surveys and governmental commission reports. First, I use the Indian National Sample Survey (NSS), which is a representative repeated cross-section carried out every five years. This data set has information on educational attainment, field of graduate study, caste, age, and host of other labor market outcomes along with a comprehensive consumption expenditure module. The nationally-representative 'thick' rounds of the data set are enumerated every 5 years. Since this paper focuses on affirmative action policies instituted in the early 1990s, the main data set used is the 2000 module, which was also the first round to ask questions about whether the person is OBC or not.[11] The 1995 round is too early to capture the effects of policies instituted in the early 1990s, since changes in schooling decisions take time. Whereas the 2005 round is too late and may suffer from other confounding policies

[9]This is because the federal level implementation of OBCs in higher educational institutions only happened recently in 2009.

[10]In the US, there is a literature that shows that schemes like the federal contractor program, under which the targeted groups of women and African-Americans, were given preferential treatment, increased their employment and the demand for them in such sectors (Leonard, 1984; Smith and Welch, 1989), but there is little consensus on the impacts of court-ordered affirmative action in the US (Donohue, 1991).

[11]The NSS 55th Round was collected between July 1999 and June 2000 using a stratified two stage sampling design. First, clusters (rural villages or urban blocks) were sampled, and then 12 households within each cluster were sampled.

introduced in the interim years, and changes in definitions of the OBC group across waves of the survey.[12] In my robustness checks I use the 2005 rounds as well to show that my results are consistent with either round.

The data set has information on *level* of education, rather than years of education. The various levels of education are (a) illiterate, (b) literate without formal schooling, (c) literate with formal schooling, (d) primary school, (e) middle school, (f) secondary school, (g) higher secondary school, (h) college educated. Even though, we may expect the level to matter for eligibility for jobs and colleges, I discuss how to translate these levels into years to be consistent with the rest of the literature, and present results for both changes in the levels of education and years of education.

Primary source data was compiled on affirmative action policies instituted by the federal government and the various Indian states. I did this by obtaining government reports via the Right to Information (RTI) Act. This dataset is comprehensive in that it has information on reservation policies for all states in the country. Furthermore, detailed knowledge on classification and identification of OBCs was found for a few states. The states in question had Committee Reports that laid out the methodology for identifying Other Backward Classes (OBCs) and their recommendations for reservation policy. Therefore, some of the estimation procedures will allow me to look at the effect on the entire country, while more detailed analysis has been done for the states where the in-depth reports were obtained.

The third source is the ARIS-REDS (Additional Rural Incomes Survey and Rural Economic and Demographic Survey 1999) data set. Unlike the NSS data, ARIS-REDS has information on disaggregated sub-castes. While the NSS is nationally representative, it only has information on four broad caste categories. Despite having a smaller sample, the ARIS-REDS asks respondents their sub-castes, and thus has social-group information at a finer level. Unfortunately, neither of these data sets have information on educational aspirations and expectations, nor test scores, which would have been useful for additional analysis.

2.4 Descriptive Evidence

By the end of the decade, in 2000 – when the NSS dataset was collected – there were about 3.9 million Central Government jobs, and 45% of them were in the Indian railways. Only 15% of these jobs were in large cities, 53% of them were in rural areas, and the rest were in small towns. In the 2000 NSS survey, 60% of all organized sector jobs and 14.8% of all enterprise-based jobs, and 2.5% of all jobs were in the public sector (both state and central government).

[12] One big policy change in 2000 was the introduction of OBC level scholarships under the Ninth Five Year Plan (see Gupta(2004)). Another was the implementation of the Millennium Development Goals in 2000. If the 2005 data set was used, then this policy would make it impossible to disentangle the direct effects of reservations in government jobs, because of the coincidental presence of scholarships and MDGs

Table 2.1: Social Groups in India. $Source$: NSS 2000

	SC	**ST**	**OBC**	**Others**	**Total**
Sample Size	94098	66798	195579	237102	593577
Proportion of Sample (%)	15.85	11.25	32.95	39.94	100
Mean Education Level	3.62	3.90	4.26	5.55	4.63
Mean Years of Education (Approx.)	3.037	3.4042	3.9035	5.678	4.4186
Illiterate (%)	46.20	50.93	42.13	28.02	38.34
College Educated (%)	1.96	1.63	2.79	8.42	4.76
Household Month Exp (Rs.)	1245.89	1444.92	1440.81	2074.11	1609.51
Per Cap Month Exp (Rs.)	398.67	427.32	446.33	519.02	465.67
Urban (%)	30.87	22.61	33.21	48.17	37.62
Work in Agriculture (%)	56	72.89	53.31	40.36	51.53
Wage work (not Casual) %	26.61	32.65	40.46	68.38	45.85

'Others' are general category individuals (i.e. not SC, ST or OBCs). 'Mean Education Level' covers 8 levels of education from illiterate to college graduates. Nominal exchange rate: approx Rs. 50 to \$ 1. Household Monthly Expenditure deflated by rural-urban-region-wise CPI.

Table 2.1 uses NSS data to summarize the primary variables of interest by social groups. About one-third of the sample was self-reported to be OBCs. The proportion of SCs (16%) and STs (11%) are smaller. Looking at the mean education level by social group, it is clear that SCs and STs have the lowest rates of educational attainments, whereas OBCs do slightly better than them, but worse than the non-OBC/SC/ST group (known as 'Others' or 'upper-caste'). The mean education level for the upper-caste group is 5.5, indicating that not a large proportion of them are in college either.[13] The mean expenditure for schooling is \$6.85 on average, but about \$16.8 for private school goers (Das et al., 2013b).

SCs, STs and OBCs have lower monthly expenditure compared to the rest of the population (Table 2.1). Furthermore, these three disadvantaged groups are predominantly rural and work in the agricultural sector. They are also more likely to be employed as casual labor rather than in formal wage employment. Kohli (2001) discusses how the Indian growth story has largely been concentrated on urban, English-speaking, educated middle-class families, and the large-scale reforms of 1991 have been unable to bridge the inequalities between these social groups.

In the 2000 NSS data, the yearly wage premium for public sector jobs relative to other jobs with any enterprise is \$485 for OBCs, and \$465 for all people. Over the next decade, government wages were still higher than private sector wages, but private sector wages were rising at a faster rate. Changes in the education distribution and increasing the supply of skilled work to the public sector may have effects on the public sector wage. This is however unlikely in this scenario since government wages are fixed for a decade at a time by the Pay Commissions (1983, 1994 and 2006). These Pay Commissions fix wages

[13] This is important, since if we expect the upper-caste to have reached the ceiling in educational attainment then the reservation policy should have no impact.

and compensation for all public-sector employees and tie them to the rate of inflation.

Since different levels of government jobs, require different educational qualifications, I can study the educational levels for public sector employees using the NSS data. The education distribution for OBCs with government jobs displayed a skew towards higher levels of education. 15.45% of them had less than or a primary school degree, 14.42% finished middle school, 25.13% finished secondary school, whereas 16.63% finished higher secondary school and 28.37% finished college.

Calculations using the NSS data show a large increase in public-sector (and semi-public) employment for OBCs over time. In 1999-00 (six or seven years after implementation of the law), 22% of government and semi-public sector jobs were held by members of the OBC community, whereas in 2004-5 this number was about 27.7%. Representation of OBCs in government jobs has steadily increased since the implementation of the policy to match the 27% amount of seats reserved for them.[14]

In the decade before the quotas were implemented (1981-91), all public sector employment grew by 24% whereas Central government jobs grew by only 7%. Since 1997 Central government full-time employment actually slightly shrank because they began outsourcing and cut full-time jobs to part-time contract work. Between 1998 and 2002, 31% of all public sector vacancies were filled by OBCs. Using, the 2005 round of the National Sample Survey (NSS), one can also look at public sector employment by cohort. Only 23.4% of public-sector employees for cohort born in 1946-55 are OBCs, whereas this is number is 27.4% for the cohort born in 1976-85. By 2011, 6.9% of Central Government Group A workers and 17% of Group D workers were OBCs. The Ministry of Personnel, Public Grievances and Pensions, in 2012 claims that the "*representation of OBCs is still low for the reason that reservation for them started only in 1993. Moreover, those OBCs recruited before 1993 have not shown themselves as OBCs [since] they were recruited/ appointed as general candidates.*"

2.5 A Simple Theoretical Framework

In this section, I set up a simple dynamic optimization problem to highlight the possible effects of quotas on the incentives for students in school, and the empirical challenges. Every period an agent chooses to dropout or stay in school. If the returns to s years of schooling is $w(s)$, then for a discount rate β, the agent's value function is:

$$V_s = \max_{Dropout,Stay} \left(\sum_{t=s}^{\infty} \beta^{t-s} w(s), \quad -c_i + \alpha_i + \beta \left(p_s \frac{1}{1-\beta} B + (1-p_s) V_{s+1} \right) \right) \tag{2.1}$$

[14]For a large fraction of the OBC population, a government job may be their best option given the level of job-security, and non-pecuniary benefits in addition to the pay. Furthermore, facing discrimination in the private sector employment market may make a government job the only option for some.

In this equation, B is the expected net benefits from going to college or getting a government job.[15] The simplest version of the model doesn't allow for on-the-job search, and once the agent drops out and gets a job at wage $w(s)$ he/she earns that wage forever.[16] The cost of an additional year of schooling for person i is c_i. This varies by person, and is drawn from a distribution $F(c)$. α_i captures the preferences for schooling, and would then be affected by 'role-model' effects, if present.[17] The probability of getting into college or getting a government job p_s is a function of various factors:

$$p_s = p_s(schooling,\ caste,\ quotas,\ grades,\ peers,\ ability,\ Z_i) \tag{2.2}$$

Where Z_i is a vector of other individual characteristics that generate heterogenous responses to changes in the probability. Heterogenous levels of optimal schooling are captured by a cost function c_i. The probability function can be different in various contexts. For example, different levels of government jobs require different levels of education, which means that the probability p_s could discretely jump across levels of schooling. Getting into college, however, requires one to complete high-school, therefore $p_s = 0$ for any s below the last year of schooling.

For an increasing and concave wage function $w(s)$, the value function converges to an optimum under certain regularity conditions,[18] and can be solved by using backward induction. The optimal policy function has a threshold strategy, whereby a student chooses to drop out of school when his marginal value from an additional year of schooling is less than the cost he/she must bear. Let this threshold level of education be s^*.

When the probability of getting into college increases, it raises the expected value of an additional year of schooling and lowers dropouts. Thus reservations in colleges can incentivize the marginal student to stay in school for more years. $\frac{\partial s^*}{\partial quota} > 0$.

The effects of reservations in government jobs is less apparent, and depends on the shape of the probability function. Since different government jobs have different thresholds – literacy, finishing primary school, finishing middle school, high school or college – it depends on the distribution of where students would expect to drop out in the absence of the quota. On the one hand, it could incentivize students just below the job-qualification threshold to get at least as much education as the government job requires. On the other hand, it may induce students just above a threshold to drop out at the threshold.

[15] Government jobs may not only pay a higher wage, over and above the going wage, but also provide job security.

[16] Currently the agent has perfect foresight. In alternative specifications, expectations of benefits can be made to depend on the information set of the agent. This will allow us to see the impacts of a change in the peer group that benefits from affirmative action, and incorporate the possibility of 'role model' effects.

[17] It is also easy to add an on-the-job ability that raises wages, that is correlated with costs of schooling.

[18] We need the slope of the wage function to be steeper than probability function:

$$\frac{1}{1-\beta} w_s(s) \geq B p_s$$

For jobs that require only primary schooling, a student may be encouraged to drop out as soon as they finish their primary-school rather than stick around for longer. The shape of the probability function, and the different wages at each level of government job, therefore, determines how students respond to quotas.

There are, however, factors that may confound the identification of these effects. If the quality of schooling and the number of schools increase, then costs of attending are lower, which also tends to discourage the agent from choosing to drop out of school $\frac{\partial s*}{\partial c_i} < 0$. This is an important result as it will be the primary confounding factor in the empirical specification. The government of India made large investments in schooling at around the same time that affirmative action policies were expanded. These investments were made under the National Policy of Education (NPE) program in 1986, under two schemes – Operation Blackboard and the District Primary Education Program (DPEP). Though the education reforms were not caste/class specific, I compile original data and control for both schemes in robustness checks. Furthermore, placebo tests with other minority-groups will be shown to provide evidence that no other coincidental policies will produce these results.[19]

Another implication is the result on test scores. The factors that determine the probability of getting into college p can be seen as substitutes. Marginal students who had a low probability of entering college may now seek to improve their test scores when it is easier to get into college due to reservations. Whereas the marginal student who has a high probability of getting into college may actually lower their effort input when it is easier to enter college.[20] This may lower the variance of the distribution of test scores for the minority group in question.[21] However, due to a lack of reliable data, I will be unable to examine how test scores change in response to these policies.

There are few things that jump out from the model. First, the shape of the probability function p_s determines whether students are incentivized or disincentivized to get more education. Since different government jobs have different educational requirements, students may be encouraged to drop out earlier or later depending on which margin they lie on. Second, the probability p_s depends on the extent of the quotas, and more reservations will increase this probability even more. Therefore, the size of the impacts are a function not only of if seats are reserved or not, but also how many seats are reserved. To get at this, the paper will also exploit variation in the different intensities of reservations across states. Crucially, at the time the Government implemented the quotas in 1993, it expanded the total number of jobs so as to keep the number of jobs in the general category unchanged.

[19] One aspect, that may also affect the probability function, is the peer group. If changing the composition of classrooms has peer effects, then that affects the interpretation of the results. There is little evidence of these negative peer effects, however (Rao, 2016).

[20] In the model this can be seen if 'quotas' and 'test scores' are substitutes: $\frac{\partial testscores}{\partial quota} = \frac{\partial testscores}{\partial p} \frac{\partial p}{\partial quota} < 0$

[21] Consistent with this result, Assuncao and Ferman (2015) show that in Brazil, there is a fall in high-school proficiency of the minority group.

2.6 Identification and Results

I use three different identification strategies to provide a consistent picture of what the incentive effects of affirmative action policies are. While the Difference-in-Differences estimator will identify the Average Treatment Effect on the Treated (ATET), the RD will identify a localized effect – in the neighborhood of the cutoff – for the marginal sub-caste. The state-level variation shows how the treatment effect varies with the intensity of treatment.

2.6.1 Difference-in-Differences

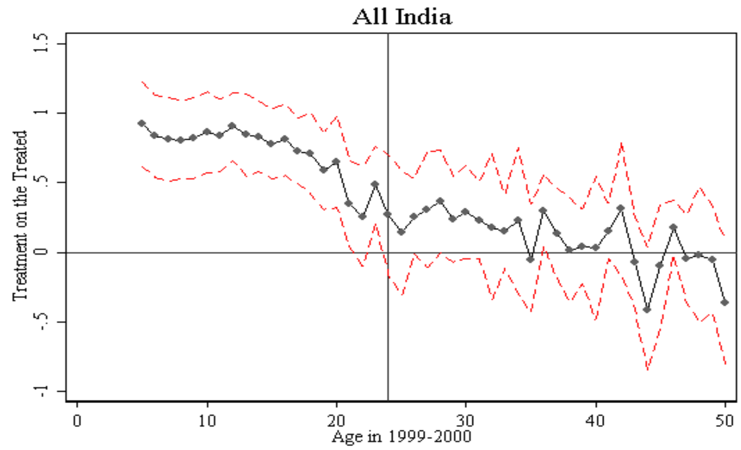
The double difference estimator will exploit variation on two fronts: (a) age and (b) social group. Some cohorts were too old to be affected by changes in the reservation policy. Others will be young enough (and still in school) and can thus change their level of educational attainment. Furthermore, only certain social-groups were eligible, providing variation in policy implementation on the social-group front. As discussed above, the federal government implemented reservations for OBCs in government jobs in 1993, whereby 27% of all public sector jobs were reserved for this group.

If such measures incentivize attaining a higher level of education, we should expect to see this for the OBC group. The average age for entering the last year of high-school lies between 17 and 18 years. In Figure 2.8.5 one can look at the enrollment rates by ages and see a sharp drop off at the age of 18, which is when most students finish high-school. This represents a 16 percentage point drop. The Factories Act of 1948 and Mines Act of 1952 were the first laws to ban employment of persons under the age of 18. Many public sector jobs are therefore only available to people who are at least 18. At the time that reservations were implemented, anyone under the age of 17 or 18 years could have changed their educational attainment. Since the data was collected 6 to 7 years after that, by that time anyone who is above the age of 24 would not have been able to change their level of education.

Furthermore, there would be many high-schoolers who have already dropped out of school and will thus find it hard to change their educational attainment. We should then see the impact of this policy being larger for younger individuals. For instance, the impact on 15 year olds will be smaller than the impact on 10 year olds, since many 15 year olds would have already dropped out of school.

Equation (2.3) is the Differences-in-Differences regression, where $\boldsymbol{\alpha_a}$ and $\boldsymbol{\kappa_c}$ are vectors of dummies for age cohort and caste respectively. Here $\boldsymbol{\beta_{0ac}}$ is a vector containing the relevant coefficients, and is allowed to vary by age cohort a and caste c. The coefficient identifies the Treatment on the Treated ($\boldsymbol{\beta_{0ac}} = ATET_{ac}$) for caste c in cohort a. For $a > 24$, we expect $ATET_{c,a>24} = 0$. If this condition is violated, then we may not be satisfying the parallel trends assumption, which would then bias our coefficient of interest. For $a \leq 24$ and for the OBC group, we would expect $ATET_{c=OBC,a\leq 24} > 0$. Furthermore, the younger is the members of the OBC group, the greater would the expected impact be

Figure 2.1: OBC Coefficients of Diff-in-Diff Regression Across Age Cohorts



Plot of coefficients from Difference-in-Differences regression on 'levels of education'. Standard errors clustered at state-level. People above the age of 24 should be unaffected by implementation of policy. Vertical red line indicates the year of implementation.

(i.e. $ATET_{c=OBC,a} > ATET_{c=OBC,a+1}$) since younger members find it easier to change their schooling. A person who was sixteen at the time the law was passed may have already dropped out of schools many years ago, and would therefore find it difficult to change their level of schooling; whereas a ten-year old now knows the law is in place and can stay in school for much longer.

$$edu_{ac} = \boldsymbol{\alpha_a} + \boldsymbol{\kappa_c} + \boldsymbol{\beta_{0ac}}\boldsymbol{\alpha_a} * \boldsymbol{\kappa_c} + \epsilon_{ac} \tag{2.3}$$

Since the NSS data only asks for education levels and not years of education, the results are produced for both measures.[22] The NSS data set has four broad social groups: (a) SCs, (b) STs, (c) OBCs and (d) Others. The 'Others' category includes the upper-caste section of the population ineligible for any reservations (i.e. not OBCs, SCs or STs). They comprise of 33% of the sample, have a higher monthly per capita expenditure than OBCs, SCs and STs, and are more likely to be urban and salaried (Table 2.1). More than 67% of Muslims fall into this category, and almost 70% of them are Hindu. The above regression specification was run where the the omitted social group was SC-STs and upper-caste members, and ages above 50 were the omitted cohort category.[23] Figure 2.1 plots these coefficients $\boldsymbol{\beta_{0ac}}$ for the OBC group across the various age cohorts, and confidence intervals based on standard errors at the state-level.

[22]A rough translation from the level of education to the years of education maintains the results. However, because there is no clear way to go from the level of education to the precise year of education, both forms have been presented. Furthermore, in the Indian contexts, certain changes in levels of education may be more relevant than the years of education. For example, the difference between being illiterate and literate without formal schooling will change the chances of acquiring a low-level government job.

[23]The results are similar, but slightly larger when the omitted category does not include SCs and STs, and only includes them as a separate non-omitted category instead (see Figure 2.8.6).

The coefficients are close to zero and statistically indistinguishable from 0 for age cohorts above the age of 24. This confirms the absence of pre-treatment differential trends. For those below 24, however, the coefficient is positive and statistically significant. Furthermore, it is larger for younger cohorts as the model predicted. This is because younger cohorts have more time to change their educational decisions, whereas many in the older cohorts would have already completed or dropped out of school and will thus find it difficult to change their decisions. At its highest point, the coefficient is below 1, indicating that the reservation policy caused an increase of at most one level of education for the OBC group – this could be even something informal from a transition away from illiteracy to basic literacy.[24]

Since this is a difference-in-differences coefficient, it captures both the changes over cohorts and across groups relative to older age cohorts. Young OBCs still receive less education than young upper caste students, but the gap is smaller smaller than in the older cohorts by 1 level of education.[25] In the Appendix, there are a few other graphs. One of them reproduces the Difference-in-Differences figure but uses the omitted category to be only upper-caste individuals (Figure 2.8.6) while controlling for the SC-ST categories. This merely makes the impacts slightly larger. Another figure translates the dependent variable from levels of education to years of education (Figure 2.8.8).

Table 2.2: Difference-in-Differences – Years of Education

Education Years	**OBC vs. Others**		
	Younger Cohort	**Older Cohort**	**Difference**
OBC	4.439	4.129	-0.310
	(0.013)	(0.016)	(0.020)
Others	5.579	6.654	1.076
	(0.013)	(0.016)	(0.020)
Difference	-1.140	-2.526	**-1.386**
	(0.018)	(0.022)	**(0.029)**

Using NSS 1999-2000 data. Standard Errors in Parentheses. Difference-in-Differences value in bold. ‘Others’ are general category individuals (i.e. not SC, ST or OBCs).

The Difference-in-Differences tables can be made by dividing the sample into younger and older cohorts; and OBCs and upper-caste. Looking at the Difference-in-Differences results in Table 2.6.1 we see that being an OBC in an age category eligible for education corresponds with a statistically and economically significant increase in educational attainment of about 1.07 educational levels on average (Table 2.3),

[24] While the effects seem to be plateauing, the data also artificially truncates any larger effects for younger cohorts - since those cohorts had not yet reached schooling age at the time of the survey.

[25] At the same time that the government implemented OBC reservations, they also upheld the decision to provide reservations in job *promotions* for the SC-ST groups, and established various National Commissions for the SC and ST groups. In the early 1990s, new policies were initiated to ensure that vacant quota seats were being filled by SC-STs and that upper-caste members were not appropriating the seats for themselves. These policies may attenuate any impacts I find on OBCs.

and 1.38 years of education (Table). The education gap that existed between older cohorts of OBCs and upper-caste individuals is merely bridged by one level of education in the younger cohorts, but upper-caste individuals still get more education than OBCs on average. The ATET in the tables is the weighted average of all the $ATET_{ca}$s seen in the figures, where the weights are proportional to the cohort sizes.

2.6.1.1 Addressing Possible Concerns with Differences-in-Differences

One possible concern with the Difference-in-Differences strategy is that of violating the parallel trends assumption. By looking at the figures we can see that older unaffected cohorts do not have a trending education gap with respect to the omitted categories ($ATET_{c=OBC;a>24} = 0$), suggesting that the parallel trends assumption holds in this context. There may also be the concern of mean reversion. Since OBCs have less education than the general category, a theory of mean reversion would predict that over time this gap will fall. It is hard to see why this mean reversion should kick-in at exactly the same time as the reservation policy is implemented.[26] Nonetheless, I present evidence to show that mean reversion did not affect other disadvantaged social groups, and the other estimation strategies discussed in this paper will be unaffected by this issue of mean reversion.

Another concern arises if the omitted group is simultaneously 'treated.' Despite the fact that reservations are only applicable to OBCs, we may see a change in behavior of upper-caste members of society for various reasons. One possibility is akin to the John Henry effects discussed in the experimental literature, where the control group may react adversely because they were denied treatment. Upper-caste members of society may feel discouraged by the reservations and lower their educational attainment.[27] Another possible reaction by upper-caste members is to view these policies as increasing the competitiveness of getting a job, and thus working harder and attaining more education in order to compete for these spaces – these would attenuate the results downwards. As far as the federal law change is concerned, these reactions are unlikely since the number of government jobs were expanded to ensure that general category applicants were unaffected. Many states also expanded seats in colleges and jobs in order to accommodate the quotas and ensure that general category applicants had the same number of seats as before. However, there still existed a few states where quotas were implemented without the expansion of seats and thus this may be a concern when interpreting some of the state-wise graphs that I will present.[28]

The general equilibrium effects of such policies may also affect the interpretation of the coefficients found. Increasing the number of seats could lower the wages paid in the government jobs, which may then attenuate the ATETs found. Since the government wages are fixed (tied to inflation) for significant

[26] Furthermore, models of intergenerational transfers of human capital would predict the opposite trend.

[27] In the US context, there is no evidence of this happening (Ogbu, 2003).

[28] Furthermore, it is not clear how well the expansion of seats were handled at the federal level and could lead to additional costs, and what margin these costs would lead to other cut-backs in expenditure.

periods of time by the Pay Commissions, and for large sections of society this is the best possible job, the changes in wages may be of little concern. On the other hand, there may be peer effects in the classroom, which may affect the incentives for upper-caste students in attending school. Rao (2016), however, shows that in a different context, where Delhi public schools were required to admit poorer students, there were positive impacts on the prosocial behavior of richer peers, and only modest negative impacts on the test scores in one out of three subjects.[29]

One possible remaining concern is that of simultaneous policy changes. As discussed above, in 1986 the Indian government revamped the National Policy of Education (NPE) program and started spending on the improvement of schooling infrastructure and the building of new schools and recruited more teachers all across the country. They also expanded scholarships, provided access to adult education, and provided incentives for poor families to send their children to school regularly. This program was not OBC specific, but may still pose a problem to the double-difference identification strategy. The program will lower the costs of attending school, and may matter the most for communities that have a higher cost of schooling. The results could merely be picking up this declining gap because of this additional spending.

One of the largest expenditures was in hiring more primary school teachers under Operation Blackboard. Chin (2005) shows that despite hiring new teachers, teachers-per-school didn't increase and class sizes didn't decrease. There was merely a redistribution of teachers from larger to smaller schools. And for girls, she finds that this may have impacted the primary school completion rate in states that had a higher 'intensity' of redistribution. I re-construct the measures for the intensity of Operation Blackboard, and control for flexible forms of it in my analysis, and it doesn't affect my results (Figure 2.8.11a).

The other large policy at that time was the District Primary Education Program (DPEP), originally piloted in 1994, but expanded over the next decade expanded to other districts. Khanna (2016) uses a regression discontinuity to show the program increased education at the cutoff, but this result did not differ by caste. Figure 2.8.11c controls for a flexible polynomial of the DPEP intensity and an indicator for whether any state received DPEP funds, and the results are identical to the graphs without controls.

Lastly, these school-building policies should also affect other disadvantaged groups like the low-income upper-caste population, and the Muslim population.[30] From Table 2.8 in the Appendix, we can look at the educational attainment and per capita expenditure for the Muslims and poorer members of the upper-caste category (non-OBC/SC/ST). Both categories have mean per-capita expenditures and land assets that are *lower* than those of OBCs, and should thus be a relevant comparison group. While the poorest-fifth

[29] Other general equilibrium effects include states changing policies in light of the federal government policy change. In order to tackle this I drop the states that introduced affirmative action policies around the same in a 5-year span around the federal government policy. The results remain identical. Some states that had affirmative action policies for more than 20 years prior to the federal government change made minor changes to the amount of quotas, thus the parameter identified here may include that – giving us the policy relevant parameter that includes the inducement of minor state-level law changes. However, as I show, controlling for state-level laws, does not in any way affect the impacts of the federal level law change.

[30] Desai and Kulkarni (2008) propose a similar test – when looking at the trends in indicators for SC-ST they compare them to the trends for Muslims. They have similar levles of socio economic indicators, and also have similar geographic dispersion.

of the upper-caste category have very slightly more years of education than the average OBC; Muslims have less years of education, which would imply the possibilities of a larger impact on Muslims.

Table 2.3: Difference-in-Differences Table - Levels of Education

	Panel A:	**OBC vs. Others**	
	Younger Cohort	**Older Cohort**	**Difference**
OBC	4.279	3.673	-0.605
	(0.010)	(0.013)	(0.016)
Others	5.185	5.652	0.467
	(0.009)	(0.012)	(0.016)
Difference	-0.906	-1.978	**-1.072**
	(0.014)	(0.018)	**(0.023)**
	Panel B:	**Hindus vs. Muslims**	
	Younger Cohort	**Older Cohort**	**Difference**
Muslim	3.965	3.565	-0.400
	(0.015)	(0.021)	(0.025)
Hindu	4.527	4.293	-0.234
	(0.007)	(0.009)	(0.011)
Difference	-0.562	-0.728	**-0.166**
	(0.017)	(0.024)	**(0.029)**
	Panel C:	**Rich vs. Poorer Others**	
	Younger Cohort	**Older Cohort**	**Difference**
Others-Poorest 20%	4.059	4.420	0.361
	(0.020)	(0.030)	(0.035)
Others-Richest 80%	5.500	5.909	0.409
	(0.010)	(0.013)	(0.017)
Difference	-1.441	-1.489	**-0.048**
	(0.022)	(0.032)	**(0.039)**

Dependent variable is levels of education. Using NSS 1999-2000 data. Standard Errors in Parentheses. Difference-in-Differences value in bold. 'Others' are general category individuals (i.e. not SC, ST or OBCs). Levels of education determined by NSS.

Table 2.3 compares the Difference-in-Differences impacts on OBCs (Panel A) to the analogous impacts on Muslims (Panel B), or the poorer upper-castes (Panel C). While the impact on Muslims is statistically significant, it is economically small, being less than one-sixth of the effect on OBCs.[31] The impact on the poorest 20% of the upper-caste category is both economically and statistically insignificant. The Difference-in-Differences result therefore indicates that the policies incentivized a rise in education by 1.07 levels (approximately 1.38 years) of education on average.

[31]The slight impact on Muslims can also be explained by the fact that some Muslim groups are also categorized as OBC and could benefit from affirmative action policies

Comparing the graphs in Appendix Figure 2.8.12, we can see that the largest impact is on the OBCs. Muslims seem to experience little or no-impact, but there is a slight change in trend in the poorest-fifth of the 'Others' category many years after the policy was implemented. Nonetheless, the effect on this population is much smaller than the total impact on the OBCs. There still remains a large impact on OBC schooling that can only be explained by the affirmative action policies. If Figure 2.8.13 I restrict the sample to be only OBCs, Muslims and poorest upper-caste, and plot the differential impact on OBCs where the omitted categories are Muslims and the poorest upper-caste member. We can see that OBCs still have a substantial differential effect. Furthermore, the other identification strategies used in this paper, will not be threatened by this issue of simultaneous treatment. The figures also show little or no immediate differential impact of launching the National Policy of Education in 1986 since persons between the ages of 24 and 30 should be affected by the National Policy of Education but not by reservations.

These tables and pictures can also be produced by excluding college-goers. Artificially truncating the sample by dropping all people who have college education allows us to focus on human capital accumulation at the pre-collegiate level.[32] Looking at Figure 2.8.10, in the Appendix, the impacts on the OBC group is still starkly significant. It is natural that the impacts of these reservations lead to an increase in educational attainment even at the pre-collegiate level, since many lower-level government jobs don't require collegiate education.

On can also look for secular trends across age groups and castes in the data. Since the 1999-2000 wave was the first to ask the OBC identifier question a pre-policy analysis of this cannot be done. In the 2005 wave, one can see that the Difference-in-Differences graph is shifted about 5 to 6 years to the right, as expected (Figure 2.8.7). This may help negate fears of age-specific caste trends that kick in at exactly the age of 24. However, this result should be interpreted with caution, as between 1999 and 2005 a few more sub-castes were added to the list of OBCs, and states added more reserved seats to state-level quotas.

2.6.1.2 The Creamy Layer

The *Indira Swahney v. Union of India, 1993* case prompted the Supreme Court to exclude the relatively wealthier members of the OBC group from being eligible for these reservations. This excluded group was referred to as the 'creamy layer,' and consists of sons and daughters of people with high-ranking Constitutional Posts (the President, Supreme Court Judges, etc.), high-ranking civil service posts, and large landowners. It also excludes sons and daughters of richer members of certain occupations (doctors, lawyers, dentists, film professionals, authors, sportsmen, etc.). The members of these occupations are subject to an income test, where their annual household income must be below Rs. 100,000 (approx.

[32] Since we are look at reservations in jobs and not colleges (in this section) there is no *a priori* reason to drop college goers, other than to focus on pre-collegiate education.

\$2000) in order to be eligible for reservations.[33]

Using the NSS Labor Force Survey data, I can identify certain occupational groups and industrial sectors, and classify persons as whether they should be classified as 'creamy layer' or not. Then using the income-information in the Labor Force Survey, I construct the total household income for adults. However, this constructed measure will be far from perfect as (a) the labor force survey only identifies broad occupational groups and not the specific occupations, and (b) persons close to the income cutoff may find it easy to manipulate their bank statements and income tax returns, in order to qualify for reservations.[34] Therefore, the creamy layer indicator will be at best, a close approximation of whether the persons took advantage of these policies or not.

Table 2.4 produces the Difference-in-Differences tables for the creamy layer and non-creamy layer groups separately. While there is some impact on the creamy layer group – which could be a result of income-reporting manipulation, or other ways of getting around the eligibility criteria – the impact on the non-creamy layer group is more than double the size than that of the creamy layer group. The triple -difference estimator is the difference between the two double difference estimates in the table, and is a statistically and economically significant 0.614 years of education. One can also run the triple-difference regression (results shows in Appendix table 2.11), which produces the same estimate. The tables therefore show that the bulk of the impact was on the non-creamy layer households.

2.6.1.3 Transition Between Education Levels and Quantile Effects

While the Difference-in-Differences estimate shows that on average, there was an increase of about 1.4 years of education for OBCs, it says little about the transition between the different levels of education. In being eligible for government jobs, these levels of education are important milestones in the qualification criteria. In order to see how the transition takes place, one can make Difference-in-Differences tables for each level of education (using the highest attained grade as a 1/0 indicator). For example, looking at Secondary School grade attainment in Table 2.9, it can be seen that only 8.7% of older OBCs had secondary school as their highest grade attained, whereas this number is 14.6% of the older individuals in the upper-caste category. The difference-in-difference coefficient (0.0281) shows that there is a relative (to the upper-caste group) transition of the OBCs into having secondary school as their highest grade attained.

These tables can be produced for every level of education to look at the relative transition of OBCs in and out of certain grade levels. The difference-in-differences coefficients for each grade level are reproduced in Table 2.10. These tables were also made for the sample excluding college goers, by artificially truncating the data and dropping all college-goers, in order to focus on transitions in the pre-

[33]Since then this threshold has been raised and now stands at Rs. 600,000 (approx. \$12000).

[34]Furthermore, the law stipulates that the income criteria will be applicable to 'household' income, where the definition of 'household' is also subject to manipulation.

Table 2.4: Years of Education: Creamy Layer vs. Non-Creamy Layer

Panel A:	**Creamy Layer**		
	Old	**Young**	**Difference**
OBC	10.674	7.263	-3.412
	(0.13)	(0.16)	(0.21)
Others	11.734	7.711	-4.023
	(0.06)	(0.08)	(0.10)
Difference	-1.060	-0.448	0.612
	(0.14)	(0.18)	(0.22)
Panel B:	**Non-Creamy Layer**		
	Old	**Young**	**Difference**
OBC	4.046	4.415	0.370
	(0.02)	(0.01)	(0.02)
Others	6.406	5.523	-0.883
	(0.02)	(0.01)	(0.02)
Difference	-2.360	-1.108	1.253
	(0.02)	(0.02)	(0.03)
Triple Difference			**0.641**
			(0.274)

Dependent variable is years of education. Standard Errors in parentheses. Panel A consists of 9133 observations and Panel B has 370500 observations. Households with no income or occupational information are excluded. The Triple Difference estimate is the difference between the two double difference estimates: 0.614 years of education. The triple difference estimator has standard errors clustered at the state-level.

collegiate level. The table indicates that the relative transition of OBCs before and after the policy, has been away from illiteracy (and away from below primary and primary levels of education) and into secondary school and college.[35]

The policy allows for reservations at any of the four classes of public sector jobs. While on average students are incentivized to get another level of education, there may be parts of the distribution where students get less education and drop out early to get a lower level public sector job. Figure 2.8.9 shows

[35]Between 2001 and 2002, the government tried to implement policies to universalize elementary education and the Millennium Development Goals, but the effects of these policies are not being captured here since the dataset was collected before the MDG projects were implemented.

the quantile treatment effects. While there are a few quantiles where the treatment effect is negative, it is mostly always positive. Furthermore, the biggest jump in the distribution is at the primary-level and college levels.

2.6.2 The Intensity of Treatment and State Level Variation

Since different states have, over time, passed different laws reserving state-level seats and jobs for OBCs, this kind of analysis can be done for each state separately. In the graphs in Appendix Figure 2.8.15, I perform a meta-analysis of all the state-law changes, where the vertical red line represents a marked change in reservation policy for the OBC group in that state. By restricting the sample to the corresponding state and plotting the coefficients, we can see that the state-wise changes in reservation policy have impacts similar to the federal law change.

This cross-state meta-analysis can also be used to study how the difference-in-difference treatment effect varies by the intensity of reservations, and can address the issues of mean reversion and simultaneous policy changes mentioned above. While variation in social group and age were exploited in the difference-in-differences section, it is possible to investigate another dimension of variation: 'the intensity of reservation policy.' Since each state has its own reservation policy, there is variation in terms of which states are more pro-reservations and which are less so. Let us define the 'intensity of reservation' as the ratio between the percentage of quotas and the population percentage for each group: $\frac{quotas\%}{population\%}$. For instance, in the state of Karnataka, this ratio is $\frac{53}{36} = 1.47$, whereas in the state of Madhya Pradesh it is only $\frac{13}{40.5} = 0.32$, thus making the intensity higher in Karnataka than in Madhya Pradesh.

In the following regression specification, edu_{ics} is the education level obtained by a person i belonging to caste c and residing in state s. OBC is an indicator for whether the person is OBC or not. Most states made significant changes to reservation policies in the early 1990s.[36] The variable $young$ equals 1 for cohorts that were still in school or will attend school after the changes in reservation policy have been implemented. $\boldsymbol{Z\beta}$ is a vector of controls.[37] The parameter γ captures how the ATET varies with intensity of treatment. It should be positive in sign if the older members of the reserved castes have relatively less education than the younger members, and this disparity should be higher in states that had larger changes to the intensity of reservations.

[36] The reason that the law changes from the early 1990s are used (as opposed to previous changes) is because the federal law also changed at that time. The federal law change should not differentially impact residents of different states because people are competing for federal seats with people all over the country. If state-law changes from periods both before and after the 1990s were studied, then they would be confounded by other changes like the federal law change.

[37] I present results with and without controls, where the controls include the intensity of SC and ST reservations, and the interactions with the young indicator and SC ST indicators – this is a fully saturated model.

$$edu_{ics} = young_i + intensity_{cs} + OBC_c + \beta_{0c}OBC_c * young_i + \beta_{1cs}intensity_{cs} * young_i + \beta_{2cs}OBC_c * intensity_{cs} + \gamma OBC_c * intensity_{cs} * young_i + \boldsymbol{Z\beta} + \epsilon_{cs} \quad (2.4)$$

This specification is akin to a continuous form of the triple-difference estimator, where the three dimensions of variation are age, caste and intensity of reservations. As Gruber (1994) explains, such an approach allows us to control for caste-specific trends ($\beta_{0c}OBC_c * young$), and state specific trends in laws ($\beta_{1cs}intensity_{cs} * young$). Controlling for these trends allows us to tackle the issue of simultaneous timings of policy; there is no reason to believe that the state-wise intensity of reservations should be correlated with investments in schooling infrastructure. Furthermore, this method also allows us to control for state-specific caste preferences $\beta_{2cs}OBC_c * intensity_{cs}$, since certain states may care more about certain castes, and the intensity variable would then be picking up these preferences. Last, this approach also solves the automatic mean-reversion problem, since there is no reason to believe that non-policy driven mean-reversion should be higher in states that have more favorable reservation policies than others.[38]

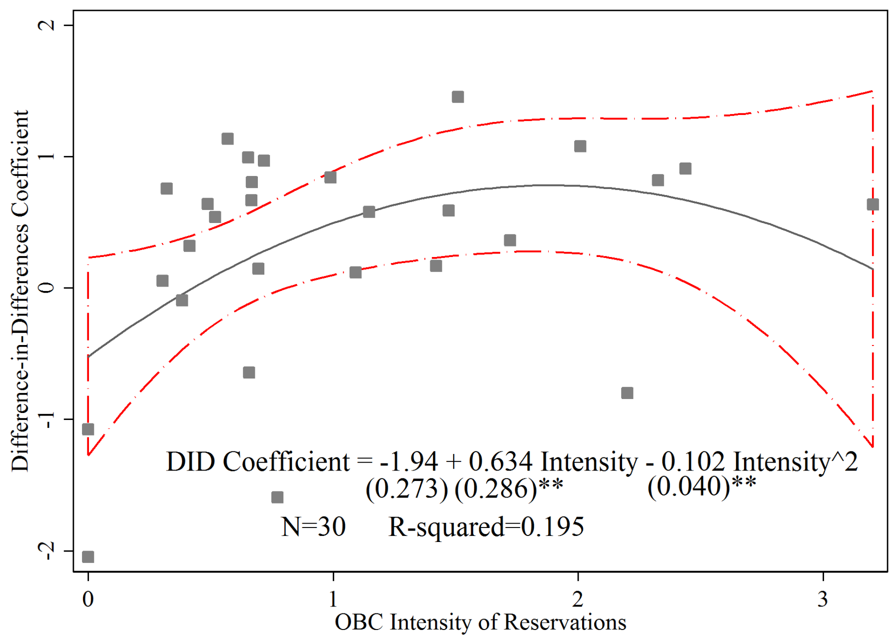
Table 2.5: The Intensity of Treatment and State Level Variation

Education Level	Full Sample	Non 0 Intensity	SC-ST Controls
OBC*Intensity* Young	0.182	0.178	0.168
Standard Errors:			
State Level	(0.0985)*	(0.0983)*	(0.0957)*
Region Level	(0.0788)**	(0.0788)**	(0.0780)**
SC-ST Controls	N	N	Y
Observations	508,410	503,981	508,410
R-squared	0.026	0.027	0.087

Dependent variable is level of education. Standard Errors in parentheses clustered at the state-level (30 states) or region level (77 regions). Specification 'Non 0 Intensity' drops the 2 states that have no reservations for OBCs. Controls in all specifications include an OBC indicator, a young indicator, the state-level intensity of reservations, and all double interactions between these variables. The final column that has SC-ST Controls also include an SC indicator, an ST indicator, state-level intensities of SC reservations, state-level intensity of ST-reservations, a young indicator and all double interactions between these variables.

[38] It may be interesting to see if the 'intensity' variable is correlated with the minority group's situation in society. If greater socio-economic disadvantage in the state is positively correlated with more intensity, then the treatment effect will probably be larger (since there is potentially a larger gap to bridge). If however, more advanced minority groups, can (say via political power) influence greater 'intensity', then the treatment effect will potentially be smaller. I find no evidence of such heterogeneity and no significant correlation between intensity and the baseline socio-economic characteristics of the minority groups.

Figure 2.2: Relationship Between Treatment Effect in a state and state-level Intensity of Quotas



Auxiliary regression of relationship between the ATET and intensity of OBC reservations by state.

In Table 2.5 we can see that the effect of affirmative action policies is larger in states that have a higher intensity of reservations. An increase in the intensity by one unit, increases the treatment effect of these policies by between 0.17 and 0.18 levels of education for OBCs. This coefficient is stable across specifications with and without additional controls.

This regression specification, however, imposes a linear functional form. In order to explore non-linearities in the relationship between intensity and the effect of reservations, I use a method proposed by Donald and Lang (2007). This method also tackles the issue of having a small number of clusters. I use a two-stage estimation procedure by first computing the treatment effect for each state, and then regressing that treatment effect on the intensity of reservations. In order to find the treatment effect in each state, I do a simple difference-in-differences using only the sub-sample of each state. I then plot the difference-in-differences coefficient across the intensity of reservation by each state in order to find the relationship. In Figure 2.2, I plot the relationship and display the auxiliary regression that captures this relationship – which is increasing at a decreasing rate.[39]

[39]Figure 2.2 drops outlier states that have very large intensity values because of almost non-existant OBC populations. These states are amongst the smallest in the country (Goa and Mizoram).

2.6.3 Regression Discontinuity and Difference-in-Discontinuities

For my final empirical strategy, I exploit a state-determined methodology of identifying/classifying OBCs, and obtain a Regression Discontinuity (RD) estimate of the the impacts of reservations. Such an analysis is new to the literature, and provides a causal impact of affirmative action policies. The biggest advantage of an RD estimate is that it is not encumbered by issues such as mean reversion and simultaneity of government policy. Government spending on school infrastructure should have uniform impacts on castes just below and above the cutoffs determined by the eligibility methodology. Thus there should be no confounding effects of the government's investment in schooling program. There is also the benefit of identifying a different and interesting parameter – the effect of such policies on a student from the $marginal$ sub-caste.

Classification and identification of OBCs for state-level reservation policies is the prerogative of the state government. States appoint committees to determine who the OBCs are and what reservations they should be eligible for. Some committees conduct a socio-economic survey and collect data. They use this data to rank the different sub-castes on the basis of socio-economic indicators. Castes above a certain cutoff of 'backwardness' are eligible for reservations. This set-up allows us to estimate the impacts of the reservation policy using a regression discontinuity design. If we have information on the index of 'backwardness,' we can compare sub-castes just above to those just below the cutoff to see what the causal impacts of reservations are. The analysis in this section will focus on the state of Haryana, which had one such methodology for classifying the OBCs.

The RD can then be aided by an additional source of variation – once again, certain cohorts were too old at the time the policies were implemented to be affected by these reservations. I then perform a difference-in-discontinuities analysis, using the sub-caste index to identify the discontinuity, and the age cohorts to identify the difference in the discontinuities for each cohort.

In the state of Haryana, an 'index of backwardness' for each sub-caste is published, making it possible for me to conduct a sharp RD. The Singh (1990) Haryana Backward Classes Commission Report was the first ever committee in the state. Being the first is an added bonus, since it prevents any lingering policies from contaminating the before-after analysis.[40] The Committee conducted a survey and created a score out of a total score of 60. Any caste that had more than half the total score was considered an OBC. A half-way mark is an intuitive cut-off point and it is thus unlikely that the cut-off itself was manipulated to include certain castes. It is also unlikely for people of different castes to manipulate their score as the index is based on survey data where the respondents were probably unaware of the utilization purpose of this data. I observe no bunching of castes just above the cutoff.[41] Manipulation of the methodology from

[40]The handful of other states that used similar methodologies had lingering policies; furthermore, the other states don't publish the tables used to formulate the index.

[41]This would not be a valid way of testing manipulability if there were certain groups that wanted to move in opposite directions. But since it is reasonable to believe that the marginal caste wants to be eligible for reservations, in the presence of manipulation we should see bunching just above the cutoff.

the government's side is also unlikely, since they use the same methodology formulated by the Mandal (1980) Federal Commission. Lastly, I test that the treatment is discontinuous at the cutoff and all other baseline characteristics vary continuously.[42]

Singh (1990) identifies the OBCs by creating an index of 'backwardness' based on (a) social, (b) educational, and (c) economic disadvantage. The social disadvantage criterion looks at 10 indicators, including employment in manual labor, the unorganized sector, and lack of access to proper sanitation and other civic amenities. The educational criterion studies 10 other indicators related to drop-out rates, female literacy, test scores and vocational education. And the economic index looks at 15 indicators such as family assets, consumption expenditure, maternal mortality rates, unemployment rates, etc. The survey was done in 53 villages and 4 towns, and the report produces caste-wise tables on each of the 35 indicators used in the final index. From the raw data tables, I can reconstruct the index and it matches the final index produced. While the RD analysis is done for only one state, it confirms a powerful causal finding on affirmative action.

The data set used for the RD analysis is the ARIS-REDS 1999 data set. The nationally representative NSS data cannot be used since it doesn't have disaggregated sub-caste categories, which we require for the RD analysis. Unlike the NSS data set, ARIS-REDS collects information on years of education rather than levels. Therefore, the results in this section analyze the impact of affirmative action policies on the years of education. In the RD results in Figure 2.3, the dependent variable is the difference in mean years of education between the older and younger members of that caste.[43] Once again 'older' is defined as being too 'old' to enjoy the benefits of this reservation policy. There are 27 sub-castes for which the ARIS-REDS and the Haryana Committee Report have matching caste names.[44]

If, however, we look at the mean education level of the population that is too old to be affected by the reservation policy, we see no discontinuity at the cutoff (Figure 2.4). Furthermore, we can see a slight downward trend, since a higher index indicates a larger socio-economic disadvantage. Looking at regression Table 2.17 in the Appendix, we can see that the cutoff isn't significant for the educational attainment of the older population.

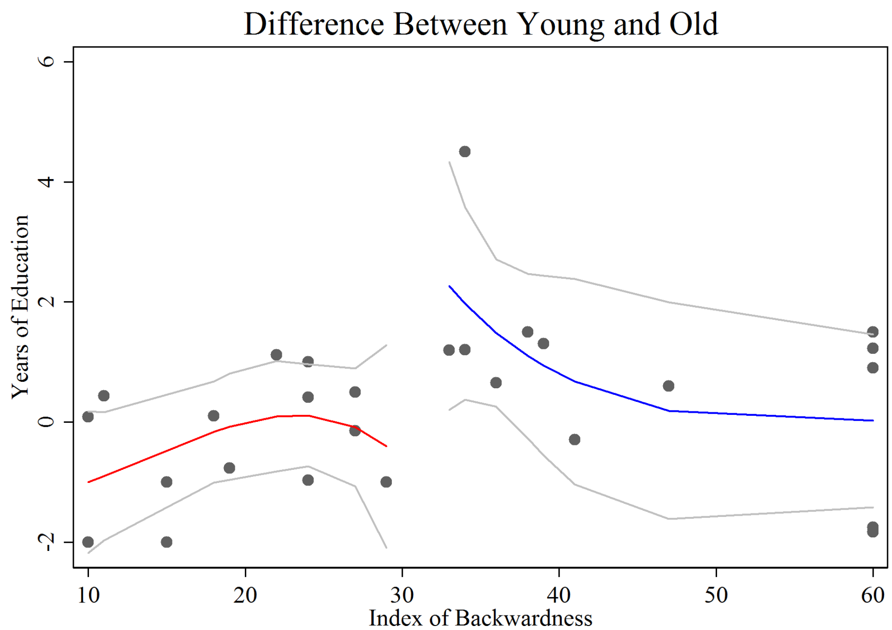
Based on the RD literature, I explore various regression specifications. Imbens and Lemieux (2008) suggest changing the bandwidth and seeing if the results are robust to restricting the sample to a small area just around the cut-off, whereas van der Klaauw (2008) suggests using a semi-parametric approach of local linear regressions with higher order polynomials. I use higher order polynomials of the index

[42] In 1995, the Ramji Lal Committee – the second backward classes commission in Haryana – added 4 more castes to the list. In the dataset used, this adds two castes below the cutoff to be eligible for reservations. However, since these castes will have only felt the benefits for less than 3 years (the data was collected in 1999), they have been coded as ineligible. Doing so doesn't change the results.

[43] Dependent variable = mean education of young in subcaste c − mean education of old in subcaste c.

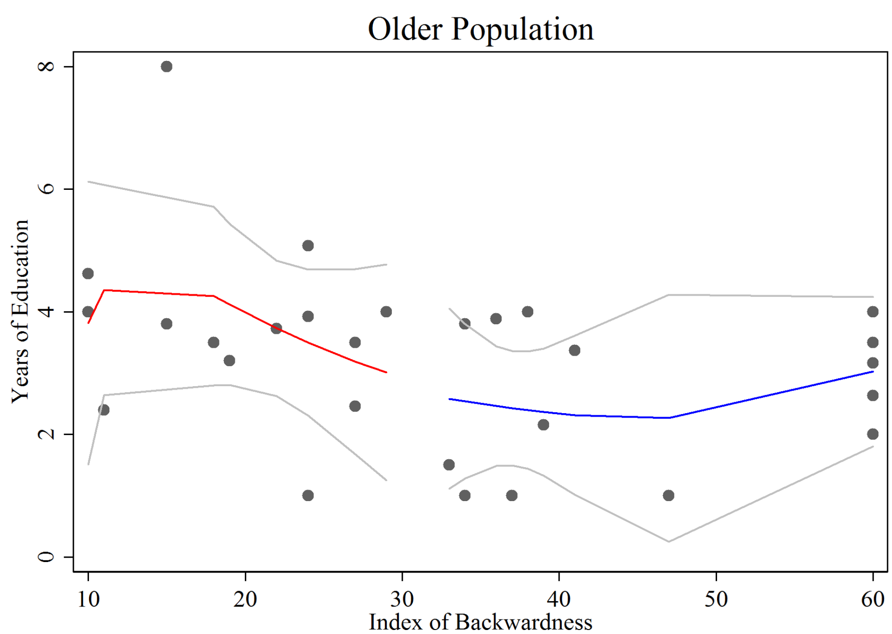
[44] While 34 of the ARIS-REDS castes could be matched to the names in the Committee Report, there wasn't any data on 7 of these castes in the data-set – since the data set is much smaller than the NSS. Furthermore, the Scheduled Castes were assigned highest possible index values as they were already eligible for these policies (and are therefore not castes near the threshold) – this shows up in the bunching at the highest possible index value of 60.

Figure 2.3: RD: Average change in years of education by caste



Regression Discontinuity on d_{ed}, defined as: mean years of education young$_c$ $-$ mean years of education for old$_c$.

Figure 2.4: RD: Mean Education by Caste of Older Population



Sample restricted to population above the age of 27 (too old to benefit from the reservation policy). Regression Discontinuity on years of education.

value, in the spirit of the Heckman and Robb (1985) 'Control Function Approach,' and sub-samples restricted around the cutoff, in the specifications below. Column headings have the index's degree of polynomial order, and if 'Restricted' is mentioned, then the sample only includes half the index-span around the cutoff (index values of 15 to 44). 'Flex Slope' indicates that the slope is allowed to vary on either side of the cutoff.[45]

The RD framework identifies the localized treatment effect in the neighborhood around the cutoff. Students from sub-castes that are close to the cutoff may behave differently from the average student. The coefficient of interest is the Neighborhood Average Treatment Effect (NATE) since it is the average impact of the policies in the neighborhood around the cutoff of eligibility.[46] In this context, there is a large mass of students who get no schooling whatsoever. These students presumably belong to sub-castes further away from the cutoff, having the highest levels of 'backwardness.' This NATE therefore may be larger than the ATET found via the Difference-in-Differences methodology since the ATET is pulled down by students who have the highest costs of schooling. The students that are far from the RD cutoff with high values of 'backwardness,' would presumably only respond to extremely large changes in the returns to education to first budge them on the extensive margin of attending school. Since a large number of high-cost students will not increase their schooling despite these policies, the ATET may be lower than the NATE.

I use three distinct regression specifications. The first – a Discontinuity-in-Differences – is what Figure 2.3 plots, where the dependent variable is the mean difference in education between the younger and older members within a caste – this regression is at the sub-caste level. The second, restricts the sample to only the young, and estimates the discontinuity for the young sample. The third – a Difference-in-Discontinuities – combines a difference-in-differences approach with the RD approach to estimate the differential discontinuity for the young.

In my first, approach – a Discontinuity-in-Differences – I control for a flexible polynomial of the index $f(index)$ and estimate the following regression:

$$\overline{ed\ young}_c - \overline{ed\ old}_c = \beta_0\ \mathbb{1}_{index>0} + f(index) + \epsilon_c \tag{2.5}$$

In Appendix Table 2.12, it is clear that despite having a small number of observations, the coefficient of interest is both economically and statistically significant. Looking at the third order polynomial column, the coefficient shows that the causal effect of reserving seats for backward classes is to increase their high-school education by about 2.6 years.

In my second approach, I restrict the sample to young cohorts who would be able to change their school-

[45] Data-driven bandwidth selection procedures, like the one discussed in Calonico et al. (2014b) cannot be used in this context as there are only twenty-seven mass points of the index.

[46] Another way to think about this parameter is to think of it as a weighted average treatment effect, where the weights are the probability that each caste's assignment value lies in the neighborhood of the threshold (DiNardo and Lee, 2011).

ing decision, and control for the mean education level of the older population in that caste. I cluster the standard errors at the caste level, and also show the p-values with a Wild-t small-cluster bootstrapping procedure (Cameron et al., 2008). The added advantage of just performing the regression for the younger population is that we can make sure that the discontinuity isn't merely arising out of the older populations education, and reconfirm the results ins Figures 2.3 and 2.4. In some specifications, I restrict the sample to index values around the cutoff. The regression of interest is:

$$edu_{ic} = \beta\ \mathbb{1}_{index>0} + f(index) + \overline{ed\ old}_c + \epsilon_{ic} \quad for\ \{-v_1 < index < v_2\} \tag{2.6}$$

Table 2.14 shows similar coefficients as before: the causal impact of reservations is to increase years of high-school education by about 3 years for the average student in the neighborhood of the cutoff.

My preferred specification, however, is a Differences-in-Discontinuities approach, which is relatively new to the literature (Grembi et al., 2016). This incorporates the discontinuity along the caste index, and differences it across the older and younger age cohorts. In Equation (2.7), I interact the cutoff with the variable $Young$. This interaction term should have a positive sign since the younger group will benefit from the reservation policy. When $Young = 0$, the discontinuity should be insignificant (as seen in Figure 2.4 and Table 2.17), but when $Young = 1$ those above the 'backwardness' cutoff should increase their educational attainment. Therefore, the model predicts that the coefficient β_1 will be positive:

$$edu_{ic} = \beta_0\ \mathbb{1}_{index>0} + \beta_1\ \mathbb{1}_{index>0} * Young_i + \beta_2\ Young_i + f(index, Young_i) + \epsilon_{ic}\ for\ index \in \{-v_1, v_1\} \tag{2.7}$$

Table 2.6: Difference in Discontinuities

Polynomial	β_1	Clustered SE	CGM pval	Flex Slope	Restricted Sample	R-squared
3rd	1.892***	(0.725)	(0.100)			0.104
3rd	1.986*	(1.056)	(0.150)		X	0.103
4th	1.681***	(0.627)	(0.0300)			0.105
4th	1.156	(1.267)	(0.358)		X	0.108
5th	2.307**	(0.999)	(0.0360)			0.105
5th	1.850	(1.892)	(0.470)		X	0.111
1st	1.832**	(0.724)	(0.104)	X		0.106
1st	1.418**	(0.647)	(0.138)	X	X	0.103
2nd	0.988***	(0.211)	(0)	X		0.105
2nd	1.603***	(0.467)	(0.004)	X	X	0.113

Dependent variable is years of education. Level of significance (based on clustered standard errors): *** 0.01; ** 0.05; * 0.1. 'Restricted Sample' consists of half the index span around the cutoff (index values 15 to 44). 'Flexible slope' allows the slope of the regression specification to vary on either side of the cut-off. Standard Errors clustered at the Caste-level, and Cameron et al. (2008) small-cluster Wild Bootstrap p-values also presented.

Table 2.6 shows an increase in education for the OBCs. Once again, the results are economically and statistically significant and of a similar magnitude, giving us a consistent story across all the different possible specifications. The average value for the coefficients is about 2.22 years of education. This last specification is the preferred specification, since it utilizes the entire data set and has the highest power.

I conduct numerous robustness checks to validate these results. Tables 2.13 to 2.19 check the robustness to the age-cutoff by dropping three years before and after the used cutoff. The results are stronger than before. Tables 2.15 and 2.18 drop any person who ever attended college. It is important to check that other factors are not discontinuous at the same cutoff, since the RD design requires all other factors (income, assets, etc.) to vary smoothly at the cutoff. Looking at Table 2.21 in the Appendix, we see that there are no other significant discontinuities at the same threshold (an index value of 30). The table presents results on total expenditure, medical expenditure and work in casual labor, but robustness checks were done on other variables as well. We can also look for educational discontinuities at any other value of the index. When studying the impact on the average change in education between the younger and older cohorts, no other values of the index have significant discontinuities. The only value that the discontinuity is visible, is at the true cutoff (a value of 30 on the index). Figure 2.8.16 shows cutoffs at 20, 25, 35 and 40 – none of which have significant discontinuities.

2.6.4 Summary of Empirical Section

So far this paper has used three different approaches to answer the primary question of interest: does affirmative action incentivize students to stay in school? The double-difference strategy uses a nationally representative dataset to identify the average treatment effect on the treated ($ATET_{ca}$) for each age cohort a and caste c. Exploiting variation in caste and age, the estimator finds that the students eligible for reservations increased educational attainment. This was found to be true in response to the change in federal law, as well as various states' changes in law. The concern of confounding policies – like large expenditures on school building – was tackled by looking at placebo groups that should have been affected by school building but not by reservations. There wasn't an impact on educational attainment for the Muslim population and the low-income high-caste population, suggesting that the bulk of the impact on lower-caste members was due to the affirmative action policies. I also compare the differential impact between the excluded creamy layer and included non-creamy layer populations – the non-creamy layer population was seen to have more than twice the impact than that of the creamy layer group.

The state-wise analysis exploits variation on a third front: the intensity of reservation policy. The paper finds that the impact is larger in states that have a more generous reservation policy. Using a 2-stage estimation procedure, I show the treatment effect of quotas is increasing, at a decreasing rate, with the intensity of reservations.

The regression discontinuity approach looks at the introduction of OBC reservations in the state of Haryana. OBCs were classified on the basis of an index of 'backwardness,' which provides the running variable for the RD. Across specifications, I consistently find an increase in educational attainment for the population eligible for reservations, and having no discontinuities in other dimensions. While the RD is the most 'internally valid' of all the identification strategies used here, it could be less externally valid than the other strategies that use the nationally-representative data since it focuses on the state of Haryana. Haryana is similar to most North Indian states, but may be quite different from some South Indian states. In Appendix Table 2.20 the means of the major variables are studied comparing Haryana with the Rest of India. The one stark difference is that Haryana has virtually no Scheduled Tribes (STs). The other differences are economically insignificant: the average Haryanvi is richer than the average Indian by about $1.1 a month, is half a year older, and has about one-tenth more levels of education. This differences are small, and not statistically significant, suggesting that it is a representative North Indian state.[47]

The Difference-in-Differences estimator identifies the Treatment on the Treated (ATET) of about 1.38 years of education. Whereas the NATE from the RD strategy is somewhere around 2.2 and 3.1 years of education, the average impact is only about 1.38 years of education.[48] The NATE may be the relevant estimate of interest if the government is considering adding another sub-caste to the list of OBCs; whereas the ATET may be the relevant estimate if the government wants to know the overall impact of changing the amount of quotas for all OBCs. The state-level intensity variation tells us how increasing the intensity of quotas may affect the magnitude of the treatment effect.

2.7 Conclusion

Using three different empirical strategies, I find that affirmative action policies encourage students to increase their education by one more schooling-level. This has major implications as it indicates the possibility of encouraging students just below the threshold of a certain level of education to cross that threshold and get to the next highest level. Similarly, Shreshta (2016) finds that the Gurkha community in Nepal attains one more year of education, on average, in response to a change in the mandatory employment eligibility law of the British army. However, it is hard to translate the results in most of the other literature into years of educational attainment. Kazianga et al. (2013) show that enrollment rises by 20% when schools are made more girl-friendly, and Dinkelman and Martinez (2014) show that absenteeism falls by 14% when information is provided about financial aid. Using the booming IT sector near Delhi as a sign of an increase in returns to education, Oster and Steinberg (2013) find that school enrollment rises by 4% to 7% when a new IT center is opened in the area. On the other hand, Jensen and Miller (2015), and de Brauw and Giles (2008) show that schooling investments may actually fall in response to higher returns.

[47] Nonetheless, South Indian states are culturally different from Haryana and have a history of reservations unlike Haryana.

[48] However, it is important to remember that the RD and Difference-in-Differences are looking at the impacts of different policies: the difference-in-differences looks at the impact of reservations only in governmental jobs and not colleges.

Contrary to the expectation of 'complacency' effects, I find that lowering standards may actually have some positive incentive effects. There is, however, a non-linearity – very high levels of reservations may lead to a 'patronizing equilibrium' (Loury, 1992). Furthermore, a patronizing equilibrium may get strong with time as information updating takes place. While the model may be generalizable to other contexts, like the US, Sri Lanka, Malaysia, the empirical results would probably drastically differ by context. For instance, in the US, affirmative action lacks the backing of certain salient features – like explicitly reserving seats for certain groups – found in the Indian context. Furthermore, interventions in the US are minor in size, compared to a 27% reservation of seats in all government jobs. This lack of salience, and difference in policy-size may lead to different results.

The policies may also come with certain costs if there is a crowding out of educational attainment for upper-caste members. The government has tried to mitigate this concern by increasing the seats in colleges and government jobs so that the absolute number of seats available to the upper-castes does not change, but it is not clear what the possible costs of increasing seats are. Furthermore, such a large policy is sure to have other general equilibrium effects in terms of the work-force composition and composition of classrooms. Negative peer-effects, due to these changing compositions, may play a role, Although in such a context they may be less of a concern – Rao (2016) finds that interaction with poorer students actually encourages more prosocial behavior with small negative impacts on the test scores in only one subject.

In terms of the benefits for the minority group however, an increase in 2.2 years of education can translate into high wage gains since the estimated returns to education in developing countries vary between 7% and 14% (Behrman, 1999; Duflo, 2001; Khanna, 2016; Psacharopoulos and Patrinos, 2004; Strauss and Thomas, 1995). There are also various non-pecuniary benefits of education, like greater participation in the political process and better health. Lastly, lowering educational inequalities – and possibly wealth inequalities – may be something intrinsically valuable to policy-makers. In light of these results, therefore, policy-makers should keep in mind the externalities of such affirmative action policies.

2.8 Additional Tables and Figures

Table 2.7: Difference-in-Differences Table – Years of Education

OBC vs. Others	Younger Cohort	Older Cohort	Difference
OBC	4.439	4.129	-0.310
	(0.013)	(0.016)	(0.020)
Others	5.579	6.654	1.076
	(0.013)	(0.016)	(0.020)
Difference	-1.140	-2.526	**-1.386**
	(0.018)	(0.022)	**(0.029)**
Hindus vs. Muslims	**Younger Cohort**	**Older Cohort**	**Difference**
Muslim	4.031	3.961	-0.071
	(0.018)	(0.025)	(0.030)
Hindu	4.762	4.952	0.190
	(0.009)	(0.011)	(0.014)
Difference	-0.731	-0.992	**-0.261**
	(0.021)	(0.030)	**(0.036)**
Rich vs. Poorer Others	**Younger Cohort**	**Older Cohort**	**Difference**
Others-Poorest 20%	4.111	5.083	0.972
	(0.025)	(0.037)	(0.044)
Others-Richest 80%	5.989	6.983	0.993
	(0.014)	(0.017)	(0.023)
Difference	-1.879	-1.900	**-0.022**
	(0.030)	(0.041)	**(0.051)**

Using NSS 1999-2000 data. Standard Errors in Parentheses. Difference-in-Differences value in bold. 'Others' are general category individuals (i.e. not SC, ST or OBCs).

Table 2.8: Social Groups and Religions

	Years of Education	Land owned (acres)	Per Cap Month Exp (Rs.)
ST	3.40	1.17	427.32
SC	3.04	0.39	398.67
OBC	3.90	0.99	446.33
Others	5.68	1.10	519.02
Richest 80%	6.07	1.14	578.00
Poorest 20%	4.10	0.96	283.17
Hindu	4.46	1.01	465.57
Muslim	3.61	0.49	424.17

'Others' are general category individuals (not SC, ST or OBCs). Nominal exchange rate: approx Rs. 50 to $ 1. Household Monthly Expenditure deflated by rural-urban-region-wise CPI.

Table 2.9: Proportion of Students with Secondary School

Secondary School	Old	Young	Difference
OBC	0.0877	0.0736	-0.0141
	(0.001)	(0.001)	(0.001)
Others	0.1463	0.1041	-0.0422
	(0.001)	(0.001)	(0.001)
Difference	-0.0586	-0.0305	**0.0281**
	(0.001)	(0.001)	**(0.002)**

Standard Errors in Parentheses. Difference-in-Differences value in bold. 'Others' include that section of the population ineligible for reservations (i.e. not SCs, STs, or OBCs)

Table 2.10: Relative Transition of OBCs between grades

	Difference-in-Differences Coefficient	
Level of Education	**Including College**	**Excluding College**
Illiterate	-0.0862	-0.0661
	(0.003)	(0.003)
Below Primary Education	-0.0132	-0.0083
	(0.002)	(0.003)
Primary Education	-0.0149	-0.0065
	(0.002)	(0.002)
Middle School	0.0077	0.0217
	(0.002)	(0.002)
Secondary School	0.0281	0.0455
	(0.002)	(0.002)
Higher Secondary School	0.0047	0.0139
	(0.002)	(0.002)
College Graduate	0.0741	
	(0.002)	

Standard Errors in Parentheses. Levels of education determined by NSS 1999-2000. Sample of 'excluding college' drops all people with at least some college education.

Table 2.11: Creamy Layer v Non Creamy Layer

VARIABLES	Years of Education
OBC	-2.360***
	(0.297)
SC-ST	-3.392***
	(0.274)
Young	-0.883***
	(0.107)
Creamy Layer	5.328***
	(0.155)
OBC*Young	1.253***
	(0.0944)
SC-ST * Young	1.682***
	(0.125)
OBC* Creamy layer	1.301***
	(0.305)
SC-ST * Creamy Layer	1.647***
	(0.364)
Young* Creamy Layer	-3.140***
	(0.170)
OBC*Young* Creamy Layer	**-0.641****
	(0.274)
SC-ST*Young* Creamy Layer	-0.943***
	(0.297)
Constant	6.406***
	(0.150)
Observations	521,063
R-squared	0.093

Dependent variable is years of education.
Standard Errors in Parenthesis
Level of significance: *** 0.01; ** 0.05; * 0.1

Table 2.12: Discontinuity in Differences

Polynomial	3rd	4th	5th	1st	2nd
Cutoff	2.604*	2.342	4.105**	5.158**	18.07
SE	(1.347)	(1.387)	(1.959)	(2.031)	(12.83)
Nonparametric bootstrap	0.0605	0.198	0.0505	0.0252	0.359
Wild-t p-value	(0.0560)	(0.148)	(0.154)	(0.0280)	(0.196)
Flex Slope				X	X
R-sq	0.400	0.427	0.439	0.381	0.417

Dependent variable is years of education. Standard Errors in Parenthesis. Level of significance : *** 0.01; ** 0.05; * 0.1. Nonparameteric bootstrapped p-values, and Wild-t p-values presented. Regressions consist of 27 sub-castes in the state of Haryana. Flex-slope specifications allow the slopes to vary across the cutoff.

Table 2.13: Discontinuity in Differences – No 24 to 30 year olds

Polynomial	3rd	4th	5th	1st	2nd
Cutoff	3.274**	2.908**	3.862*	5.716**	21.25
SE	(1.256)	(1.285)	(1.947)	(2.183)	(12.57)
Nonparametric bootstrap	0.0203	0.0860	0.533	0.009	0.294
Wild-t p-value	(0.0200)	(0.0540)	(0.140)	(0.0220)	(0.136)
Flex Slope				X	X
R-sq	0.400	0.427	0.439	0.381	0.417

Subsample of people no 24 to 30 year olds. Dependent variable is years of education. Standard Errors in Parenthesis. Level of significance : *** 0.01; ** 0.05; * 0.1. Nonparameteric bootstrapped p-values, and Wild-t p-values presented. Regressions consist of 27 sub-castes in the state of Haryana. Flex-slope specifications allow the slopes to vary across the cutoff.

Table 2.14: Young Sample

Polynomial	β_1	Clustered SE	CGM pval	Flex Slope	Restricted Sample	R-squared
3rd	2.640***	(0.747)	(0.050)			0.108
3rd	2.900**	(1.192)	(0.108)		X	0.120
4th	2.393***	(0.738)	(0.022)			0.109
4th	1.727	(1.261)	(0.266)		X	0.121
5th	3.004***	(1.103)	(0.078)			0.109
5th	2.728*	(1.570)	(0.206)		X	0.122
1st	0.520	(0.593)	(0.434)	X		0.106
1st	6.952***	(2.578)	(0.066)	X	X	0.119
2nd	17.66**	(7.477)	(0.194)	X		0.109
2nd	-30.52	(32.15)	(0.438)	X	X	0.121

Dependent variable is years of education. Level of significance (based on clustered standard errors): *** 0.01; ** 0.05; * 0.1. ‘Restricted Sample’ consists of half the index span around the cutoff (index values 15 to 44). ‘Flexible slope’ allows the slope of the regression specification to vary on either side of the cut-off. Standard Errors clustered at the Caste-level, and Cameron et al. (2008) Wild Bootstrap p-values also presented.

Table 2.15: Young Sample – No college

Polynomial	β_1	Clustered SE	CGM pval	Flex Slope	Restricted Sample	R-squared
3rd	3.113***	(0.743)	(0.002)			0.108
3rd	3.113***	(0.743)	(0.002)		X	0.120
4th	2.882***	(0.840)	(0.024)			0.109
4th	2.882***	(0.840)	(0.024)		X	0.121
5th	4.290***	(1.129)	(0.004)			0.109
5th	4.290***	(1.129)	(0.004)		X	0.122
1st	1.365***	(0.301)	(0.000)	X		0.106
2nd	3.645***	(1.016)	(0.034)	X		0.109

Subsample of people never going college. Dependent variable is years of education. Level of significance (based on clustered standard errors): *** 0.01; ** 0.05; * 0.1. ‘Restricted Sample’ consists of half the index span around the cut-off (index values 15 to 44). ‘Flexible slope’ allows the slope of the regression specification to vary on either side of the cut-off. Standard Errors clustered at the Caste-level, and Cameron et al. (2008) Wild Bootstrap p-values also presented.

Table 2.16: Young Sample – No 24 to 30 year olds.

Polynomial	β_1	Clustered SE	CGM pval	Flex Slope	Restricted Sample	R-squared
3rd	3.105***	(0.746)	(0.000)			0.108
3rd	3.513**	(1.231)	(0.050)		X	0.120
4th	2.959***	(0.854)	(0.014)			0.109
4th	2.378	(1.490)	(0.126)		X	0.121
5th	3.448***	(1.222)	(0.034)			0.109
5th	3.951**	(1.572)	(0.038)		X	0.122
1st	0.473	(0.650)	(0.580)	X		0.106
1st	5.985*	(3.173)	(0.106)	X	X	0.119
2nd	19.11**	(8.051)	(0.122)	X		0.109
2nd	-13.83	(46.45)	(0.786)	X	X	0.121

Subsample of people no 24 to 30 year olds. Dependent variable is years of education. Level of significance (based on clustered standard errors): *** 0.01; ** 0.05; * 0.1. 'Restricted Sample' consists of half the index span around the cut-off (index values 15 to 44). 'Flexible slope' allows the slope of the regression specification to vary on either side of the cut-off. Standard Errors clustered at the Caste-level, and Cameron et al. (2008) Wild Bootstrap p-values also presented.

Table 2.17: Education for the Older Population – Not eligible for reservations

Polynomial	Flex Slope	Restricted	β_1	Std Err	R-sq
3rd			-0.165	(0.109)	0.189
3rd		X	-0.0615	(0.152)	0.175
4th			-0.112	(0.106)	0.189
4th		X	-0.0361	(0.129)	0.175
5th			-0.153	(0.109)	0.189
5th		X	-0.465*	(0.240)	0.175
1st	X		0.0545	(0.0641)	0.189
1st	X	X	-0.648	(0.427)	0.175
2nd	X		-1.876**	(0.869)	0.189
2nd	X	X	7.224	(5.678)	0.175

Dependent variable is years of education. Sample restricted to older population. Level of significance: *** 0.01; ** 0.05; * 0.1. 'Restricted Sample' consists of half the index span around the cutoff (index values 15 to 44). 'Flexible slope' allows the slope of the regression specification to vary on either side of the cut-off.

Table 2.18: Difference in Discontinuities – No college

Polynomial	β_1	Clustered SE	CGM pval	Flex Slope	Restricted Sample	R-squared
3rd	2.424***	(0.623)	(0.00)			0.104
3rd	2.424***	(0.623)	(0.00)		X	0.103
4th	2.194***	(0.572)	(0.01)			0.105
4th	2.194***	(0.572)	(0.01)		X	0.108
5th	3.826***	(0.892)	(0.00)			0.105
5th	3.826***	(0.892)	(0.00)		X	0.111
1st	2.354***	(0.623)	(0.02)	X		0.106
2nd	1.127***	(0.134)	(0.00)	X		0.105

Subsample of people never going college. Dependent variable is years of education. Level of significance (based on clustered standard errors): *** 0.01; ** 0.05; * 0.1. 'Restricted Sample' consists of half the index span around the cut-off (index values 15 to 44). 'Flexible slope' allows the slope of the regression specification to vary on either side of the cut-off. Standard Errors clustered at the Caste-level, and Cameron et al. (2008) Wild Bootstrap p-values also presented.

Table 2.19: Difference in Discontinuities – No 24 to 30 year olds

Polynomial	β_1	Clustered SE	CGM pval	Flex Slope	Restricted Sample	R-squared
3rd	2.128***	(0.650)	(0.022)			0.104
3rd	2.806**	(1.020)	(0.040)		X	0.103
4th	1.923***	(0.607)	(0.018)			0.105
4th	2.313*	(1.317)	(0.006)		X	0.108
5th	2.890***	(0.929)	(0.00)			0.105
5th	3.722**	(1.623)	(0.054)		X	0.111
1st	2.020***	(0.652)	(0.020)	X		0.106
1st	1.509**	(0.588)	(0.064)	X	X	0.103
2nd	1.031***	(0.152)	(0.00)	X		0.105
2nd	1.797***	(0.378)	(0.00)	X	X	0.113

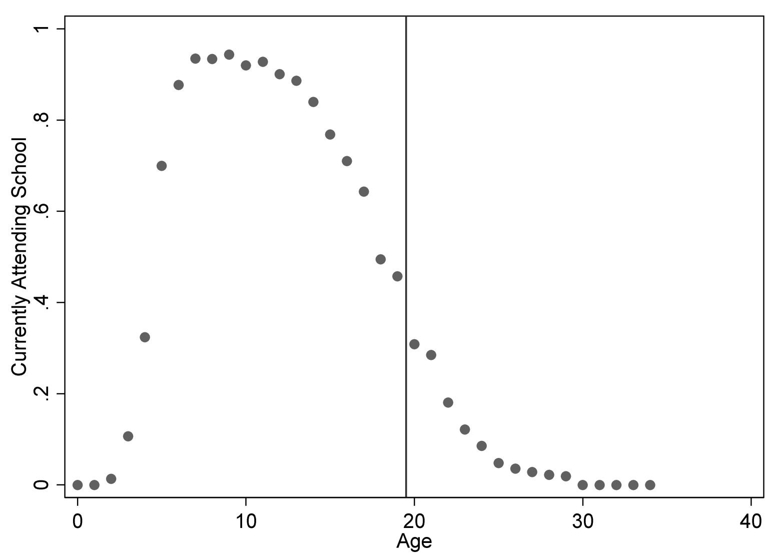
Subsample of people no 24 to 30 year olds. Dependent variable is years of education. Level of significance (based on clustered standard errors): *** 0.01; ** 0.05; * 0.1. 'Restricted Sample' consists of half the index span around the cut-off (index values 15 to 44). 'Flexible slope' allows the slope of the regression specification to vary on either side of the cut-off. Standard Errors clustered at the Caste-level, and Cameron et al. (2008) Wild Bootstrap p-values also presented.

Table 2.20: Comparing Haryana to the Rest of India

	Rest of India	Haryana	Difference
Sample Size	583422	10155	
Mean Education Level	4.629	4.763	-0.134
Monthly per cap Expenditure (Rs.)	464.612	526.433	-61.821
Male (%)	0.515	0.529	-0.014
Age	26.133	25.551	0.582
Urban (%)	0.376	0.363	0.013
Agricultural sector (%)	0.516	0.503	0.013
OBC (%)	0.33	0.278	0.052
SC (%)	0.158	0.165	-0.007
ST (%)	0.114	0.007	0.107

Using NSS 1999-2000 data. 'Mean Education Level' covers 8 levels of education from illiterate to college graduates. Nominal exchange rate: approx Rs. 50 to 1 dollar. Household Monthly Expenditure deflated by rural-urban-region-wise CPI.

Figure 2.8.5: Enrollment Rates by Age



National Sample Survey 1999. The largest drop in enrollment occurs at the age of 18 - representing a 16% point fall.

Table 2.21: RD: Robustness Checks: Other Discontinuities?

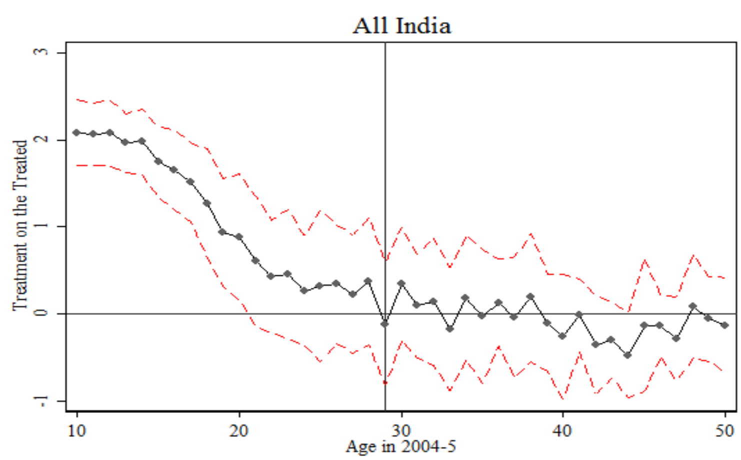
Polynomial	Restricted	Variable	β_1	Std Err	R-sq
3rd		Edu Expenditure	-0.0587	(0.0502)	0.016
3rd	X	Edu Expenditure	-0.0772	(0.0809)	0.013
4th		Edu Expenditure	-0.0834*	(0.0471)	0.017
4th	X	Edu Expenditure	0.0315	(0.0798)	0.014
5th		Edu Expenditure	0.0230	(0.0588)	0.019
5th	X	Edu Expenditure	0.0491	(0.116)	0.014
3rd		Med Expenditure	0.0273*	(0.0156)	0.008
3rd	X	Med Expenditure	0.00234	(0.0166)	0.005
4th		Med Expenditure	0.0223	(0.0142)	0.008
4th	X	Med Expenditure	-0.0309*	(0.0172)	0.006
5th		Med Expenditure	0.00726	(0.0187)	0.009
5th	X	Med Expenditure	-0.0181	(0.0152)	0.006
3rd		Total Expenditure	30.73	(30.97)	0.072
3rd	X	Total Expenditure	-29.21	(25.01)	0.094
4th		Total Expenditure	9.403	(20.71)	0.094
4th	X	Total Expenditure	-59.37*	(32.23)	0.097
5th		Total Expenditure	10.06	(35.61)	0.094
5th	X	Total Expenditure	-95.38***	(29.56)	0.100
3rd		Casual Labor	-0.686*	(0.390)	0.151
3rd	X	Casual Labor	0.289	(0.226)	0.189
4th		Casual Labor	-0.381	(0.259)	0.194
4th	X	Casual Labor	0.253	(0.303)	0.189
5th		Casual Labor	-0.360	(0.372)	0.194
5th	X	Casual Labor	0.489	(0.287)	0.190

Variable ‘Edu Expenditure’ is expenditure on education-related goods. ‘Med Expenditure’ is medical expenditure. ‘Total Expenditure’ is expenditure on all items. ‘Casual Labor’ is 1/0 indicator of occupation. Level of significance: *** 0.01; ** 0.05; * 0.1. ‘Restricted Sample’ consists of half the index span around the cutoff (index values 15 to 44).

Figure 2.8.6: OBC Coefficients of Diff-in-Diff Regression across age cohorts (Controlling for SC-STs)

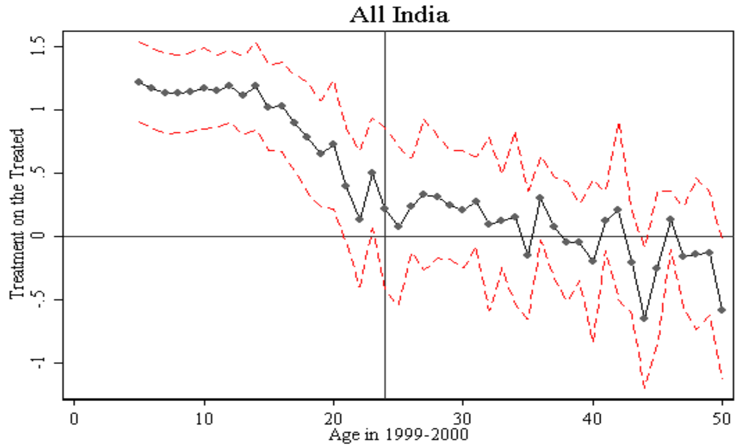
Unlike the figure in the main text, the omitted group in this picture is only upper-caste individuals. Whereas SC-ST indicators are used as controls. Plot of coefficients from Difference-in-Differences regression on 'levels of education'. Standard errors clustered at state-level. People above the age of 24 should be unaffected by implementation of policy. Vertical lines indicate year of implementation of policy.

Figure 2.8.7: Five years later - OBC Coefficients of Diff-in-Diff Regression across age cohorts (2005)



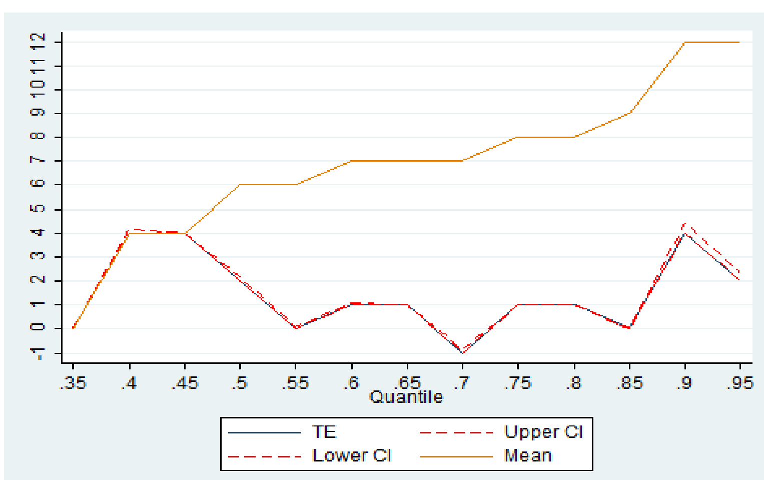
Source: 2004-5 NSS data. Standard errors clustered at state-level. Vertical lines indicate year of implementation.

Figure 2.8.8: OBC Coefficients of Diff-in-Diff Regression across age cohorts (Years of Education)



Dependent variable is years of education. Standard errors clustered at state-level. Vertical lines indicate year of implementation.

Figure 2.8.9: Quantile Treatment Effects

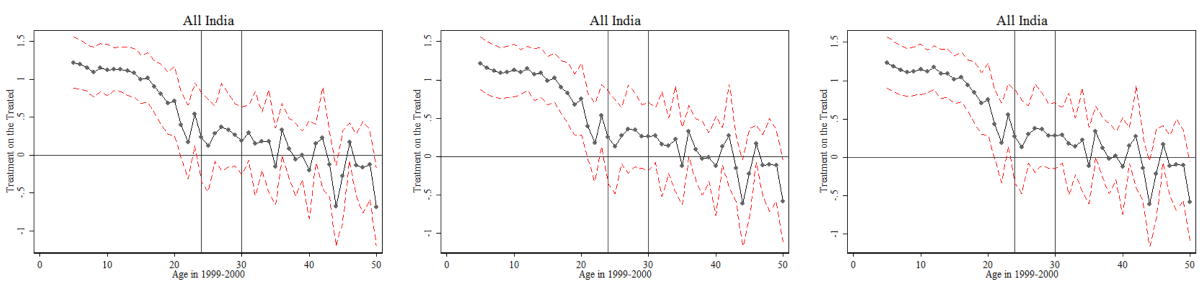


Quantile Treatment Effects using the Difference-in-Differences method, across caste and age cohort.

Figure 2.8.10: Pre-Collegiate Sub-sample

Standard errors clustered at state-level. Vertical line indicates year of implementation. Sub-sample of those without a college education.

Figure 2.8.11: Operation Blackboard and District Primary Education Program (DPEP)



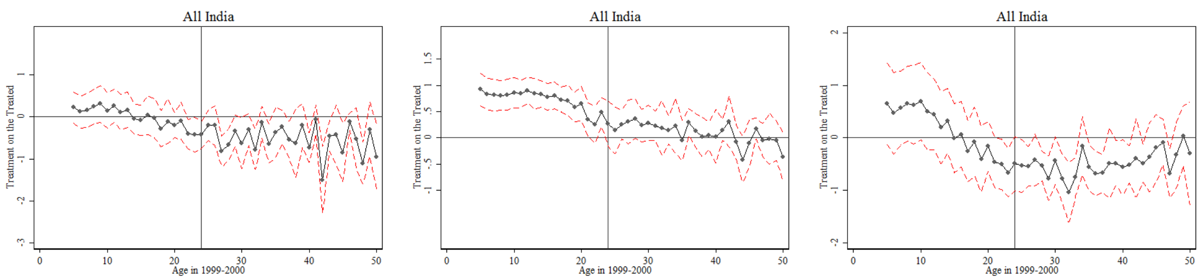
(a) Controlling for Operation Blackboard

(b) No controls

(c) Controlling for DPEP

Years of Education. Standard errors clustered at state-level. Persons above 24 unaffected by Reservations, persons above 30 unaffected by National Policy of Education. ‘Others’ defined as non-(OBC, SC or ST). Controls include a quadratic in the ‘intensity of the policy,’ an indicator for being above median intensity, and an indicator for whether the state received any funds under the policy. Intensity of Operation Blackboard is taken from Chin (2005). Intensity of DPEP is defined as the proportion of districts within a state that received funds under DPEP.

Figure 2.8.12: Comparing Impacts on OBCs with Muslims and the Poorest 20% of Others



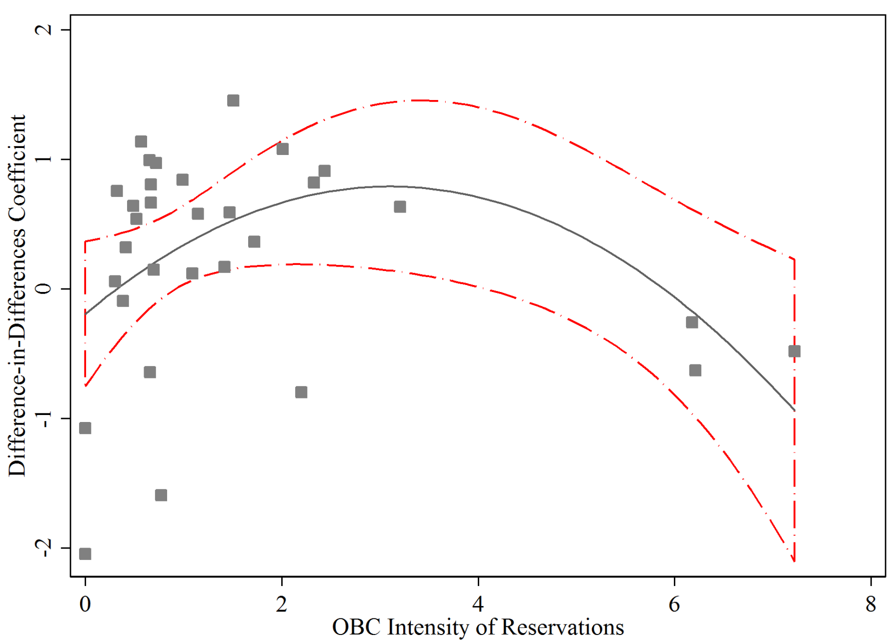
(a) Muslims (b) OBCs (c) Others - poorest 20%

Axes scaled for consistency across graphs. Dependent Variable: Level of Education. Standard errors clustered at state-level. Persons above 24 unaffected by Reservations; Persons above 30 unaffected by National Policy of Education. ‘Others’ defined as non-(OBC, SC or ST). In the regression with the Muslims, the omitted category is Hindus. In the regression with the poorest 20% of upper-caste members, the omitted category is the richest 80% of the upper-caste population.

Figure 2.8.13: Impacts on OBC: Control Group – only Muslims or Poorest Upper Caste

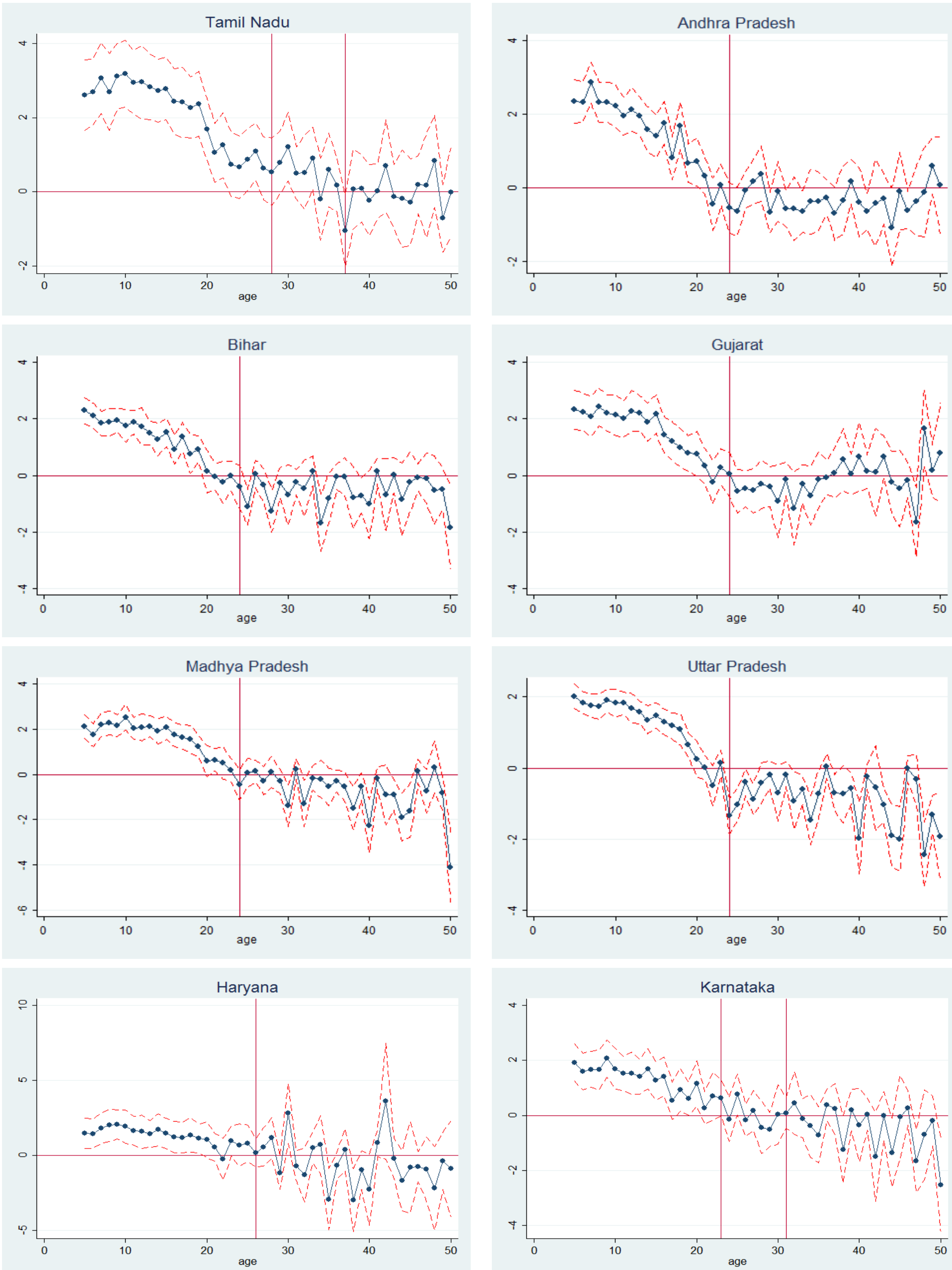
Standard errors clustered at state-level. Vertical line indicates year of implementation. Sub-sample of OBCs, Muslims, and bottom two income deciles of non-OBC/SC/ST population.

Figure 2.8.14: State-wise Auxiliary Regression With Outliers



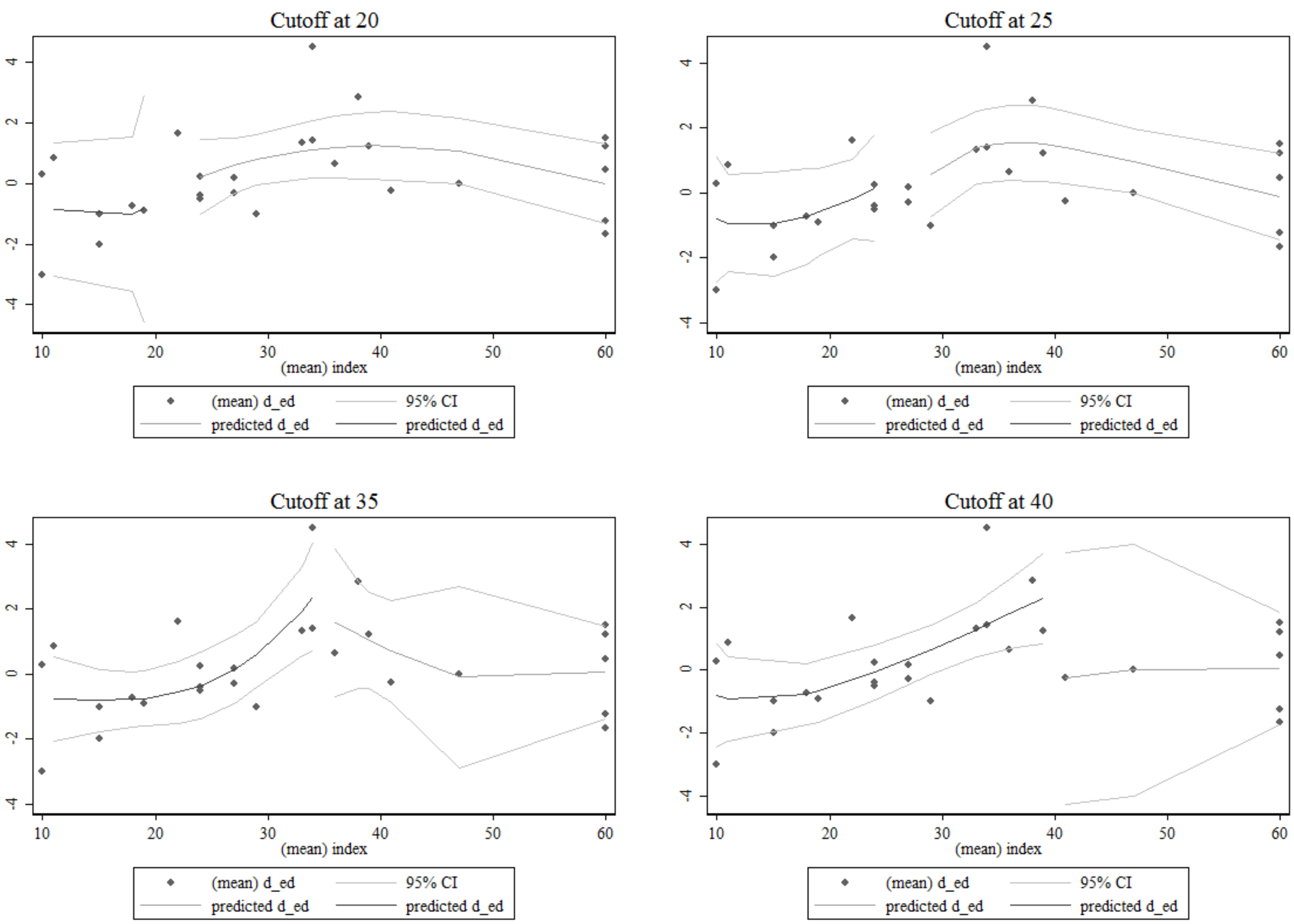
Auxiliary Regression of state-by-state relationship between the ATET and intensity of OBC reservations in each state.

Figure 2.8.15: State-Wise Changes in Reservation Policy for OBCs



Vertical lines indicate year of implementation of significant changes in state-wise policies. Primary source data on reservation policy changes collected via Right to Information (RTI) Act petitions.

Figure 2.8.16: Looking for other Discontinuities at index values: 20, 25, 35 and 40



Placebo test: looking for discontinuities at other values of the index. True cutoff value is 30.

CHAPTER 3

Road Oft Taken: The Route to Spatial Development

Most estimates of the economics impacts of transit networks compare regions along such routes to neighboring regions away from the route. In the presence of spatial spillovers in economic activity, such a method will underestimate the true effect of roads and railways. I model and estimate the overall impacts of transit networks in India by first measuring the extent of such spillovers. I use an empirical strategy relying on the historical placement of major cities: by connecting nodal cities with straight-lines I instrument for the endogenous placement of these networks. Using night-time luminosity data as a measure for economic activity, I estimate the parameters of the model, which does a good job of fitting the data. I find that being close to such transit networks led to greater economic activity in the 1990s, and such activity spread to neighboring regions, substantially increasing the overall impacts of such routes. Ignoring the spillovers produces income elasticities that are only 27% of the true overall effects of such routes, and such geographic externalities led to a rapid rate of convergence in incomes across regions. Transit networks and spatial spillovers together, therefore, strongly determine the geographic spread and temporal changes in the economic development of the region.

One reason why some regions within a country are rich while others are poor is differences in access to infrastructure. This is especially true in developing countries, where large regional disparities in income levels can arise as a result of these infrastructure differences. This paper examines how access to transit networks can affect the level of income and growth across regions in India. Importantly, unlike cross-country convergence in incomes, intra-regional growth is often promoted by spillovers across regions. I study the impact of road and rail routes connecting the four major Indian cities, the pattern of regional development produced due to access to these networks, and the corresponding spatial spillovers in economic activity.

Past work on transit networks that compares regions along the route to neighboring regions overlooks the fact that these neighboring regions are also affected by these routes because of spillovers across regions. If the spillovers are positive, then such an empirical strategy would understate the true impacts of these routes. If the spillovers are large, then it may seem like there are no impacts of road and rail routes, when in fact the overall effects are in fact, even larger. Therefore, to understand the full extent of the impacts of a route, the amount of spillovers need to be quantified, and the entire pattern of spatial development needs to be studied in a dynamic setting.

Within the Indian context, I focus on the two decades starting in 1992. This was a period of rapid growth, economic liberalization, and a period of upgrading the existing highways that connected major cities. Consistent with decreasing income-differentials for regions along the route, I find that while significant spatial inequalities existed at the beginning of the period, there was a rapid convergence across regions in the 1990s, as economic activity spreads from these routes to neighboring regions allowing them to catch-up. I find that ignoring these externalities would lead to estimates of income elasticities that are only 27% of the true overall impact of these routes. In 2012, a 1% decrease in distance from these routes raises incomes by only 0.06% ignoring spillovers, but overall incomes rise by 0.23% once spillovers have been incorporated. These spillovers also drive the rapid rate of convergence in incomes across regions – I estimate a 4% rate of β-convergence which is twice as high as the rate estimated in the cross-country literature.

Better infrastructure is generally thought to aid development by reducing the costs of trade and migration, equalizing prices and facilitating the spread of ideas and technology. However, causal impacts of transit networks are hard to find for two reasons. First, the true placement of these routes is endogenous. More developed regions have funds to build better roads and upgrade their infrastructure. And routes may be built to connect regions that were already developed, or expected to develop in the near future. It is easy to see that richer regions have better roads, but this is not only because roads may lead to more development, but also because these regions have the capability to build better roads, and many of the roads were built in regions that were starting out on a path to prosperity. On the other hand, these routes may have been built to help struggling regions recover, and are more likely to be built in regions with suitable terrain and easier land acquisition policies. In order to obtain causal estimates, I use straight-line paths between nodal cities as instruments for the existence of transit networks. In the first half of the paper (Section 3.3 and 3.4), I discuss the empirical reduced form impacts of these routes on spatial development. I combine this with a multi-period differences-in-differences specification to find that being connected to transportation networks causes a region to be more developed than neighboring regions, but neighboring regions catch up over time.

Secondly, there is the issue of finding an effective 'control' group. To reiterate, a neighboring region will provide an underestimate of the true impacts due to the spillovers. So, in the latter half of the paper (Section 3.5), I develop a model that captures these reduced form patterns, and then structurally characterize income growth and regional convergence in the presence of economic externalities. The model produces strong predictions for various parameters – namely, that the reduced form elasticity of income with respect to distance is a specific function of the extent of spillovers in economic activity and the direct impact of distance from these routes. I test these predictions and show that the model does a good job of fitting the data. With the help of the model and the empirical estimates, I determine the parameters of convergence and the extent of spatial spillovers across regions. I am able to use these causally estimated parameters to quantify the overall impact of these transit networks. Ignoring these spillovers would severely underestimate the true overall impacts of these routes.

Estimating such spillovers is essential for policy analysis, because positive externalities across regions suggest that the benefits of infrastructure projects are larger than previously thought. I also find that while the early construction of transit networks lead to regional development, there were diminishing returns to continue investing in and upgrading these highways. It would then be a better policy to connect other regions with road and rail routes, rather than upgrade already existing networks.

The road-map of this paper is as follows: Section 3.1 discusses the literature and the background motivating the identification strategy. One notable absence from the literature is a credible way to estimate and quantify the extent of spillovers across regions that magnify the overall impacts of these routes. In Section 3.2, I discuss the data I use, where I merge the night-time lights data with data on road and rail routes, and other sub-district level indicators. Section 3.3 studies the reduced form effect of being along these straight-line paths that connect major cities, and Section 3.4 describes what happens to this relationship over time, including the effects of upgrading the highways. After which, I discuss a model in Section 3.5 that explains the patterns in the data described in Sections 3.3 and 3.4. The first part of the model focuses on standard Neoclassical growth theory (Section 3.5.1) and the corresponding convergence in incomes across regions (Section 3.5.2). In Section 3.5.3, I introduce the role played by spillovers, and estimate the parameters of the model and rigorously test the predictions in Section 3.5.4. I use the estimated parameters of the model, including the extent of the spillovers, to determine the overall impact of theses routes in Section 3.5.5, and discuss the importance of these results in the context of the long literature on infrastructure projects in Section 3.6.

3.1 Context and Literature

A number of papers follow the methodology first established by Chandra and Thompson (2000) and later built upon by Michaels (2008) to estimate the impacts of US inter-state highways.[1] These papers focus on non-metropolitan areas that lie on highways, as these areas are more likely to be on routes only because they happen to be lying in regions between the metros. This strategy, therefore, may help tackle many endogeniety issues. In other contexts, like India, such a strategy may not be enough. There are many possible routes between two metros, and the actual placement of the highway is therefore endogenous – trying to connect growing rural areas, or areas that are particularly poor, or areas with suitable terrain, land cover and easier land acquisition.

Two papers in the Chinese context help tackle these endogeniety issues. In a significant contribution, Banerjee et al. (2012) use 'straight-lines' to connect historical cities in China, and use this to predict the existence of transportation networks. They find that this can explain moderate differences in GDP per-capita, but has no effect on income growth. A recent paper by Faber (2014) uses the construction of the Chinese National Highway system and combines it with a spatial instrument based on the 'least-cost' path of connectivity, that depends on terrain, water bodies and land cover.[2] Similarly, I connect historical cities in India, which were selected as nodal cities for a large highway-upgrading project, and study the impact on indicators that are closely related to economic activity, like the amount of night-time lights emitted (luminosity) captured by satellites in this period.[3] However, rather than just comparing regions along these paths to neighboring regions (a 'control' group), I estimate the spillovers to the neighboring regions so as to pin down the overall causal impact of the transit network and the dynamic nature of convergence in incomes across regions.

Unlike other papers I, therefore, focus on spatial spillovers and how they cause poorer regions to catch up. One crucial difference between the Chinese and Indian cases is the mobility of factors – while migration was highly regulated for many decades in the Chinese context,[4] there is unrestricted labor mobility in India, which would be especially useful in an analysis of spillovers across neighboring regions. In so far as access to transportation will have large impacts via migration, the Indian context would allow for studying these effects and be more relevant to contexts that do not have migration restrictions. While labor mobility in India is low Munshi and Rosenzweig (2016), this may be due to the high costs of migra-

[1] See Redding and Turner (2015) for a discussion of models and identification methods in this literature.

[2] Two other recent papers in the Chinese context study the decentralization of Chinese cities Baum-Snow et al. (2014) and market access Baum-Snow et al. (2015).

[3] I refrain from using the 'least cost' path since in the Indian context, since an instrument that relies on land cover, water bodies and terrain may introduce other sources of endogeniety. A region that is flat may grow faster than a hilly region for reasons unrelated to transit networks. A region that has less land cover may have been cleared for development purposes. Furthermore, regions with less land cover and flatter terrain are also more likely to be near cities. In my context, therefore, a simple 'straight-line' is arguably a cleaner instrument.

[4] From 1958 to 1978 it was restricted, and then some reforms were put into place to loosen but regulate mobility till the late 1990s.

tion for regions that are not well connected to transportation networks.[5] The second big difference is the data used – similar to the Chinese case, the Indian data on GDP at a sub-regional level is poor and problematic. Changes in data-collection methodology over time and across regions may well be correlated with regional development and access. Banerjee et al. (2012) also highlight other issues with the Chinese context – that a non-random sample of regions report GDP numbers, and which years those regions chose to report is also endogenous. It may therefore be better to use data collected from an external source – like the night-time lights data used by this paper. Lastly, unlike the Banerjee et al. (2012) paper, I conduct a before-after analysis to look at the impact of a large upgradation under the National Highway Development Project (NHDP) to test whether additional investments into already existing networks matter, and find little to no benefits of these upgrades.

The highway system studied here connects the four nodal cities forming what is called the Golden Quadrilateral (GQ). Three of the four cities (Mumbai, Kolkata and Chennai) were chosen to be capitals of the British Presidencies because they were natural harbors and therefore could be used as ports for trade. There was little economic activity in these three regions prior to the British, and were therefore not on any pre-existing road network. The fourth (Delhi) was a major historical capital of various pre-Colonial empires, and was a British cantonment during the Raj.[6]

I focus on the period between 1992 and 2012. While the decades leading up to this period was burdened with sluggish growth, these two decades were a time of high and rapid development following the reforms of 1991 that came under a proclamation of 'Liberalization, Privatization, Globalization.'[7] Starting in 1999, the Golden Quadrilateral (GQ) project upgraded about 5,846km of already existing highways in India. The NHDP invested about US $71 billion in order to widen the national highways, and strengthen them for heavy traffic and truck transportation. While the proposal was approved in 1998, many projects started only as late as 2001. Most of the delays had to do with issues of land acquisition, which makes the placement of the final roads endogenous, prompting the use of the 'straight-lines' between the nodal cities.[8]

Ghani et al. (2015) focus on the upgrades in the late 1990s and look at the behavior of manufacturing firms. They find an increase in entry-rates for organized manufacturing firms within 10km of the high-

[5] Even though labor mobility is low, capital is relatively more mobile in India Ghani et al. (2015).

[6] Three cities were then chosen to be major British capitals for their natural harbors and strategic positioning of coastal forts – all reasons unrelated to being near land-based routes. A member of the British East India Company arrived near modern day Kolkata in 1690, and the British established Fort William in 1698, which gave rise to the modern Kolkata. A few decades before that, in 1639, the British had set up Fort St. George which grew into modern day Chennai. While on the other side of the peninsula, Francisco de Almeida, a Portuguese explorer, sailed into the deep natural harbor of the Mumbai islands in 1508, and the Portuguese acquired the islands in 1534. In 1661 the islands were given to the British as part of the dowry for Catherine of Braganza's wedding to Charles II. Delhi, on the other hand, only passed over to British hands in 1803.

[7] The reforms opened up major sectors of the economy to foreign trade and eventually some sectors to foreign investment, privatized many industries and cut down what is well known as the 'license-permit raj.'

[8] The junior Highways Minister told the Parliament that "Projects have been delayed mainly due to problems associated with land acquisition, shifting of utilities, obtaining environment and forest clearance, approval for road over bridges, poor performance of some contractors due to cash flow constraints and law and order problems in some states." The bulk of the projects were over by the end of 2006, but some alterations on additional phases of the project continue even as late as 2014.

ways, and modest impacts on other indicators.[9] They are careful to point out that their OLS results may be affected by the fact that the route of the highway may have been chosen to connect regions that were (a) expected to develop or attract businesses, or (b) were struggling and needed investments to turn them around, or (c) had other systematic differences like lesser land acquisition issues, and therefore less agricultural regions. My paper is, therefore, complementary to the Ghani et al. (2015) work. Firstly, while their paper analyzes how manufacturing firms respond to upgrades, I estimate the effects of historical connectivity on $overall$ economic activity associated with night-time lights. Secondly, since the highway project updated an already existing network of different forms of transportation, unlike the Ghani et al. (2015) paper, my paper focuses on the long-run economic impacts of being connected to these historical transportation networks, and the eventual dispersion and dissipation of these impacts. When I also look at the upgrades, our results together suggest that organized manufacturing responded differently to any new upgrades than other agents in the economy. Last, I study and quantify the extent of neighborhood spillovers and show how they affect the rates of convergence in incomes across regions by formalizing regional development in a growth-model framework.

Within the Indian context, there has also been work using other identification strategies, like using the historical expansion of railroads (Donaldson, 2014) to look at price equalization and regional development, or more recent rural road construction programs (Asher and Novosad, 2016) to look at village employment. Papers on highway infrastructure in the US context (Atack et al., 2008; Donaldson and Hornbeck, 2015; Michaels, 2008) look at market access, urbanization, population movements and the demand for skill across different regions. Lastly, while the literature on neighborhood spillovers in economic activity is scant, there is a growing literature on Solow-style convergence within countries (for an analysis of US counties see Higgins et al. (2006)) which my paper addresses by estimating the rate of β-convergence across regions driven by these regional externalities.

3.2 Data and Sample

The primary dependent variable of interest is night-time lights as measured by satellite imagery. This has been used as an indicator for economic development, especially in developing countries that have issues with disaggregated income data (Henderson et al., 2012; Michalopoulos and Papaioannou, 2013). Researchers at the National Oceanic and Atmospheric Administrations (NOAA) National Geophysical Data Center (NGDC) process data from weather satellites that circle the Earth 14 times a day and take pictures between 2030 and 2200 hours at night. They use algorithms to filter out other sources of natural light using information about the lunar cycles, sunset times and the northern lights, and other occurrences like forest fires and cloud cover. Given a lack of reliable sub-regional level GDP data in the India, this

[9]For other contemporaneous work related to the upgrades on market access within the context of a Ricardian trade model, see Adler (2016); Asturias et al. (2015).

measure is ideal to capture overall economic activity.

Figure 3.7.1a zooms in on the region connecting Delhi-Mumbai and Delhi-Kolkata where a stream of lights is associated with the National Highway that connects the nodal cities, whereas Appendix Figure 3.8.1 shows the entire geographic distribution of night-time lights along with the straight-lines between the four nodal cities. The Golden Quadrilateral Highways and the sub-districts used in the analysis are shown in Figure 3.7.2. Distance to the nearest straight line is calculated using standard ArcGIS software, and in some regressions I also use data on actual highways and roads that are obtained from the Digital Chart of the World (DCW) database. DCW provides detailed information on road and rail routes based on the content in the US Defense Mapping Agency (DMA).

The lights data is calculated at approximately every one square kilometer, but I aggregate the results to the sub-district level in order to account for issues of spatial correlation.[10] I regress this lights data on distance to the nearest straight line connecting the nodal cities.[11] Epanechnikov kernel density plots of light density over time are presented in Figure 3.7.1b. As we can see, as the country grows between 1992 and 2002 there is an increase in mean light density and reduction in the variance, but this is not true of the following decade.

In all calculations, I drop the nodal cities and 26 adjacent sub-districts so as to not capture the impact of being a neighbor to a big city – doing so only slightly attenuates the results. It is also important to include year fixed effects to capture the change in average light density due to changes in satellites across years.[12] In all specifications, I also flexibly control for other geographic features like distance to closest nodal city, coastline, latitude and longitude.

3.2.1 The Elasticity between Light Density and Domestic Product

In order to get at spatial development at a finer level, this paper studies the impacts at the administrative level of the 2253 sub-districts. Gross Domestic Product (GDP) is not calculated at this level, but there are GDP numbers available for about half the 594 districts in the country, and for all the 32 states. The elasticity between State Domestic Product (SDP) and light-density will be an underestimate of the true

[10]Sub-districts are third largest administrative unit of aggregation with a population of about 460,000 people on average. Results will be presented with standard errors clustered at even higher levels of aggregation. There are about 2253 sub districts in 594 districts which are in 35 states and union territories. Results are statistically significant even at standard errors clustered at the level of 35 states.

[11]Following the conventions established in the literature (Michalopoulos and Papaioannou, 2013), the lights data is transformed to be of the form: $Log(0.01 + Luminosity)$ to account for the fact that some areas have no luminosity. About 1.6% of the total sample, and less than 1% of the last 3 years of the sample have sub-districts that had no luminosity. Furthermore, the results are robust and significant in using Poisson regression specifications of luminosity. I also present results of $Log(Luminosity)$ for the regions that never have 0 recorded lights - which can be interpreted as an impact on purely the intensive margin.

[12]These include Delhi and surrounding areas like Ghaziabad and Gurgaon, Mumbai and greater-Mumbai, Kolkata and Haora, and Chennai and its neighbors.

elasticity between sub-district domestic product and light-density because of the measurement error introduced in aggregating the 2253 sub-districts into 32 states. The relationship between District Domestic Product (DDP) and light-density, while suffering from some measurement error as well, will however bring us closer to the true parameter. Appendix Tables 3.14 and 3.15 show the state-level and district-level elasticities between GDP and luminosity. While the state-level elasticities are a little below 0.2, the district-level elasticities are a little above 0.3 in general, and there is no trend over time in the elasticities. Due to measurement error in aggregating lights and domestic product to a higher administrative unit, the true sub-district level elasticities should be higher. The cross-country literature (Henderson et al., 2012; Michalopoulos and Papaioannou, 2013) has elasticities of about 0.3 for sub-samples of low and middle income countries, suggesting that 0.3 would be a reasonable lower bound for the sub-district level elasticities.

3.3 Historical Connectivity to Transit Networks

In this section, I study the effect of being close to transit networks that have historically existed for many decades, and In the following section, I see how this long-run relationship is changing over time. These two sections characterize the reduced form impacts of roads. In Section 3.5, I parameterize these relationships in order to estimate the overall impacts of these routes in the presence of regional externalities.

3.3.1 Empirical Strategy

While the actual path of these routes is endogenous to regional characteristics, being on a straight-line between two major cities should not be correlated with anything other than being close to the routes connecting them. In order to examine the long-term general impact of connectivity to historically determined transportation networks, one can look at the impact on lights for cities closer to the straight-lines using the following regression specification, for sub-district i:

$$LogLights_i = \alpha_i + \gamma \mathbf{X} + \beta Distance_i + \epsilon_i \tag{3.1}$$

Equation (3.1) is the reduced form specification where $Distance_i$ is the distance between the sub-district and the nearest straight line, and $\mathbf{X}$ are geographic controls (distance to nearest nodal city, coastline, latitude and longitude).[13] Similarly the OLS formulation of this same equation would replace the $Distance_i$ variable with distance to the nearest highway, or distance to the nearest rail-line. The OLS estimates,

[13] Some results are presented as distance in kms to be comparable to the Ghani et al. (2015) paper, and other results are presented as Log(distance) to calculate the elasticity as in the Banerjee et al. (2012) paper.

however, will be biased because the highways and rail-lines will be laid according to where cities and economic centers are located, or lagging regions where the government wishes to induce economic activity. It is interesting, however, to study the direction of the OLS bias. If road and rail-lines are laid to be closer to economic centers, then the OLS estimates will show large impacts of being close to a highway or rail-line. If, however, land acquisition for construction of rail-lines and roads forces the government to move away from economic centers, then the impact of distance to these lines would be attenuated.

Finally, one can derive the upper-bound of the effect of roads and rail lines by performing a two-staged least squares exercise of the following form:

$$\begin{aligned} DistanceToRoad_i &= \pi Distance_i + \mu_i \\ LogLights_i &= \alpha_i + \delta_i + \gamma \mathbf{X} + \beta \widehat{DistanceToRoad}_i + \epsilon_i \end{aligned} \tag{3.2}$$

For this to be a valid instrument, it must be that regions along these straight-lines do not systematically have a different light-density for any reason unrelated to road or rail routes. Since the impact of the distance to the straight-lines will work through both rail lines and roads, the coefficient β is not an instrument for each separately, but rather an instrument for transit networks in general. It is, however, possible to see the strength of the instrument for roads and railways separately. I find that the straight-lines are strong predictors of highway placement, but not for rail lines, suggesting that the effects I capture are more likely to be driven by these highways. Later, the paper will seek to isolate the impact of upgrading the highway system by comparing districts close to and far away from the straight-lines, before and after the highways were upgraded.

Table 3.1 shows the relationship between distance to the straight-lines and distance to transit networks. While the distance to the straight-line is a good predictor of distance to the nearest GQ highway, it only does moderately well in predicting distance to closest rail line. This is hardly surprising, since while the GQ highways were built in order to connect the nodal cities, the rail lines were built to connect other cities as well.[14] Throughout the paper, the results will be clustered at higher-level administrative units like districts or states in order to account for possibilities in spatial correlation and for other sources of correlations in outcomes within administrative units.

3.3.2 Results: Distance to GQ Highway and Railways

The OLS, reduced form and IV-2SLS relationships between light-density and proximity to the nearest GQ highway are shown in Table 3.2 for every decade – the years 1992, 2002 and 2012.

In 1992, the OLS relationship between light-density and distance had a coefficient of -0.391 (Table 3.2).

[14] Interestingly enough, the first set of railway lines laid in India (in the 1860s) were built by Lord Dalhousie to connect the five major provincial capitals – four of which are the current nodal cities.

Since the distance variables are in 100km, this means that a 100 km increase in distance from the highway was associated with a fall in light-density of 0.391 log points. As discussed previously, a reasonable lower-bound for the elasticity between sub-district domestic product and lights is 0.3; in the literature, a 1 log point increase in light density is usually related to the a 0.3 log point increase in income for the sample of low and middle income countries (Henderson et al., 2012; Michalopoulos and Papaioannou, 2013). Therefore, a 100 km increase in distance from the highway would be correlated with a 12% fall in income. By the year 2012, this had halved to about a 6% difference in income.

The most commonly used metric for luminosity as a predictor of development is the light-density variable (Henderson et al., 2012; Michalopoulos and Papaioannou, 2013). In Table 3.11, I present the OLS relationship between distance and $other$ measures of night-time lights, including the standard deviation of lights within a sub-district. In order to look at the extensive margin, the last column in Table 3.11 is a linear probability model (LPM) of the probability of having the majority of recorded light-emission pixels in a sub-district be greater than 0. This relationship is greater in 1992, again suggesting that over the two decades the relationship between distance to the highway and development has weakened.

The OLS relationship, however, could be biased as the exact path of the highway will depend on the government's wish to connect some areas, and the ease with which land could be acquired for construction. The second column of Table 3.2 shows the reduced-form relationship between lights and the distance to the straight-lines that connect the nodal cities. Once again, the relationship is much larger in magnitudes in 1992 than in 2012, and of similar size as the OLS relationship. Similarly, Table 3.12 shows the analogous reduced form relationship for other measures of night-time lights and distance to the straight-line.

The two-staged least squares (2SLS) estimates are presented in the final column of Table 3.2 (and Table 3.13 shows the corresponding relationship for other measures of lights). In Table 3.1 we can see that the excluded distance-to-line variable has an extremely high F-stat no matter what the level of clustering, displaying a strong first-stage relationship. The 2SLS estimates are slightly more negative than the OLS estimates in some cases – for example, in 1992, the 2SLS results say that a 100km increase in distance from the highway leads to a 0.5 log point fall in light density. Assuming the same elasticity between lights and income, this is a 0.15 log point or about a 16% fall in income. The 2012 light-density coefficient, however, is identical to the OLS result.[15]

In Table 3.3, I estimate the elasticity between distance and economic activity over time. By using a log-log specification in each year, I calculate how this elasticity between lights and distance to the nearest straight line is changing between 1992 and 2012. Given that a lower bound for the elasticity between

[15]Tables 3.16 and 3.17 show the analogous OLS and 2SLS results for distance to the nearest railway line. Keeping in mind that distance to the straight line is only a moderately good predictor of distance to the rail line, the bulk of the reduced-form impact is coming via roads. For the railway lines, we can see that the 2SLS results are much larger than the OLS results. The reduced form is in Table 3.12, and the 2SLS results are magnified by the fact that the excluded distance-to-line variable is not a very good predictor of the distance to nearest railway line.

lights and GDP per capita is constant at 0.3, the results indicate that the elasticity between GDP per capita and distance is falling over this period from about 0.15 to 0.06. The range subsumes the Banerjee et al. (2012) elasticity of 0.07, but for much of the period is higher. The results therefore indicate that in 1992 this elasticity was high, and that historical connectivity played a large role in regional and spatial development. However, this elasticity more than halves to fall to an economically insignificant relationship by the end of the period. If we merely estimated the impact in 2012 we would say that transit networks do not significantly affect incomes.

This dissipation of the impacts hasn't been investigated much in the literature, and is studied below in Section 3.4 in more detail, where large spatial spillovers can lead to a dissipation in the differential increases in income for the regions along these routes.

3.4 Dynamics and Changes Over Time

In this section, I examine the dynamics of this dissipation in more detail. I focus on how the economic impacts of these routes change over time, and the effect of the major highway upgrading program which took place between 1998 and 2006. First, I estimate the following regression for sub-district i in year t:

$$LogLights_{it} = \alpha + \delta_i + \tau_t + \gamma \mathbf{X} + \beta_{\mathbf{t}} Distance_i * \tau_t + \epsilon_{it} \tag{3.3}$$

The regression includes year fixed-effects τ_t, region fixed effects δ_i and the usual geographic controls $\mathbf{X}$, similar to a multi-period difference-in-differences specification. The coefficient of interest $\beta_{\mathbf{t}}$ is on the interaction term between $Distance_i$ and τ_t. In the regressions, the omitted year is the first year of the sample – 1992. One can plot this coefficient $\beta_{\mathbf{t}}$ to look at how the impact of distance to the line is changing over this period, relative to the year 1992.

Figures 3.7.4b shows how the effect of distance on light-density changes relative to 1992. Positive values of $\beta_{\mathbf{t}}$ indicate that the impact of distance on lights is falling relative to 1992. In the figures it is clear that this differential impact is indeed falling over time, and especially after the upgrades to the highways system begins in 1998. It is however, not possible to reject the possibility that the dissipation of the relationship between distance and development would not have happened if the highways were not upgraded. The panel on the standard-deviation of lights shows that even though inequality within a sub-district was higher for regions closer to these routes, this difference has been shrinking over time.

Appendix Figure 3.8.4 shows this relationship for different measures of night-time lights and finds similar trends over time – the impact of being far away from these routes decreases over time. Robustness to different specifications is shown in Appendix Figures 3.8.6. The figure has four panels allowing for two-different levels of clustering errors and doing robustness checks with dropping any sub-district that ever had 0 recorded luminosity, and dropping any subdistrict that ever had any pixel with the maximum possible luminosity value. Appendix Figures 3.8.5 reproduces the main results after excluding regions that are a significant distance away from these routes.[16]

While Table 3.3 shows the 'reduced form' elasticities between light and distance to the straight-line, in Figure 3.8.2 I show the OLS and 2SLS elasticities over time. While the elasticities are similar towards the end of the period, there are stark differences in the beginning of the period. One explanation for this convergence, that is worth exploring, is that in the early 1990s, the 'distance to line' could be picking up other transit networks as well in the 2SLS regression, but by the 2000s the highways seem to become the dominant channel. And this shift in importance of which transportation networks are used may be due to the highway upgrades.

3.4.1 Upgrading the Highways

The literature on the golden quadrilateral, has so far concentrated on the upgrades that were started in the late 1990s. The NHDP upgrading projects were first finalized in 1998, and the foundation stone was laid by the Prime Minister on January 6, 1999. The first couple of years, however, were plagued with delays in certain areas because of contractual issues and problems with land acquisition. About 20% of the projects started between 1998 and 2000, whereas almost 50% of projects started in 2001.[17] While Phase I of this project officially ended in 2006, about 8% of the projects ended a few years later. Later phases added some additional upgrades, and work continued on the GQ till the end of 2011. This timing allows for a before-after analysis of this highway construction, since the lights data spans from 1992 to 2012. The period 1999 to 2006 in the sample will be considered to be the 'construction' phase, while the years after that will be the post-project phase.

In order to see how the impact of distance changes with time, one can run the following regression:

$$LogLights_{it} = \alpha + \tau_t + \beta \mathbf{X} + \delta_{1i} Distance_i + \delta_{2i} Distance_i * Construction_t + \delta_{3i} Distance_i * Post_t + \epsilon_{it} \tag{3.4}$$

where δ_{1i} is the impact of distance on lights in the pre- construction period, $\delta_{1i} + \delta_{2i}$ is the impact in the construction period, and $\delta_{1i} + \delta_{3i}$ is the effect in the post-construction period:

Table 3.6 shows the impact of distance over these three periods. In the pre-construction period, light density would fall by 3.389 log points for every increase in 1000km from the straight line, but once construction starts, this falls by 0.955 log points to about 2.434 log points, and in the post-construction

[16]The excluded regions include the states of Jammu and Kashmir, Sikkim, Assam, Arunachal Pradesh, Meghalaya, Mizoram, Tripura, Nagaland, Andaman and Nicobar Islands and Lakshwadeep.

[17]Source: National Highways Authority of India http://www.nhai.org/completed.asp

period it's even lower at about 2 log points. All measures are statistically significant, and show that the impact of distance on development $falls$ around the turn of the century, after the highway construction begins. If the upgrades made the highways more important, then we should expect the opposite result – that they should matter more for economic activity. This does not necessarily indicate that the highway construction *caused* the relationship to dissipate, as the relationship was already weakening over time. If anything, upgrading the highways did not change the rate at which this relationship was weakening over time.

To see at what distances the change in impact appears, I split up all positive distances into 8 equal quantiles Ψ_i, and interact them with indicators for being in the 'construction' phase or the 'post-project' phase. In the regression equation below τ_t are year fixed effects and $\mathbf{X}$ is a vector of geographic controls:

$$LogLights_{it} = \alpha + \tau_t + \beta\mathbf{X} + \psi_{\mathbf{1i}}\mathbf{\Psi_i} + \psi_{\mathbf{2i}}\mathbf{\Psi_i} * Construction_t + \psi_{\mathbf{3i}}\mathbf{\Psi_i} * Post_t + \epsilon_{it} \tag{3.5}$$

The omitted category in this regression are the sub-districts that are on the straight-line (a little more than 5% of all sub-districts). $\psi_{\mathbf{1i}}$ traces out the impact of distance from these sub-districts in the pre-construction period, whereas $\psi_{\mathbf{1i}} + \psi_{\mathbf{2i}}$ is the impact during the construction phase. These coefficients can be plotted for each distance quantile to look at the the semi-parametric impact of distance, and how that changes in the three time periods.

The lines in Figure 3.7.4a show the impact on light-density by distance quantiles, relative to sub-districts that touch the straight-lines. The blue lines are for the pre-construction period, the orange for the construction period, and the green lines are for when the project is over. Looking at the pre-construction period in panel (b) we can see that a district in the eighth distance quantile has about 2 log points less light density than a district that is on the line. But once construction begins, these lines start flattening out. Together these results seem to suggest that while an increase in distance from the straight line leads to less development, this relationship weakens in the later period, and especially after the construction of the highway.

Figure 3.8.5 reproduces this result for other measures of light-density. The panel on the standard-deviation of lights in Figure 3.8.5 shows that there is a larger dispersion of lights within a region that is closer to the route. This gives us some indication towards the pattern of development in these regions – that in sub-districts near the route there are a few large towns with a lot of activity, and then areas with very little activity. In regions away from the route however, there is an equal amount of low economic activity. This pattern is consistent with developed regions reflecting agglomeration economies, where activity is concentrated in certain areas but is sparse in other regions Krugman (1991).

Ghani et al. (2015), show that upgrading the highways induced new manufacturing firms to enter in regions close to the highway. The difference in results are not due to the methodology, and must therefore be because of different outcome variables under analysis.[18] Their paper looks at the organized manufacturing sector, for about half the districts in the country and shows that there was an increase in entry for such firms in regions within 10kms of the highway.[19] Together our papers, therefore show that while there was an increase in the entry of organized manufacturing enterprises, overall economic activity was still shifting away from the highways.[20]

This spread in economic activity may have been stemmed if upgrading the highways induced enterprises to stay or enter in regions on the highway at a higher rate than other regions, like it happened for firms in the organized manufacturing sector. There is, however, little evidence in this paper to show that upgrading the highway system actually turned around the trends that were already visible in the data for overall economic activity.

One way to finally determine whether the GQ upgrades caused these trends is to look at another route that did not have heavy investments in the highway system: the diagonal of the quadrilateral between Mumbai and Kolkata. Table 3.5 shows that the elasticity between economic activity and distance to a straight line connecting these two cities has been dissipating over time, despite the fact that there was no projects under the National Highway Authority of India to upgrade the routes between these two cities. If this is an indication of the trends seen along routes connecting the other nodal cities, then the construction of the highway may not have had much of an impact on the pre-existing trends under which economic activity was already spreading geographically away from the well developed regions connected to the highway. These mechanisms are explored further when studying the geographic spillovers within the confines of a modified Neoclassical growth model that I set-up in Section 3.5.

3.4.2 The Nature of Development

While light-density may be a good proxy for overall economic activity, it tells us little about what the nature of this economic activity is. Unfortunately, the only source of data that provides counts at the sub-district level is the Census of India which has a limited number of outcome variables and is only compiled once every 10 years. Table 3.8 shows results from the 2001 and 2011 Censuses. While regions away further away from the route have less population overall, they are more likely to have a higher concentration of Scheduled Tribe (ST) persons. Along with the Scheduled Castes (SCs) these are among the most socio-economically disadvantaged groups in the country. Furthermore, regions further away

[18] While Ghani et al. (2015) use an OLS specification for their main results, they show that their main specification is robust to using an instrumental variables approach. Unlike this paper, instead of using a continuous "distance" measure, they use two discrete categories – compare districts between 0 and 10km near the road to districts further away with a 1/0 indicator for whether you are within 10kms of the highway. When I use their methodology, I still find that the relationship between night-lights and distance to the road dissipates over time.

[19] Their sample is only for states that had enough manufacturing activity, and in districts that were observed over their entire panel.

[20] One possibility that Ghani et al. (2015) mention in relation to our papers is the change in the composition of economic activity – for example, organized manufacturing firms may move closer to the highway, but other firms may not.

also have a higher concentration of cultivators, but lower concentration of persons engaged in household industry. While distant regions also have lower literacy rates, they seem to have a more equitable distribution of literacy across genders – the gender-gap defined as the difference in the male and female literacy rates is lower. Unfortunately, without more detailed data at the sub-district level, it is hard to discern any intricate patterns in the nature of economic activity, but it is clear that regions farther away from the route have a higher concentration of marginalized populations, lower literacy, and have more cultivators but less persons engaged in household enterprises.

The question of what kind of economic activity the night-lights are picking up over the long-run will provide more information on the kind of development. If the regions further away from the routes were previously uninhabited and the routes allowed people to locate there, then we should see a rise in population for those regions. If, however, it is merely the composition of the population and the kind of economic activity undertaken by them, then for a given population, the increase in night-lights will represent more wealth per capita being generated.

To get at the question of whether there are population changes to less inhabited regions or whether the changes over time are picking up increases in economic activity per capita, I use the LandScan data on population estimates. The data compiled by the US Department of Energy's Oak Ridge National Laboratory, uses sub-national Census counts and primary geospatial ancillary datasets, including land cover, roads, slope, urban areas, village locations, and high resolution imagery analysis of settlements, to predict the populations at a finer geographic level than available elsewhere.

Table 3.4 presents the elasticity between population and distance to the nearest straight-line connecting the nodal cities over time (starting with the data in 2002). While the elasticity is high (between 0.18 and 0.2) for this period, there is no change in the elasticity over time. This shows that the change in the night-lights elasticity over time is driven by something other than the number of people migrating to newer areas, and instead be due to changes in per-capita economic activity.

3.5 A Simple Model of Spatial Development

Better access to transportation networks can induce development in connected regions by facilitating trade and migration, the spread of technology and ideas, and reducing price volatility. So far from the results it is clear that till the early 1990s, being near the transit network that connects the four major cities had significant impacts on regional development. However, the impact of being close to the straight lines dissipates over time, even after the highway system is upgraded. Why, then, does the effect of the transit network dissipate? One possibility is that these routes affect the initial level of development more than the steady state level – in such a situation, a simple Neoclassical growth model would predict that underdeveloped regions would grow faster and there would be convergence in incomes across regions.

In such a model, the dissipation has nothing to do with spillovers. If we build in the possibility of spillovers in economic activity across regions, this not only spurs the growth in underdeveloped regions but also increases the overall impact of these routes. As I will show with my structural estimates, the evidence strongly supports a model in which there are economic spillovers in activity across neighboring regions.

3.5.1 Distance to the Network and Neoclassical Growth

First, let us consider the model that is not driven by spillovers. While transit networks may be important for regional development, these impacts may dissipate over time on their own. After the four nodal cities were established, the regions that had the least-cost connections to these cities started developing. This helped build up a network structure whereby regions connected to these growing regions started growing. Since regions connected to the routes started growing earlier, they are at any point of time closer to their steady-state level of development than regions further away which will hence be growing faster. If we then look at long-term development, the first set of regions would have more economic activity, but their neighbors and their neighbors' neighbors would be catching up over time. This would then produce the dynamic trends seen in the data, and is consistent with the results found in Banerjee et al. (2012) where they find that Chinese regions near the straight-lines had reached a higher level of GDP per capita but were not necessarily growing faster than other regions.

There are a few ways to encapsulate transit networks into a Solow-style growth model. One possibility is that distance to networks affects the steady-state level of development, which would then predict regions closer to the highways would grow to a higher level of income than regions further away. Another possibility is that the distance to transit networks only affects the initial level of income, and all regions converge to a similar steady-state level of economic activity. This would then be consistent with a result that shows that regions further away from the route have higher growth rates. To formalize this framework, we can modify the empirical predictions of the Neoclassical growth model in the following way:

Let y_{td} be income per effective worker in sub-district d and time t. A region on the route is a simple Solow-style economy with income $\bar{y}_t$. A region that is distance D from the route can be characterized in relation to the region on the network:

$$y_{td} = D^{\alpha_t}\bar{y}_t \tag{3.6}$$

Here $-1 < \alpha_t < 0$ captures the elasticity of income with respect to distance from the route. This framework is similar to gravity-models in trade theory, where the distance would be picking up trade-costs and other frictions. Given that the evidence so far shows that this elasticity is higher in earlier years,

α_t can be simply represented by:

$$\alpha_t = \lambda + \frac{\psi}{(1+t)} \tag{3.7}$$

Here $-1 < \lambda < 0$ captures how distance to the transit network affects the steady-state level of income, and $\lambda + \psi$ the initial level of income. One test is to see whether these routes have a larger effect on the initial level of income or the steady state level. If, and only if, the routes affect the initial levels of income more than the steady state levels, then $\psi < 0$ as well, and we should see that regions further away from the route grow faster. From the results in Table 3.3 it is clear that $\lambda + \psi < -0.497$ and $\lambda > -0.212$. Together, this implies that $\psi < -0.285$.[21]

3.5.2 Convergence

In the results so far, we see that while regions along the route are richer, there is a catching-up of regions further away. The Solow (1956) model's predictions of a conditional (on parameters) convergence of per-capita incomes may be used as a framework to study these patterns in the data. Barro and Sala-i Martin (1992) are careful to distinguish between the different notions of convergence. If $\beta-$convergence holds then poorer sub-districts would be growing faster than their richer counterparts. The Neoclassical growth framework has certain predictions for the rate of convergence to steady state β. As discussed in Barro and Sala-i Martin (1992, 2004); Mankiw et al. (1992), the solution to the income dynamics can be characterized by:

$$log\ y_{td} = (1 - e^{-\beta t})log\ y_d^* + e^{-\beta t}log\ y_{d0} \tag{3.8}$$

Given this setup, there are a few ways to then estimate the rate of convergence. One possible approach is to use a cross-sectional regression based on the long difference between the first and last year in the data, as explained by Mankiw et al. (1992):

$$log\ y_{td} - log\ y_{d0} = \gamma_0 + (1 - e^{-\beta t})log\ y_{d0} + \gamma_x \mathbf{X} + u_{dt} \tag{3.9}$$

β can be estimated from the coefficient on initial income. Another approach was used by Evans (1997); Higgins et al. (2006) amongst others. Here, instead of using the cross-section, they use all the years and back out the rate of the convergence:

$$growth_d = \delta_0 + \delta_y log\ y_{d0} + \delta_x \mathbf{X} + u_{dt} \tag{3.10}$$

Here, $growth_d$ is the average growth rate across all the years for a region d. The rate of convergence over T periods is then $\hat{\beta} = 1 - (1 + \hat{\delta}_y T)^{\frac{1}{T}}$.

[21] These estimates are bounds, rather than precise estimates since we neither observe year 0 nor the steady-state, which lies in the distant future. If we assume that 1992 is the 'initial' period, or that the steady state has been reached by 2012, then $\lambda = -0.212$ and $\psi = -0.285$.

In Table 3.7, I present the results using both methods. The Mankiw et al. (1992) cross-sectional method shows a convergence rate of 4% a year, whereas the Evans (1997); Higgins et al. (2006) method displays a rate of 3.9% a year. These are both double the 2% rate that we see in cross-country convergence rates (Sala-i Martin, 1996). While capital mobility within the country can aid rapid convergence, mobility in labor across regions in India are low (Munshi and Rosenzweig, 2016). Rates of convergence could be higher within countries due to the presence of spillovers in economic activity across regions.

Lastly, $\beta-$convergence is a necessary but not a sufficient condition for a decrease in the variance of incomes across regions – also known as $\sigma-$convergence (Barro and Sala-i Martin, 1992). Figure 3.7.1b shows that between 1992 and 2002 the variance in light density had shrunk a lot, but this did not continue for the next decade.

3.5.3 Spillovers and the Direction of the Spread of Development

While a Solow-style model discussed in Section 3.5.1 is consistent with the dynamic trends, it says nothing about spillovers in economic activity across regions. It does not necessarily predict that as one region gets richer that will $lead$ to its neighboring region getting wealthier as well over time. In this section I explicitly show strong evidence of such spillovers, and estimate the parameters of a simple model that will help quantify the extent of these spillovers and the overall benefits of these routes. In such a model, distance to the highway not only directly affects the income levels (as in the Neoclassical growth framework) but also sets into motion certain dynamics that affect the spatial development of neighboring regions as well.

As a descriptive motivation, in Figure 3.7.3 we can follow the regions around the Mumbai-Chennai over time. The pictures show the regions along the highway (depicted by a blue line) and along the straight-line path (red line) every 5 years, in a Green to Red spectrum (where deeper green reflects lesser economic activity, and red indicates more activity). It is clear that the regions to develop first are the ones along the route, after which economic activity seems to fan out to neighboring areas, and then the neighbors' neighbors and eventually reaching areas farther away.

If economic activity spreads from regions close to highways to regions away from the highway, then a given region (d) should be affected more by economic activity in neighboring regions that are closer to the highway ($d-1$) than neighboring regions that are further away from the highway ($d+1$). Let $Log\ y_{td}$ be income in region d at time t, $Log\ y_{t,d-1}$ be the mean of the lights for all it's neighboring regions that are closer to the highway, and $Log\ y_{t,d+1}$ be the mean of lights for neighboring regions farther away from the highway. In the regression below, we should then expect $\delta > \gamma > 0$ if economic activity is spreading from regions closer to the highway to regions away:

$$Log\ y_{td} = \alpha_d + \delta Log\ y_{t,d-1} + \gamma Log\ y_{t,d+1} + \epsilon_{td} \tag{3.11}$$

While this regression formulation shows a contemporaneous impact, the effect could also have a period lag, when activity in a region today can affect economic activity in a neighboring region tomorrow. Furthermore, the true relationship could be one of changes, where changes in a region's economic activity affect changes in it's neighbor's activity. It is important to stress however, that the true test of the model is one where $\delta > \gamma$, and not merely if $\delta > 0$ since $\delta > 0$ can also be consistent with a model of spatial correlation in income shocks. The regressions for contemporaneous spillover effects, as well as the one-period lagged effect and the changes over time specification are shown in Table 3.9, where it can be seen that δ is always statistically and economically significantly greater than γ.

Studying the pattern of lights in regions along the route, and their neighbors would help answer the question of how spillovers are leading to convergence. In order to study this, let us define a 'degree-of-separation' (s) as how many regions lie between your region and the route. For example, $s = 1$ means the region neighbors a sub-district that lies on the route, and $s = 3$ means that the region is a neighbor of a neighbor of a neighbor of a region that lies on the route. Plotting the coefficients (β_{st}) of the regression below for each s and over time t allows us to study whether convergence takes place across neighboring regions. In the specification below, $\mathbf{1}_{s=S}$ is an indicator function that depends on the 'degree of separation' of the given region, and $Year_t$ is an indicator variable for the year:

$$Log(lightdensity)_{dst} = \alpha + \beta_{st}\mathbf{1}_{s=S} * Year_t + \epsilon_{dt} \tag{3.12}$$

In Figure 3.7.5 each point is the differential impact of light density on that region compared to other regions in a given year. It can be seen that regions on the straight-line connecting two major cities have the highest light-density compared to other regions, and the regions that are one-degree of separation away are only slightly worse off. Figure 3.7.5, therefore, indicates a few things. First, the ordering in term of 'degrees of separation' is maintained, whereby regions closer to the route have higher light-density, and even though there is convergence over time, we don't see an 'over-taking' by the regions further away. Second, while it seems like the 1990s were a period where convergence was rapid, this has slowed down in the later half of the 2000s. At the end of the period, regions still maintain their initial ordering in concentration of light-density, an ordering which directly depends on their 'degree-of-separation' from the route. Third, it is clear that distance to the route strongly affects initial income levels, as seen by the dispersion in incomes in the first year of the data. Last, distance to the route also affects the steady state level of incomes, as the same ordering remains in the last year of the data despite a slow down in the rate of convergence in the last few years.

3.5.4 Testing the Model and Estimating Spillover Parameters

A strong test for the magnitudes of these spillovers is to see how a structural model of the spillovers correctly predicts the reduced form impacts of the routes in each year. To formalize this, let y_{tk} represent income in region d and $y_{t,d-1}$ be the mean income for all it's neighbors that are closer to the route than region d. For a given distance from the route D_d, let the true relationship between income and distance be a modification of equation (3.6):

$$y_{td} = (D^{\mu_t}\bar{y}_t)\, y^{\chi}_{t,d-1} \tag{3.13}$$

In this relationship, $0 < \chi < 1$ represents the neighborhood spillover effect of economic activity from bordering regions closer to the highways than the given region, and $-1 < \mu_t < 0$ represents the direct effect of being further away from these routes. This equation can be solved forward recursively to:

$$Log\ y_{td} = (\sum_{j=0}^{d} \chi^j)\mu_t Log\ D_d + Log\ \bar{y}_t \tag{3.14}$$

$$= \alpha_t Log\ D_d + Log\ \bar{y}_t \tag{3.15}$$

We therefore recover equation (3.6), showing that the model with spillovers is a general form of the model without. Equations (3.14) and (3.15) can be used to estimate parameters χ,μ_t and α_t, and we can test the following model prediction:

$$\alpha_t = (\sum_{j=0}^{d} \chi_t^j)\mu \tag{3.16}$$

Table 3.10 shows results for both Equations (3.14) and (3.15).[22] The distance-spillover parameter $(\sum_{j=0}^{d} \chi_d^j)\mu$ according to equation (3.13) is also presented, and is almost identical to α_t as seen in equation (3.14). This is a strong test of the model that substantial neighborhood spillovers across regions exist in this context. The spillover parameter χ is relatively stable over time and lies within the range of 0.77 and 0.868, whereas the direct-distance effect μ_t is initially -0.104 and falls to about -0.064. The distance-spillover parameter $\alpha_t = (\sum_{j=0}^{d} \chi_t^j)\mu$ is initially 0.5 and halves to about 0.25 by 2012. All these parameter estimates are statistically indistinguishable from 0 in all years.

Figure 3.7.6 shows how well the model fits the data by plotting both the reduced form elasticity of light-density with respect to distance, and the corresponding structural parameter based prediction of this elasticity from a model that incorporates these spillovers. The model fit is good throughout this period, lending strong credibility to the structural assumptions here.

Finally, how this relationship changes over time tells us about how these spillovers can actually lead to convergence across regions. In the cross-country version of the Solow model, the convergence across regions could be generated without any spillovers, and often leads to rates of $\beta-$convergence of about 2% (Sala-i Martin, 1996). Within a country, however, spillovers across regions can speed up convergence, and generate rates like 6% for US counties (Higgins et al., 2006), or 4% as in the case of this paper.

[22]For this test, we need to use the fact that on average regions have about five to six degrees of separation between the route and themselves (i.e. d is approximately six).

3.5.5 Income Elasticities with Spillovers

These results together then allow us to back out the income elasticities with respect to distance from these routes, after accounting for the presence of these spillovers. To reiterate, ignoring the presence of spillovers underestimates the true effect of these routes. As economic activity spills over to regions further away, the difference in incomes for a region on a route to the region further away falls. Measuring the difference between the two regions may then produce a small number. The issue here is that by ignoring the spillover we are subtracting it from the overall effect of the route, when in fact we should be adding these spillovers to the total impact.

At the bottom of Table 3.10, I present the income elasticity ignoring the spillovers and the elasticities taking them into account. In 2012, if we ignore the spillovers in our calculations, we would have estimated an elasticity of 0.06 – a small number, suggesting that investments in transit networks may have low returns. Taking the spillovers into account raises the income effects of routes by more than three times to 0.21. This suggests, if we ignored spillovers we would say that a 10% decrease in distance from the route, raises incomes by only 0.6%, but incorporating spillovers, we can see that $overall$ incomes rise by 2.1% – a substantial amount by many measures.

In Appendix 3.9, I explore the mobility of capital and labor to see whether standard models and evidence can pin down how they contribute to this spread of economic activity over time.

3.6 Conclusion

The impact of transportation infrastructure on regional development has been a long debated discussion. In general, better transit networks have been thought to facilitate trade, migration, the spread of ideas and technology, credit and other financial opportunities, and decrease price differentials and volatility. Studying infrastructure projects in different contexts have however provided contradictory evidence. Fogel's (1964) study of US historical development argues that there were limited impacts of railways on growth relative to the transportation networks that used waterways, whereas Hirschman's (1969) treatise posits that social overhead capital, like railways, have significant linkages that promote growth in industries. For Hirschman, infrastructure projects would have forward linkages (promote industries that need roads and railways), backward linkages (promote industries that supply materials for road and rail construction) and lateral linkages (connect industries together). The Fogelian view, on the other hand, supports the idea that much of US historical investment in railways was misguided and therefore did not have impacts on development because of governmental policies that subsidized railway construction. The natural experiment under analysis in the Indian context, however, is that some regions happened to be on the path of shortest distance connecting major centers of economic activity, and we are hence less likely to find 'misguided' investments in this context.

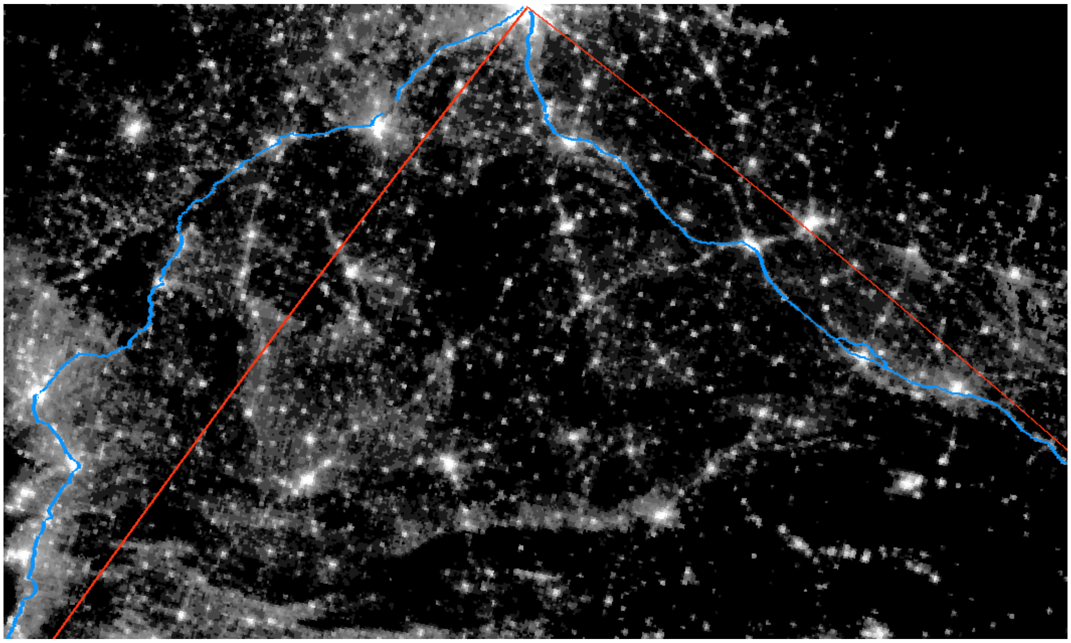
The results in this paper indicate that while distance to the straight line may have significant impacts on regional development till the early 1990s, the strength of this relationship dies out slowly over the next two decades. If one was to analyze the relationship at the end of this period, they would come to the conclusion that there is little economically significant impacts of being near a transportation network, which would support the Fogelian view. However, the relationship in the early 1990s shows how this is not the case. The question then arises, as to what happened in the two decades that weakened this relationship.

The period of study was one of rapid economic growth and development after reforms that liberalized the market structure, cut down on the license-permit bureaucracy and integrated various industries with world markets. It was also a period of upgrading the existing transportation network by strengthening the highway system. While the first set of regions were to benefit from being directly connected to the cities by being on the least-cost path of connectivity, over time other regions would establish indirect connections via these already connected regions. This would then lead to a protruding network structure that would link regions and spread development by lowering the costs of trade and exchange. This explanations can be ensconced in a growth-model framework by formalizing and estimating the extent of the spatial spillovers. The results in the paper indicate that a substantial amount of the high-rates of $\beta-$convergence can be explained by spatial spillovers in economic activity across neighboring regions.

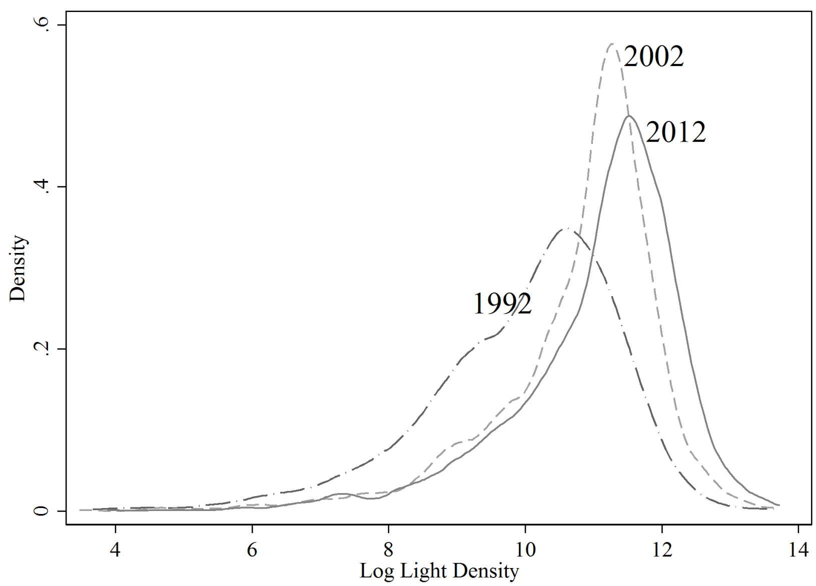
As I show, ignoring these spillovers would have produced estimates of the impacts of these routes that are only 27% the size of the true overall effect – a gross underestimate. Furthermore, the implications for policy in the light of such spillovers can be crucial. While the initial transit networks did a lot to encourage economic activity in connected regions, future investments in upgrading these highways did little to help these regions indicating that investments in these highways had reached a portion of diminishing returns. However, the initial investments in the highways not only helped develop connected regions, but also led to spillovers in activity to neighboring regions. Together, these results indicate that policy-makers should try to connect more regions rather than upgrade routes on already connected regions.

Last, the existence of large spillovers can explain why past research on roads have found little impacts. An empirical strategy that compares regions along a route to neighboring regions will provide underestimates of the true impacts because economic activity may have spread to these neighboring regions by then. The existence of the spillovers indicate that the overall impacts of routes can be larger than previously thought, since roads can affect development not only on regions along the route, but regions farther away as well. In order to capture this spread of activity, therefore, it is important to study the impacts over time and pin down the entire pattern of spatial development, as I do in this paper.

3.7 Tables and Figures



(a) Night-time lights, highways and straight-lines between Mumbai-Delhi and Delhi-Kolkata (The blue-line traces the actual route of the highway, and the red-line indicates the straight-line path between the major cities.



(b) Distribution of Log Light Density Over Time

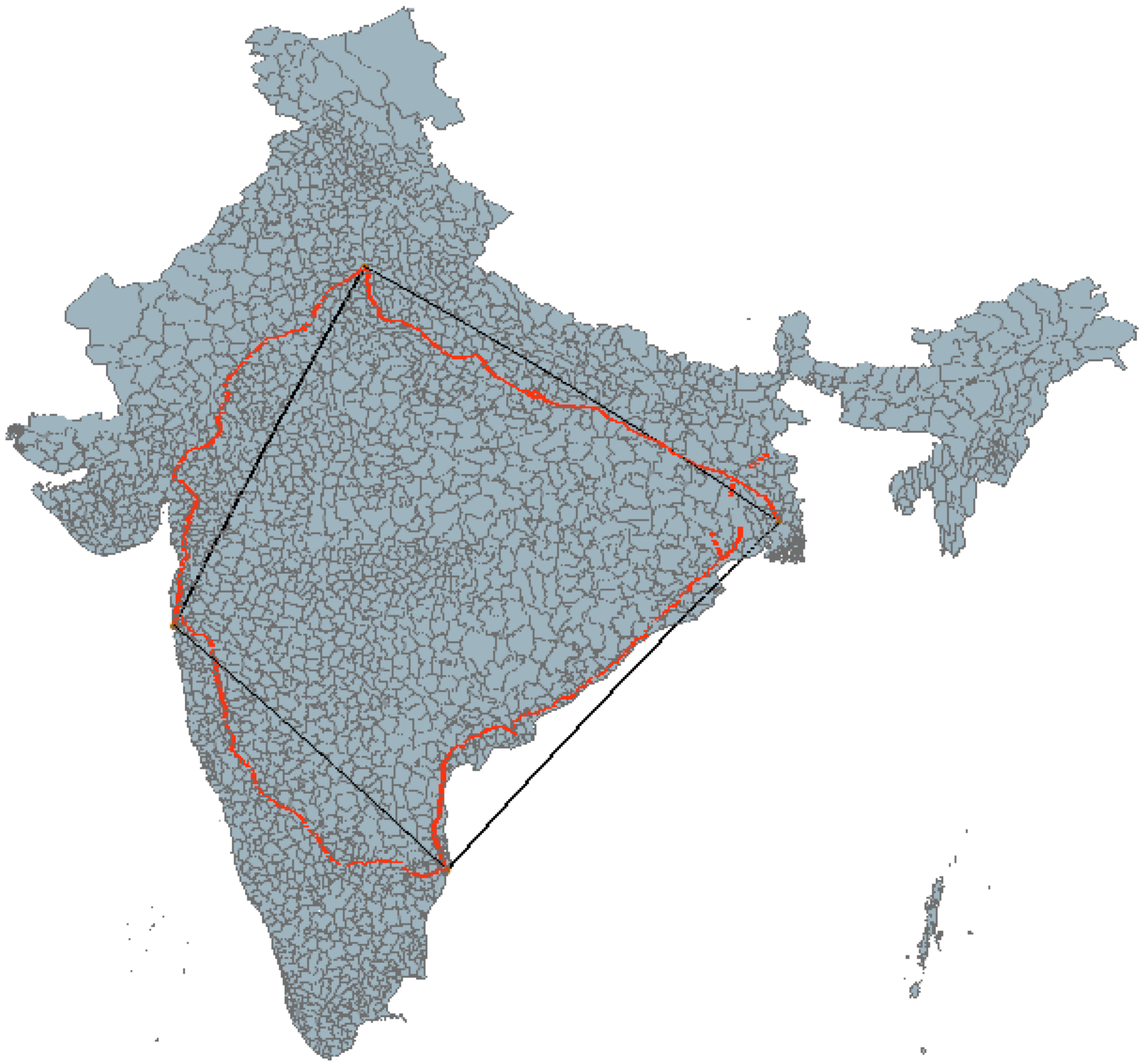
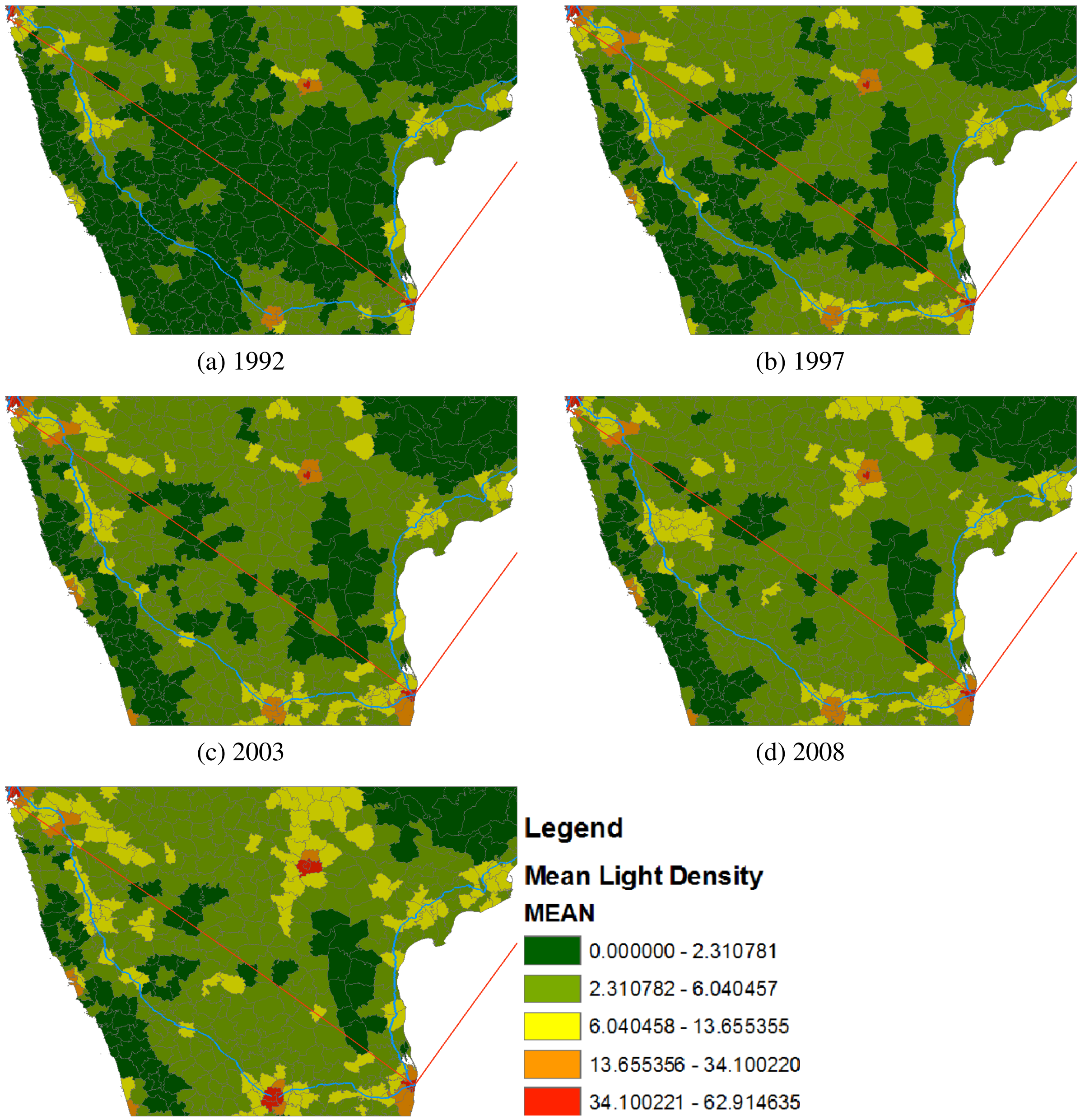


Figure 3.7.2: Golden Quadrilateral Highways and boundaries of the sub-districts used in the analysis

Figure 3.7.3: Spread of Lights From the Mumbai-Chennai Highway

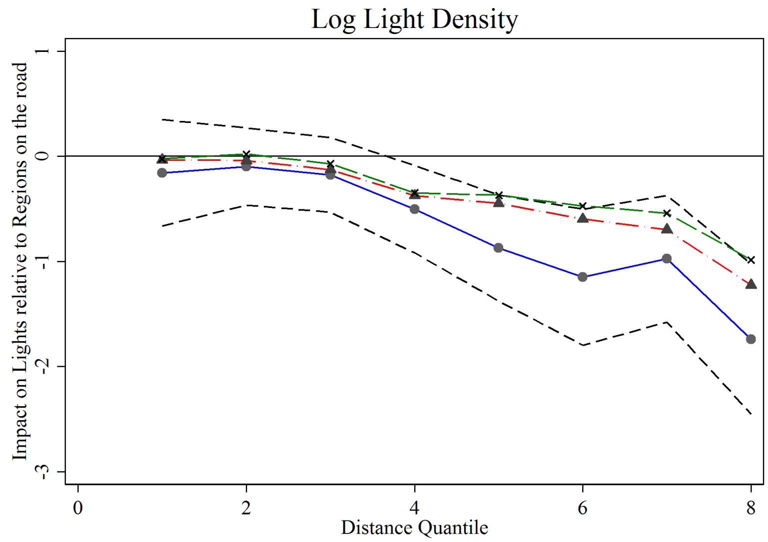


(a) 1992 (b) 1997

(c) 2003 (d) 2008

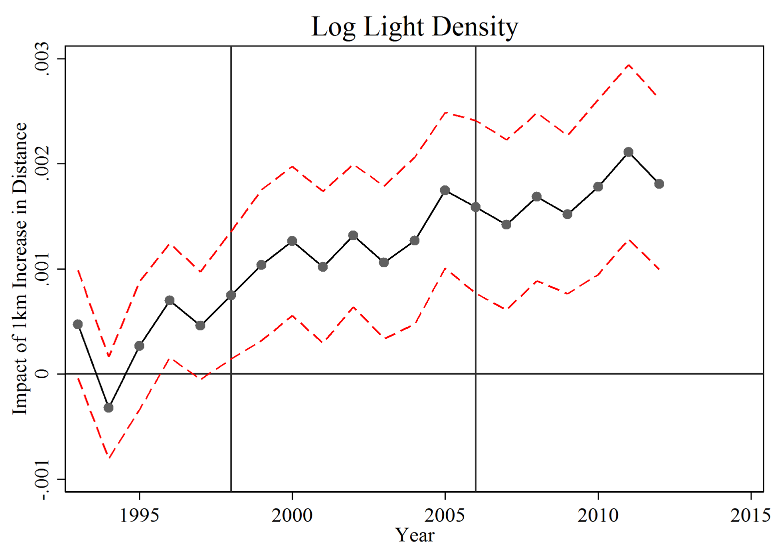
(e) 2012 (f)

Displaying the spread of night-time light density every five years. The Blue Line indicates the actual path of the Highway, and the Red Line indicates the straight line connecting 2 major cities. The legend of light-densities is on a green to red spectrum, where green is relatively less light-density and red is a higher level of light-density



(a) Light Density (per sq miles)

The graphs show the impact of distance on night-time lights relative to sub-districts that have any portion of the sub-district touching the straight-lines. The blue lines are for the pre-construction period, the orange lines for the construction period and the green lines for the post-construction period. The standard error bands are for the pre-construction (blue) lines and clustered at the district level. The 'Distance' axis consists of 8 quantiles of equal size. The distance quantile cutoffs are roughly as follows: 0 to 40kms, 40 to 90 kms, 90 to 135 kms, 135 to 200kms, 200 to 260kms, 260 to 340 kms, 340 to 440 kms, and above 440kms. Different measures of night-time lights are presented in Figure 3.8.5. Robustness checks are presented in Figure 3.8.5



(b) Light-Density:1992 coefficient -0.00406

Coefficients of change in impact relative to 1992. Standard errors calculated at the district level. Vertical lines represent the phases of construction - 1999 is when the highways started being built. There were delays till 2001 when most work started, and 2006 is when most work was completed. To interpret the graph: the mean impact of a 1km increase in distance from the highway was a 0.00406 fall in light-density, and this impact has been dissipating over time. By 2012 the impact of a 1km increase in distance from the highway had become -0.00406+0.00205, or about -0.00201. Different measures of night-time lights are presented in Figure 3.8.4. Robustness checks are presented in Figure 3.8.6.

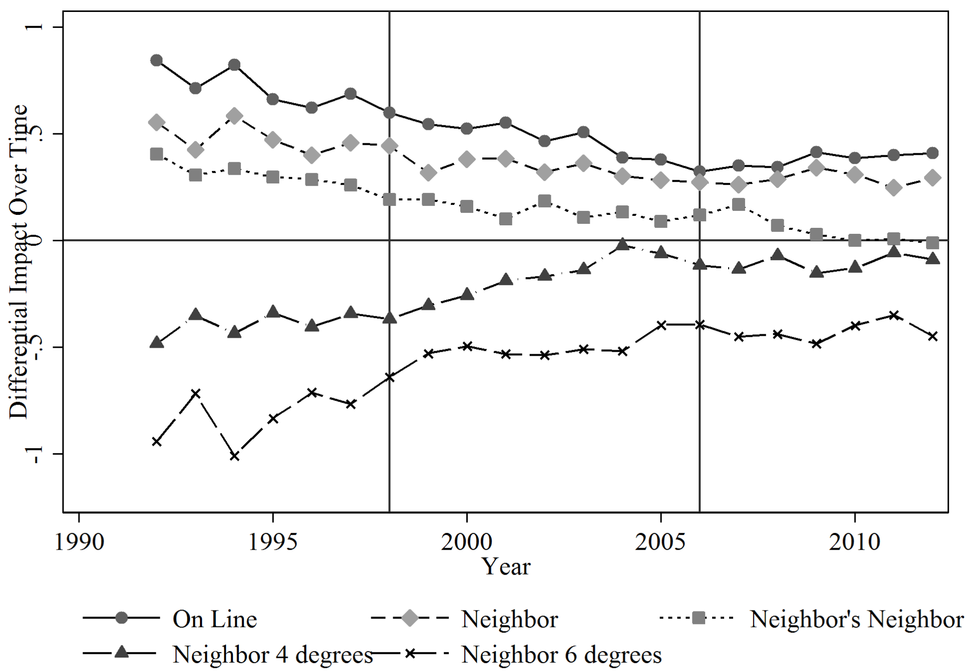


Figure 3.7.5: Relative light density for regions on the line, and their neighbors

Relative light density calculated as Log(light density) for that region relative to all other regions. “On Line” represents regions on the straight-line path between two major cities. “Neighbor” represents sub-districts that are neighbors of “On Line” sub-districts. “Neighbor’s Neighbor” represents neighbors of neighbors of “On Line” regions, and so on. “Neighbor 6 degrees” represents regions that are removed from “On Line” regions by more six-degrees of separation.

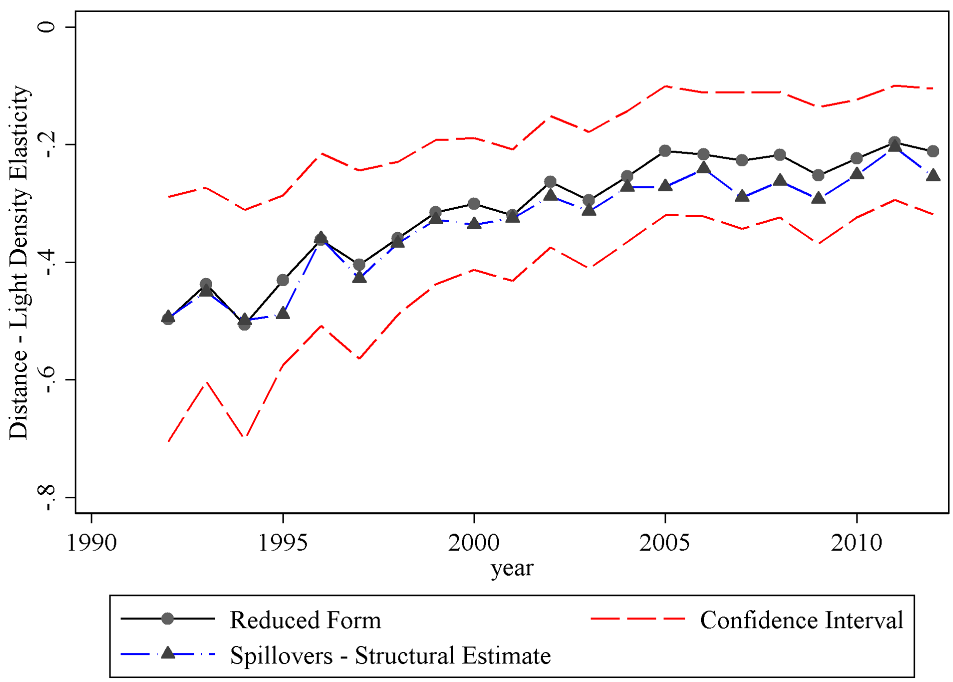


Figure 3.7.6: Spillovers & The Elasticity of Light Density with Respect to Distance

Elasticity of Light Density with respect to Distance to straight-line paths connecting historical major cities. The 'Reduced Form' line and corresponding confidence intervals plot the coefficient from the regression $Log\ LightDensity_{it} = \alpha_t Log\ Distance_{it} + \beta X + \epsilon_{it}$. The 'Spillovers - Structural Estimate' line plots the corresponding model-based elasticity as discussed in Section 3.5.4. As can be seen, the model fit is good as the structural estimates closely correspond to the reduced form estimates throughout this period.

Table 3.1: Predicting Distance to Routes with Distance to Straight-Lines

	Distance to GQ Highway		Distance to Railroad	
	Coefficient SE	First Stage F Stat	Coefficient SE	First Stage F Stat
Distance to Line	**0.81**		**0.055**	
SE clusters:				
Sub-district	(0.00983)	6796	(0.0215)	6.570
District	(0.0205)	1569	(0.0238)	5.348
State	(0.0372)	474.7	(0.0334)	2.707
R-squared	0.791	0.738	0.071	0.0516
Observations	2253	2253	2253	2253
Controls	Y	Y	Y	Y

Level of observation - sub-district

Dependent variable 'Distance to GQ Highway' is the nearest geo-distance between the the sub-district and the closest Golden Quadrilateral highway

Dependent variable 'Distance to Railroad' is the nearest geo-distance between the sub-district and the closest railway line

Independent variable 'Distance to Line' is the nearest geo-distance between the sub-district and closest straight-line connecting nodal cities

Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table 3.2: Relationship Between Night-Time Light Density and Distance to Transit Networks

	Log Light Density		
Year: 1992	**OLS**	**Reduced Form**	**IV 2SLS**
Distance (100km)	**-0.391**	**-0.406**	**-0.501**
SE Cluster:			
Sub-district	(0.0492)	(0.0479)	(0.0593)
District	(0.0784)	(0.0749)	(0.0933)
State	(0.177)	(0.159)	(0.194)
R-squared	0.144	0.152	0.141
Year: 2002	**OLS**	**Reduced Form**	**IV 2SLS**
Distance (100km)	**-0.224**	**-0.233**	**-0.287**
SE Cluster:			
Sub-district	(0.0332)	(0.0326)	(0.0404)
District	(0.0476)	(0.0442)	(0.0551)
State	(0.102)	(0.0884)	(0.109)
R-squared	0.188	0.193	0.185
Year: 2012	**OLS**	**Reduced Form**	**IV 2SLS**
Distance (100km)	**-0.193**	**-0.188**	**-0.232**
SE Cluster:			
Sub-district	(0.0316)	(0.0316)	(0.0391)
District	(0.0457)	(0.0432)	(0.0535)
State	(0.0849)	(0.0782)	(0.0965)
R-squared	0.156	0.157	0.155
Controls	Y	Y	Y
Observations	2,253	2,253	2,253

Level of observation - sub-district
Independent variable ‘Distance’ in OLS specification is ‘Distance to GQ Highway’ – the nearest geo-distance between the the sub-district and the closest Golden Quadrilateral highway
Independent variable ‘Distance’ in Reduced Form specification is ‘Distance to Line’ – the nearest geo-distance between the sub-district and closest straight-line connecting nodal cities
First stage of IV 2SLS specification is shown in Table 3.1 – Distance to GQ Highway is instrumented with Distance to nearest straight-line connecting nodal cities
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.
Dependent variable is Log(0.01+lights/area)

Table 3.3: Reduced Form and 2SLS: Elasticity of Lights, Distance and GDP

Reduced Form			**Log Light Density**			
Year	**1992**	**1996**	**2000**	**2004**	**2008**	**2012**
Log(Distance)	**-0.497**	**-0.362**	**-0.3**	**-0.254**	**-0.217**	**-0.212**
SE clusters:						
Sub-district	(0.0591)	(0.0462)	(0.0379)	(0.0365)	(0.0365)	(0.0364)
District	(0.106)	(0.0747)	(0.0571)	(0.0567)	(0.0543)	(0.0545)
State	(0.226)	(0.151)	(0.116)	(0.109)	(0.0944)	(0.0921)
R-Squared	0.135	0.160	0.170	0.196	0.154	0.147
GDP-distance elasticity	**0.1491**	**0.1086**	**0.09**	**0.0762**	**0.0651**	**0.0636**
Bootstrapped SE	(0.0164)	(0.0126)	(0.0110)	(0.0106)	(0.0119)	(0.0105)
IV - 2SLS			**Log Light Density**			
Year	**1992**	**1996**	**2000**	**2004**	**2008**	**2012**
Log(Distance)	**-0.735**	**-0.534**	**-0.444**	**-0.375**	**-0.321**	**-0.313**
SE clusters:						
Sub-district	(0.0852)	(0.0670)	(0.0546)	(0.0519)	(0.0524)	(0.0523)
District	(0.155)	(0.109)	(0.0819)	(0.0795)	(0.0772)	(0.0776)
State	(0.338)	(0.226)	(0.175)	(0.162)	(0.141)	(0.137)
R-Squared	0.124	0.148	0.164	0.202	0.159	0.155
GDP-distance elasticity	**0.2205**	**0.1602**	**0.1332**	**0.1125**	**0.0963**	**0.0939**
Bootstrapped SE	(0.0237)	(0.0207)	(0.0174)	(0.0167)	(0.0163)	(0.0140)
Observations	2,253	2,253	2,253	2,253	2,253	2,253
Controls	Y	Y	Y	Y	Y	Y

*Estimates of GDP-distance elasticity rely on elasticity of GDP-lights being 0.3 for low-middle income countries. Therefore to find elasticity of GDP-distance, multiply the coefficient with 0.3.
Level of observation - sub-district.
Independent variable ‘Log (0.01+Distance to Line)’ is the nearest geo-distance between the sub-district and the closest straight-line connecting Mumbai and Kolkata - two historical cities that have not had direct transit networks connecting them.
Dependent variables is Log(0.01 + Light density). ‘Lights per area’ normalizes the sum by the surface area of the district.
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table 3.4: Elasticity between Population and Distance to Line

		Log (Population)				
Year		**2002**	**2004**	**2006**	**2008**	**2010**	**2012**
Log(Distance)		**-0.226**	**-0.182**	**-0.178**	**-0.179**	**-0.176**	**-0.191**
SE clusters:							
	Sub-district	(0.0221)	(0.0293)	(0.0290)	(0.0290)	(0.0305)	(0.0305)
	District	(0.0404)	(0.0560)	(0.0543)	(0.0545)	(0.0562)	(0.0559)
	State	(0.0838)	(0.0991)	(0.0968)	(0.0969)	(0.0977)	(0.0970)
R-Squared		0.091	0.056	0.053	0.053	0.044	0.046
Observations		2,246	2,246	2,246	2,246	2,246	2,246
Controls		Y	Y	Y	Y	Y	Y

Population Data from LandScan, US Department of Energy
Independent variable 'Log (0.01+Distance to Line)' is the nearest geo-distance between the sub-district and the closest straight-line connecting Mumbai and Kolkata - two historical cities that have not had direct transit networks connecting them.
Dependent variables is Log(Population Density)
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table 3.5: Elasticity For non-GQ Route: Mumbai to Kolkata

				Log (Light Density)			
Year		**1992**	**1996**	**2000**	**2004**	**2008**	**2012**
Log(Distance)		**-0.248**	**-0.147**	**-0.0735**	**-0.0822**	**-0.0273**	**-0.0520**
SE clusters:							
	Sub-district	(0.0463)	(0.0351)	(0.0307)	(0.0300)	(0.0293)	(0.0294)
	District	(0.0805)	(0.0520)	(0.0436)	(0.0425)	(0.0405)	(0.0404)
	State	(0.179)	(0.118)	(0.0964)	(0.0845)	(0.0780)	(0.0743)
Observations		2,253	2,253	2,253	2,253	2,253	2,253
Controls		Y	Y	Y	Y	Y	Y
GDP-distance elasticity		0.0744	0.0441	0.02205	0.02466	0.00819	0.0156

*Estimates of GDP-distance elasticity rely on elasticity of GDP-lights being 0.3 for low-middle income countries. Therefore to find elasticity of GDP-distance, multiply the coefficient with 0.3.
Level of observation - sub-district.
Independent variable 'Log (0.01+Distance to Line)' is the nearest geo-distance between the sub-district and the closest straight-line connecting Mumbai and Kolkata - two historical cities that have not had direct transit networks connecting them.
Dependent variables is Log(0.01 + Light density). 'Lights per area' normalizes the sum by the surface area of the district.
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table 3.6: The impact of distance changing over time

	Sum of Lights	Mean Lights	Light Density	P(Majority Lights>0)
Distance to Line	-3.257	-1.796	-3.389	-0.275
SE Level of Clusters:				
Sub-district	(0.357)	(0.184)	(0.386)	(0.0497)
District	(0.539)	(0.302)	(0.561)	(0.0893)
State	(1.088)	(0.643)	(1.187)	(0.155)
Distance*Construction Period	0.867	0.35	0.955	-0.0117
SE Level of Clusters:				
Sub-district	(0.136)	(0.0521)	(0.162)	(0.0259)
District	(0.161)	(0.0872)	(0.202)	(0.0445)
State	(0.351)	(0.232)	(0.397)	(0.117)
Distance*Post Period	1.232	0.526	1.388	0.0938
SE Level of Clusters:				
Sub-district	(0.173)	(0.0733)	(0.205)	(0.0306)
District	(0.206)	(0.120)	(0.247)	(0.0512)
State	(0.437)	(0.267)	(0.487)	(0.127)
R-Squared	0.144	0.285	0.203	0.208
Year Fixed Effects	Y	Y	Y	Y
Controls	Y	Y	Y	Y
Observations	47,313	47,313	47,313	47,313

Level of observation - sub-district. Distances in 1000 kms.
Independent variable ‘Distance to Line’ is the nearest predicted geo-distance between the sub-district and the closest straight-line connecting nodal cities
Dependent variables ‘Sum of Lights’, ‘Mean Lights’ and ’Lights per area’ are of the form Log(0.01 + Lights). ‘Sum of Lights’ is the sum of pixels in a sub-district. ‘Mean Lights’ is the mean light intensity for pixels in a sub-district. ‘Lights per area’ normalizes the sum by the surface area of the district. Dependent variable P(Majority Lights>0) is a 1/0 indicator variable for if the the majority of recorded lights was greater than 0 in that year
Controls include distance to nearest nodal city, coastline, latitude and longitude, and year fixed effects. Results are robust to excluding controls.
Pre-construction period is 1992 to 1999, construction period is 1999 to 2006, and post-construction period is 2007 onwards.

Table 3.7: Beta-convergence and the Solow model

	$Log\frac{Lights(t)}{Lights(t-1)}$	$Log\frac{Lights(t=T)}{Lights(t=0)}$	Growth rate (Lights)
Log Lights (t-1)	-0.108		
	(0.0143)		
Log Lights (t=0)		-0.551	-0.0275
		(0.00998)	(0.00155)
R-squared	0.071	0.583	0.007
Controls	N	Y	Y
		Estimated Rate of Convergence	
β		**0.04**	**0.0392**
Bootstrapped SE		(0.00421)	(0.0128)

Standard errors calculated at the district level (587 districts)
Column 1 tests if unconditional convergence holds in this case using the specification discussed in Sala-i Martin (1996) and Barro and Sala-i Martin (2004)
Column 2 estimates the rate of β-convergence using the methodology discussed in Mankiw et al. (1992)
Column 3 estimates the rate of β-convergence using the OLS methodology discussed in Evans (1997); Higgins et al. (2006)

Table 3.8: 2001 and 2011 Census: population, workers, and literacy

Log Population (2011 Census)	Rural	Total	Female	Male	SC	ST
Log(Distance)	-0.239	-0.218	-0.216	-0.220	-0.226	0.285
	(0.0394)	(0.0402)	(0.0401)	(0.0403)	(0.0474)	(0.0737)
Constant	16.85	17.86	17.16	17.17	13.91	7.284
	(0.802)	(0.798)	(0.799)	(0.797)	(0.966)	(1.153)
Observations	5,259	5,290	5,290	5,290	5,032	5,149
R-squared	0.211	0.205	0.203	0.207	0.064	0.025

Log Population (2001 Census)	Rural	Total	Female	Male	SC	ST
Log(Distance)	-0.213	-0.201	-0.200	-0.203	-0.351	0.300
	(0.0390)	(0.0397)	(0.0398)	(0.0397)	(0.0619)	(0.0763)
Constant	16.65	17.56	16.88	16.86	17.76	6.358
	(0.800)	(0.799)	(0.802)	(0.797)	(1.433)	(1.232)
Observations	5,269	5,293	5,293	5,293	5,112	5,050
R-squared	0.195	0.197	0.196	0.198	0.134	0.027

Census Year	2001 Log(cultivators) per capita	2011 Log(cultivators) per capita	2001 Log(Ag Laborers) per capita	2001 Log(workers in HH ind. per capita)	2001 Literacy Rate	2001 Gender Gap in literacy
Log(Distance)	0.0612	0.0868	-0.149	-0.0698	-0.111	-1.478
	(0.0227)	(0.0222)	(0.0336)	(0.0228)	(0.700)	(0.283)
Constant	3.326	2.323	1.207*	-3.391	125.3	18.93
	(0.455)	(0.544)	(0.632)	(0.360)	(8.507)	(4.073)
Observations	5,264	5,286	5,278	5,284	4,809	4,809
R-squared	0.073	0.035	0.268	0.081	0.090	0.119

Level of observation - Census Teshils (also called Taluks, Mandals and Wards depending on the region)
Standard errors calculated at the district level (587 districts)
SC are known as Scheduled Castes, and STs are Scheduled Tribes - which the two most economically and socially disadvantaged sections.
Gender Gap is defined as the male-literacy rate minus the female-literacy rate.

Table 3.9: Neighbors closer to the road vs. farther away from the road

	Contemporaneous effect	Lagged effect	Changes*
Log(light density) of neighbors closer to road	0.490		
	(0.0302)		
Lagged		0.370	
		(0.0269)	
Changes			0.332
			(0.0427)
Log(light density) of neighbors further from road	0.351		
	(0.0356)		
Lagged		0.237	
		(0.0333)	
Changes			0.265
			(0.0287)
Constant	1.648	4.138	0.0280
	(0.304)	(0.387)	(0.00290)
Observations	45,045	42,900	42,900
R-squared	0.341	0.194	0.106
Fixed Effect Units	2,145	2,145	2,145

Fized effects regressions - Level of observation - Sub-district-year
Standard errors calculated at the district level (587 districts)
This table tests whether neighbors closer to the highway have larger impacts than regions away from the highway.
* The 'Changes' version of the equation estimates $\Delta Log(lights)_{t,k} = \beta \Delta Log(lights)_{t,k-1} + \gamma \Delta Log(lights)_{t,k+1}$

Table 3.10: Spillovers from Neighbors and the effect of distance

Panel A **Log(Light density)**	**1992**	**1997**	**2003**	**2007**	**2012**
Log(light density) of neighbors closer to road χ	**0.868**	**0.827**	**0.773**	**0.843**	**0.802**
	(0.0491)	(0.0656)	(0.0637)	(0.0560)	(0.0762)
Log(distance) μ_t	**-0.104**	**-0.101**	**-0.0853**	**-0.0652**	**-0.064**
	(0.0376)	(0.0312)	(0.0254)	(0.0235)	(0.0225)
Observations	2,219	2,219	2,219	2,219	2,219
Controls	Y	Y	Y	Y	Y
R-squared	0.493	0.470	0.462	0.502	0.440
Distance-spillover parameter α_t	**-0.495**	**-0.429**	**-0.314**	**-0.290**	**-0.254**
Bootstrapped SE	(0.166)	(0.156)	(0.0999)	(0.117)	(0.112)
Panel B **Log(Light density)**	**1992**	**1997**	**2003**	**2007**	**2012**
Log(distance) α_t	**-0.497**	**-0.404**	**-0.294**	**-0.227**	**-0.212**
	(0.0591)	(0.0488)	(0.0388)	(0.0382)	(0.0364)
Observations	2,253	2,253	2,253	2,253	2,253
Controls	Y	Y	Y	Y	Y
R-squared	0.135	0.174	0.203	0.180	0.147
Panel C **Reduced Form**	**1992**	**1997**	**2003**	**2007**	**2012**
GDP-Distance Elasticity ignoring spillovers	-0.149	-0.121	-0.0789	-0.0681	-0.0635
Bootstrapped SE	(0.0164)	(0.0132)	(0.0123)	(0.0114)	(0.0105)
Overall Income Effects Incorporating Spillovers	**-0.149**	**-0.177**	**-0.2191**	**-0.2299**	**-0.2345**
IV 2SLS	**1992**	**1997**	**2003**	**2007**	**2012**
GDP-Distance Elasticity ignoring spillovers	-0.22	-0.179	-0.117	-0.101	-0.0938
Bootstrapped SE	(0.0237)	(0.0229)	(0.0158)	(0.0154)	(0.0140)
Overall Income Effects Incorporating Spillovers	**-0.22**	**-0.261**	**-0.323**	**-0.339**	**-0.3462**

Level of observation - Sub-district

Standard errors calculated at the district level (587 districts)

This table tests the model where light density y_{td} depends on light-density of the neighbors closer to the highway $y_{t,d-1}$ and distance to the highway D, in the following way: $Log\ y_{td} = \chi Log\ y_{t,d-1} + \mu_t Log\ D$. This relationship is estimated in Panel A. Furthermore, we can recursively solve, to show that $Log\ y_{td} = (\sum_{j=0}^{d} \chi^j)\mu_t Log\ D$, which is the parameter evaluated as the "Distance-spillover parameter" between the two panels for the average number of degrees-of-separation (i.e. $d = 6$). Panel B then tests if this parameter is equal to the parameter obtained by regressing $Log\ y_{td} = \alpha_t Log\ D$.

Panel C re-estimates the GDP-distance elasticities incorporating these spillovers and compares them to the elasticities where the spillovers were ignored.

3.8 Additional Tables and Figures

Table 3.11: OLS relationship between night-time lights and distance to GQ highway: Different Measures of Lights

Year: 2012	Sum of Lights	Mean Lights	Lights per area	P(Majority Lights>0)
Distance to GQ Highway	-2.016	-1.287	-1.929	-0.0536
SE clusters:				
Sub-district	(0.302)	(0.176)	(0.316)	(0.0614)
District	(0.469)	(0.289)	(0.457)	(0.107)
State	(0.858)	(0.577)	(0.849)	(0.190)
R-squared	0.084	0.220	0.156	0.201

Year: 1992	Sum of Lights	Mean Lights	Lights per area	P(Majority Lights>0)
Distance to GQ Highway	-3.680	-1.654	-3.913	-0.217
Standard Errors				
Level of Clustering:				
Sub-district	(0.434)	(0.398)	(0.492)	(0.0582)
District	(0.701)	(0.731)	(0.784)	(0.0958)
State	(1.592)	(1.213)	(1.773)	(0.178)
R-squared	0.108	0.163	0.144	0.051
Controls	Y	Y	Y	Y
Observations	2,253	2,253	2,253	2,253

Level of observation - sub-district. Distances in 1000 kms.

Independent variable ‘Distance to GQ Highway’ is the nearest geo-distance between the sub-district and the closest Golden Quadrilateral highway

Dependent variables ‘Sum of Lights’, ‘Mean Lights’ and ’Lights per area’ are of the form Log(0.01 + Lights). ‘Sum of Lights’ is the sum of pixels in a sub-district. ‘Mean Lights’ is the mean light intensity for pixels in a sub-district. ‘Lights per area’ normalizes the sum by the surface area of the district. Dependent variable P(Majority Lights>0) is a 1/0 indicator variable for if the sub-district has any visible lights emitted in that year

Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table 3.12: Reduced-form relationship between Lights and straight-lines: Different Measures of Lights

Year: 2012	Sum of Lights	Mean Lights	Lights per area	P(Majority Lights>0)
Distance to Line	-1.919	-1.210	-1.876	-0.133
SE clusters:				
Sub-district	(0.305)	(0.171)	(0.316)	(0.0568)
District	(0.464)	(0.268)	(0.432)	(0.0942)
State	(0.778)	(0.517)	(0.782)	(0.136)
R-squared	0.085	0.220	0.157	0.202

Year: 1992	Sum of Lights	Mean Lights	Lights per area	P(Majority Lights>0)
Distance to Line	-3.805	-1.861	-4.060	-0.224
SE clusters:				
Sub-district	(0.420)	(0.216)	(0.479)	(0.0548)
District	(0.668)	(0.377)	(0.749)	(0.0923)
State	(1.402)	(0.794)	(1.588)	(0.167)
R-squared	0.117	0.173	0.152	0.052
Controls	Y	Y	Y	Y
Observations	2,253	2,253	2,253	2,253

Level of observation - sub-district. Distances in 1000 kms.

Independent variable 'Distance to Line' is the nearest geo-distance between the sub-district and the closest straight-line connecting nodal cities

Dependent variables 'Sum of Lights', 'Mean Lights' and 'Lights per area' are of the form Log(0.01 + Lights). 'Sum of Lights' is the sum of pixels in a sub-district. 'Mean Lights' is the mean light intensity for pixels in a sub-district. 'Lights per area' normalizes the sum by the surface area of the district. Dependent variable P(Majority Lights>0) is a 1/0 indicator variable for if the sub-district has any visible lights emitted in that year

Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table 3.13: Two-staged least squares relationship between lights and distance to GQ highways: Different Measures of Lights

Year: 2012	Sum of Lights	Mean Lights	Lights per area	P(Majority Lights>0)
Distance to GQ Highway	-2.368	-1.494	-1.929	-0.164
SE clusters:				
Sub-district	(0.377)	(0.211)	(0.391)	(0.0699)
District	(0.572)	(0.331)	(0.535)	(0.117)
State	(0.951)	(0.638)	(0.965)	(0.167)
Pagan-Hall Het Test	72.25	129.2	68.56	183.6
p-value of Pagan-Hall	0	0	0	0
Year: 1992	Sum of Lights	Mean Lights	Lights per area	P(Majority Lights>0)
Distance to GQ Highway	-4.697	-2.297	-5.011	-0.277
SE clusters:				
Sub-district	(0.521)	(0.266)	(0.593)	(0.0675)
District	(0.830)	(0.468)	(0.933)	(0.114)
State	(1.702)	(0.969)	(1.940)	(0.202)
Pagan-Hall Het Test	215.6	164.4	200.0	161.9
p-value of Pagan-Hall	0	0	0	0
Controls	Y	Y	Y	Y
Observations	2,253	2,253	2,253	2,253

Level of Clustering:	F-Stat	Prob>F	Hansen J	Partial R-sq
Sub-district	6796	0	0	0.738
District	1569	0	0	0.738
State	474.7	0	0	0.738

Level of observation - sub-district. Distances in 1000 kms.
Independent variable 'Distance to GQ Highway' is the nearest predicted geo-distance between the sub-district and the closest GQ highway, predicted by the distance to the closest straight-line connecting nodal cities
Dependent variables 'Sum of Lights', 'Mean Lights' and 'Lights per area' are of the form Log(0.01 + Lights). 'Sum of Lights' is the sum of pixels in a sub-district. 'Mean Lights' is the mean light intensity for pixels in a sub-district. 'Lights per area' normalizes the sum by the surface area of the district. Dependent variable P(Majority Lights>0) is a 1/0 indicator variable for if the sub-district has any visible lights emitted in that year
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table 3.14: Elasticity between Light-Density and State Domestic Product for 32 States

Per capita	GDP at	2005	prices					
Log(per cap GDP)	2005	2006	2007	2008	2009	2010	2011	2012
Log(light density)	0.19	0.198	0.19	0.191	0.183	0.198	0.192	0.186
	(0.0508)	(0.0503)	(0.0510)	(0.0545)	(0.0557)	(0.0617)	(0.0600)	(0.0589)
Constant	8.422	8.409	8.563	8.541	8.72	8.548	8.695	8.798
	(0.526)	(0.523)	(0.529)	(0.584)	(0.594)	(0.684)	(0.655)	(0.648)
Observations	32	32	32	32	32	32	32	32
R-squared	0.318	0.340	0.316	0.291	0.265	0.255	0.254	0.249

Per capita	NDP at	2005	prices					
Log(per cap NDP)	2005	2006	2007	2008	2009	2010	2011	2012
Log(light density)	0.191	0.199	0.191	0.192	0.186	0.202	0.196	0.191
	(0.0514)	(0.0508)	(0.0517)	(0.0549)	(0.0560)	(0.0626)	(0.0613)	(0.0606)
Constant	8.294	8.275	8.428	8.406	8.558	8.375	8.519	8.612
	(0.533)	(0.528)	(0.536)	(0.589)	(0.597)	(0.694)	(0.670)	(0.667)
Observations	32	32	32	32	32	32	32	32
R-squared	0.315	0.339	0.313	0.290	0.269	0.257	0.255	0.250

Regressions of Log(0.01+light density) on Log(per capita domestic product) at the state level.
State Domestic Product Sources: Reserve Bank of India
GDP indicates Gross Domestic Product of the State; and NDP is the Net Domestic Product

Table 3.15: Elasticity between Light Density and Per Capita District Domestic Product

Per capita NDP	at current	prices						
Log(per cap NDP)	2005	2006	2007	2008	2009	2010	2011	2012
Log(light density)	0.332	0.357	0.347	0.371	0.354	0.371	0.394	0.388
	(0.0236)	(0.0259)	(0.0264)	(0.0291)	(0.0249)	(0.0324)	(0.0296)	(0.0462)
Constant	6.213	5.988	6.16	5.82	6.103	5.762	5.661	5.75
	(0.241)	(0.268)	(0.274)	(0.312)	(0.264)	(0.361)	(0.323)	(0.518)
Observations	209	222	222	222	222	222	190	96
R-squared	0.488	0.463	0.439	0.426	0.479	0.373	0.485	0.428

Per capita NDP	at 2004-5	prices						
Log(per cap NDP)	2005	2006	2007	2008	2009	2010	2011	2012
Log(light density)	0.363	0.354	0.344	0.365	0.344	0.362	0.385	0.357
	(0.0230)	(0.0243)	(0.0261)	(0.0290)	(0.0244)	(0.0310)	(0.0287)	(0.0457)
Constant	5.915	6.071	6.302	6.113	6.516	6.238	6.223	6.591
	(0.238)	(0.252)	(0.271)	(0.312)	(0.260)	(0.347)	(0.315)	(0.511)
Observations	249	249	222	249	249	249	217	97
R-squared	0.502	0.463	0.440	0.391	0.446	0.356	0.456	0.391

Regressions of Log(0.01+light density) on Log(per capita domestic product) at the district level.
District Domestic Product Sources: Department of Statistics and Programme Implementation, Government of West Bengal; Planning Commission; Directorate of Economics and Statistics Government of Uttar Pradesh; Department of Economics and Statistics Government of Tamil Nadu; Directorate of Economics and Statistics Government of Rajasthan; Department of Planning Government of Punjab; Planning and Coordination Government of Odisha; Directorate of Economics and Statistics Government of Maharashtra; Directorate of Economics and Statistics Government of Kerala; Planning Programme Monitoring and Statistics Department Government of Karnataka; Directorate of Economics and Statistics Government of Bihar; Directorate of Economics and Statistics Government of Assam; Andhra Pradesh State Portal

Table 3.16: OLS relationship between lights and rail-lines: Different Measures of Lights

Year: 2012	Sum of Lights	Mean Lights	Lights per area	P(Majority Lights>0)
Distance to Railways	-9.321	-6.253	-8.82	-1.069
SE clusters:				
Sub-district	(5.341)	(2.885)	(4.656)	(0.400)
District	(5.536)	(2.954)	(4.816)	(0.403)
State	(7.584)	(3.997)	(6.552)	(0.530)
R-squared	0.096	0.232	0.164	0.207
Year: 1992	Sum of Lights	Mean Lights	Lights per area	P(Majority Lights>0)
Distance to Railways	-14.32	-7.884	-15.37	-0.862
SE clusters:				
Sub-district	(8.247)	(3.813)	(8.535)	(0.404)
District	(8.569)	(3.882)	(8.827)	(0.416)
State	(11.86)	(5.328)	(12.20)	(0.542)
R-squared	0.108	0.175	0.145	0.051
Controls	Y	Y	Y	Y
Observations	2,253	2,253	2,253	2,253

Level of observation - sub-district. Distances in 1000 kms.
Independent variable 'Distance to Railroad' is the nearest geo-distance between the sub-district and the closest Railway line
Dependent variables 'Sum of Lights', 'Mean Lights' and 'Lights per area' are of the form Log(0.01 + Lights). 'Sum of Lights' is the sum of pixels in a sub-district. 'Mean Lights' is the mean light intensity for pixels in a sub-district. 'Lights per area' normalizes the sum by the surface area of the district. Dependent variable P(Majority Lights>0) is a 1/0 indicator variable for if the sub-district has any visible lights emitted in that year
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table 3.17: Two staged least squares relationship between lights and distance to railways: Different Measures of Lights

Year: 2012	Sum of Lights	Mean Lights	Lights per area	P(Majority Lights>0)
Distance to Railways	-34.86	-21.99	-34.09	-2.412
SE clusters:				
Sub-district	(14.51)	(8.781)	(14.32)	(1.271)
District	(15.63)	(9.903)	(15.91)	(1.856)
State	(22.23)	(14.34)	(22.31)	(2.669)
Pagan-Hall Het Test	24.15	18.87	23.54	52.75
p-value of Pagan-Hall	0.000203	0.00203	0.000266	0
Year: 1992	Sum of Lights	Mean Lights	Lights per area	P(Majority Lights>0)
Distance to Railways	-69.14	-33.82	-73.77	-4.076
SE clusters:				
Sub-district	(28.20)	(13.48)	(30.25)	(1.831)
District	(30.19)	(15.06)	(33.09)	(2.293)
State	(42.09)	(21.36)	(45.97)	(3.562)
Pagan-Hall Het Test	42.88	25.49	38.05	12.99
p-value of Pagan-Hall	0	0.000112	0	0.0235
Controls	Y	Y	Y	Y
Observations	2,253	2,253	2,253	2,253
Level of Clustering:	F-Stat	Prob > F	Hansen J	Partial R-sq
Sub-district	6.570	0.0104	0	0.0516
District	5.348	0.0211	0	0.0516
State	2.707	0.109	0	0.0516

Level of observation - sub-district. Distances in 1000 kms.
Independent variable ‘Distance to Railroad’ is the nearest predicted geo-distance between the sub-district and the closest rail-line, predicted by the distance to the closest straight-line connecting nodal cities
Dependent variables ‘Sum of Lights’, ‘Mean Lights’ and ’Lights per area’ are of the form Log(0.01 + Lights). ‘Sum of Lights’ is the sum of pixels in a sub-district. ‘Mean Lights’ is the mean light intensity for pixels in a sub-district. ‘Lights per area’ normalizes the sum by the surface area of the district. Dependent variable P(Majority Lights>0) is a 1/0 indicator variable for if the the majority of recorded lights was greater than 0 in that year
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

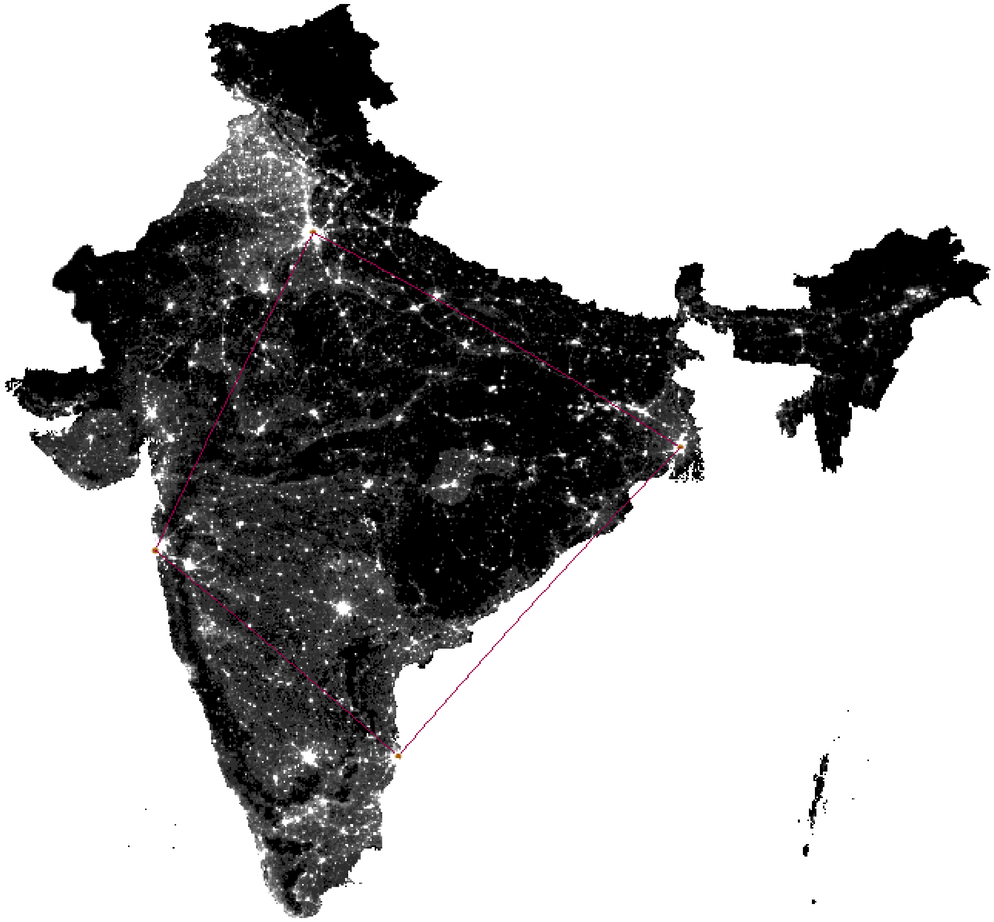


Figure 3.8.1: Night-time lights and straight-lines between four nodes

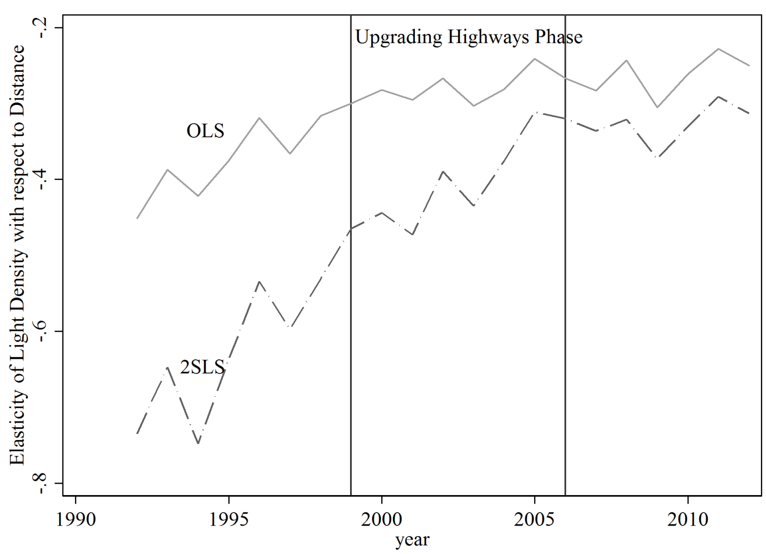
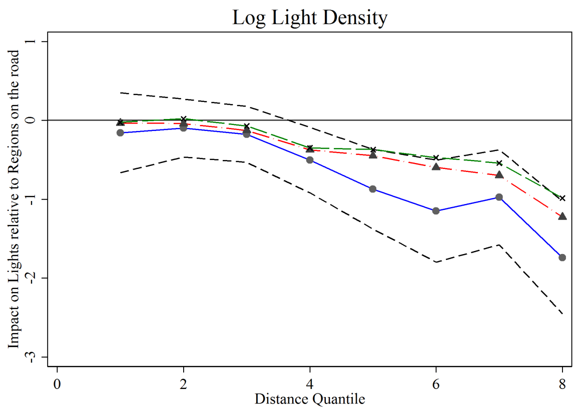


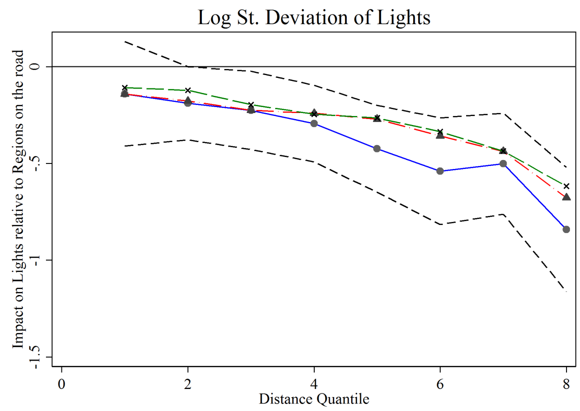
Figure 3.8.2: Change in Elasticities Over Time

Elasticities calculated by running a log-log relationship between lights and distance. Vertical lines represent the phases of construction - 1999 is when the highways started being built. There were delays till 2001 when most work started, and 2006 is when most work was completed.

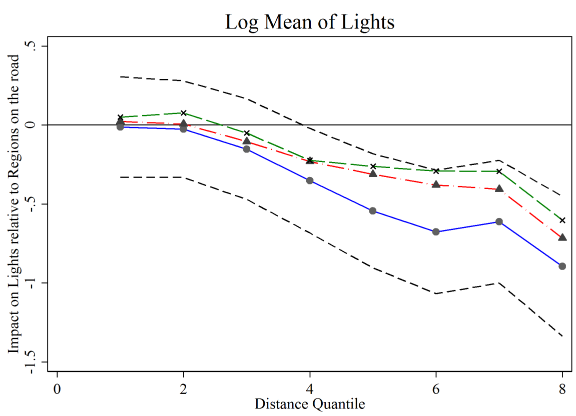
Figure 3.8.3: Impact of distance on lights: Different Measures of Light Density



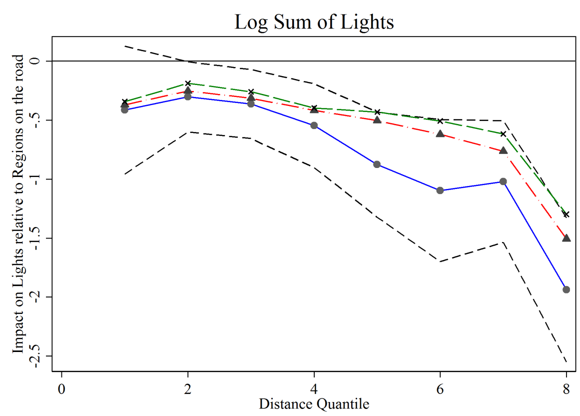
(a) Light Density (per sq miles)



(b) Standard Deviation of Lights



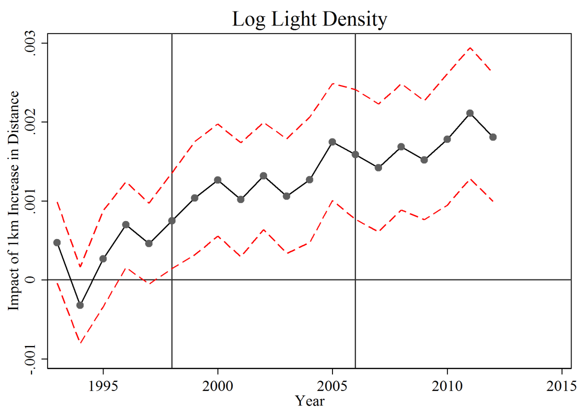
(c) Mean Lights



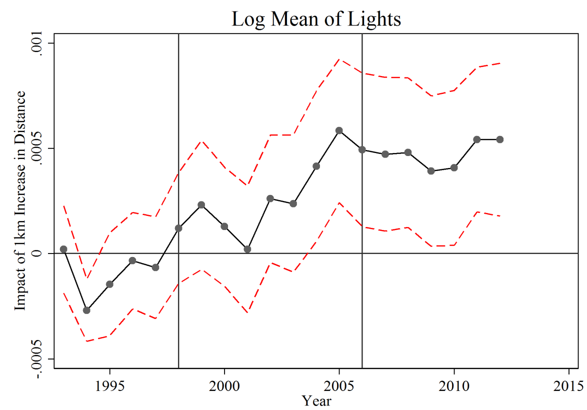
(d) Sum of lights

The graphs show the impact of distance on night-time lights relative to sub-districts that have any portion of the sub-district touching the straight-lines. The blue lines are for the pre-construction period, the orange lines for the construction period and the green lines for the post-construction period. The standard error bands are for the pre-construction (blue) lines and clustered at the district level. The 'Distance' axis consists of 8 quantiles of equal size. The distance quantile cutoffs are roughly as follows: 0 to 40kms, 40 to 90 kms, 90 to 135 kms, 135 to 200kms, 200 to 260kms, 260 to 340 kms, 340 to 440 kms, and above 440kms.

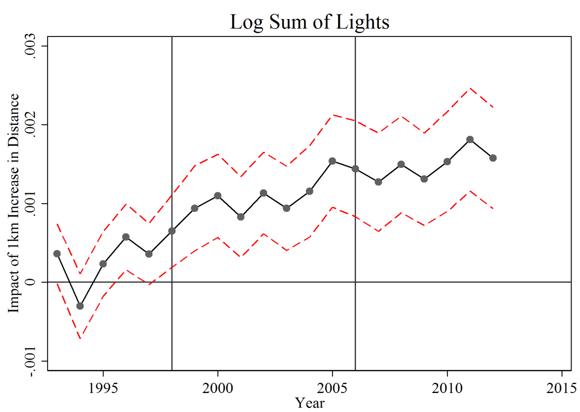
Figure 3.8.4: How the Impact of Distance on Light-Density changes over time (relative to 1992): Different Measures of Light Density



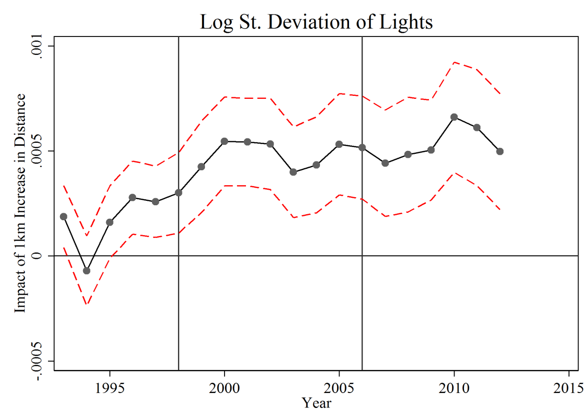
(a) Light-Density:1992 coefficient -0.00406



(b) Mean-lights. 1992 cofficient is -0.00186



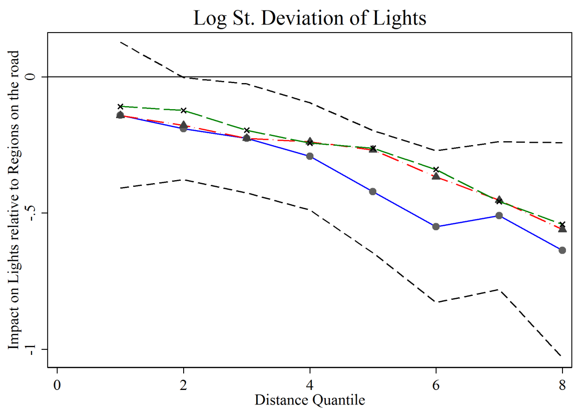
(c) Sum of lights: 1992 coefficient is -0.0038



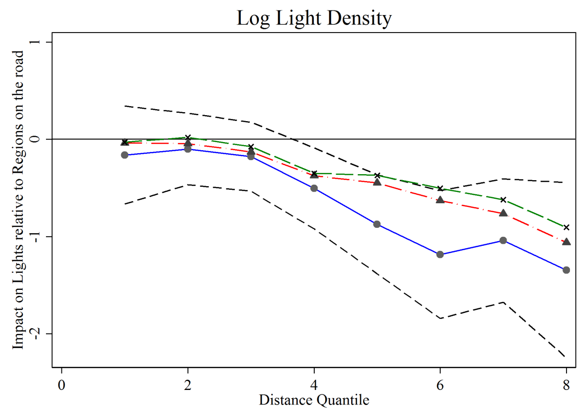
(d) Standard Deviation of Lights -0.00174

Coefficients of change in impact relative to 1992. Standard errors calculated at the district level. Vertical lines represent the phases of construction - 1999 is when the highways started being built. There were delays till 2001 when most work started, and 2006 is when most work was completed. To interpret the graph: the mean impact of a 1km increase in distance from the highway was a 0.00406 fall in light-density, and this impact has been dissipating over time. By 2012 the impact of a 1km increase in distance from the highway had become -0.00406+0.00205, or about -0.00201.

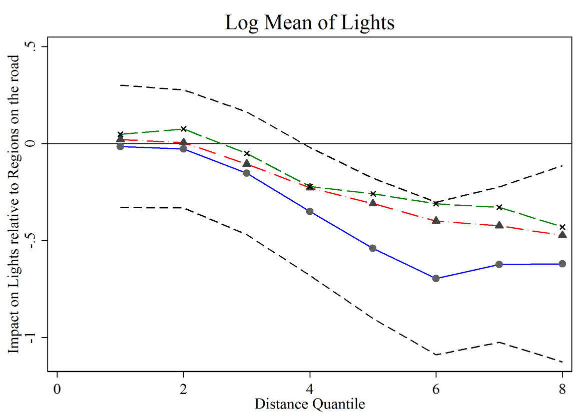
Figure 3.8.5: Robustness Checks: Impact of distance on lights excluding outlying states



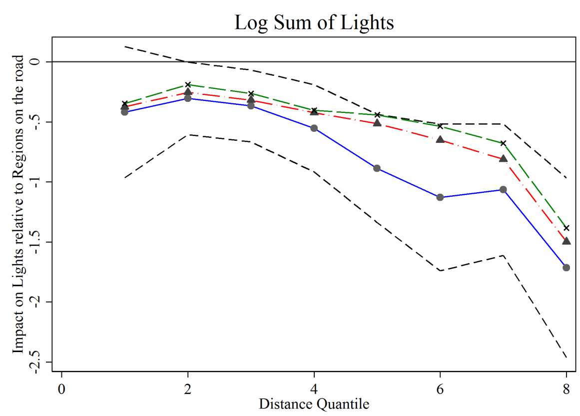
(a) Standard Deviation of Lights



(b) Light Density (per sq miles)



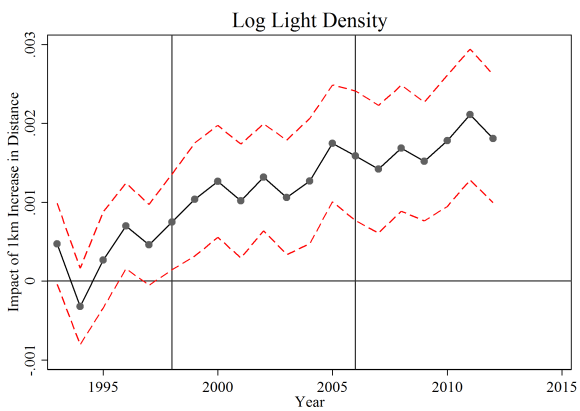
(c) Mean Lights



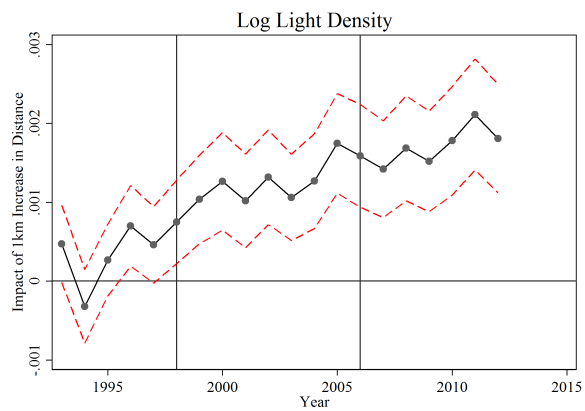
(d) Sum of lights

The excluded states in this robustness check include: Jammu and Kashmir, Manipur, Meghalaya, Tripura, Nagaland, Sikkim, Assam, Arunachal Pradesh, Mizoram, Andaman and Nicobar Islands and Lakshwadeep. The graphs show the impact of distance on night-time lights relative to sub-districts that have any portion of the sub-district touching the straight-lines. The blue lines are for the pre-construction period, the orange lines for the construction period and the green lines for the post-construction period. The standard error bands are for the pre-construction (blue) lines and clustered at the district level. The 'Distance' axis consists of 8 quantiles of equal size. The distance quantile cutoffs are roughly as follows: 0 to 40kms, 40 to 90 kms, 90 to 135 kms, 135 to 200kms, 200 to 260kms, 260 to 340 kms, 340 to 440 kms, and above 440kms.

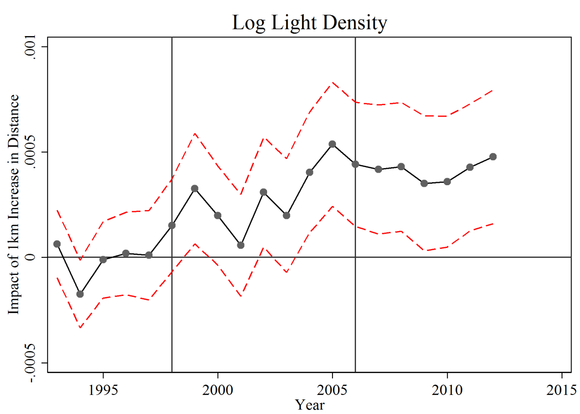
Figure 3.8.6: Robustness Checks: Different samples and specifications for: How the Impact of Distance on Light-Density changes over time



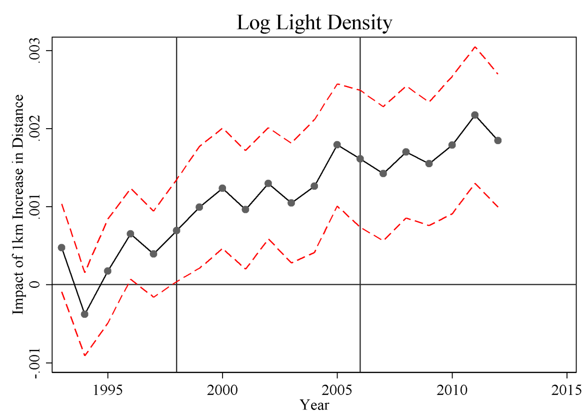
(a) District-level clusters



(b) Sub-district clusters



(c) Dropping 0s - Log(lights/area)



(d) Dropping extremely high light-density

Coefficients of change in impact relative to 1992. The impact of a 1km increase in distance in 1992 was a 0.00406. Vertical lines represent the phases of construction - 1999 is when the highways started being built. There were delays till 2001 when most work started, and 2006 is when most work was completed. The ‘Dropping extremely high light-density’ panel drops all sub-districts if they ever recorded a light pixel equal to the maximum possible value (63). This is 12% of the sample.

3.9 Mobility and Convergence

One factor that may contribute to the dissipation of the effects of these routes may lie with the mobility of firms and people. However, such a model of mobility patterns would have to explain why capital or workers move *away* from the road over time. If the upgradation of highways actually allowed firms and workers to move away from the highway, then that may explain the spread in economic activity to regions further away.

3.9.1 Movement of Capital and Firms

The first is a location choice model for firms and capital. While the long-run positive impacts of the road suggest that firms locate to regions along these routes, somewhat more complex models are needed to explain why firms may move away over time. In such a model, upgrading the highway and therefore better connectivity on the highway system allows firms and households to locate further away from the highway. A firm or household enterprise wishes to be connected to the four nodal cities for purposes of trade and exchange. For locations that are not near the nodal cities, the cost of being connected to the city is the sum of the cost of being connected to the highway $c_1(d)$ and the cost of using the highway to get to the city c_2, where $c_1(d)$ is an increasing function of the distance between the region and the closest highway d.[23] Firms have returns to investment that is drawn from a distribution $R \sim F(.)$ and choose to locate in region d if $R \geq c_1(d) + c_2$. The fraction of firms that locate there are therefore $1 - F(c_1(d) + c_2)$. Since the cost functions are increasing in the distance, this would indicate that more firms will locate closer to the highways. Once the highway is upgraded, this reduces c_2, thereby allowing the same firm to locate at a distance d' further away from the highway. After the NHDP upgrades, firms and households can therefore move into regions that would earlier have been too costly for them to locate in. Note that in such a model, firms are no longer profit maximizing – lowering the credibility of such a model.

The strongest evidence against this model is the result shown in Ghani et al. (2015) who find that when the highways are upgraded, organized manufacturing firms actually move closer to the highway. Similarly, Redding and Turner (2015) discuss how it is likely that firms would move from 'untreated' to 'treated' regions when highways are built.

3.9.2 Movement of People

Another model is that of lowering costs of seasonal migration. Once the highways are upgraded, this allows people from near the highways to migrate to the city at lower costs. If the migration is permanent,

[23] c_2 can also be made to depend on the distance on the highway from the city, but is irrelevant for this analysis

there will be a fall in population and economic activity in the region close to the highway. In the Indian context, migration for work is low Munshi and Rosenzweig (2016), and Table 3.4 shows that at least long-run migration cannot drive the dynamic results. However, a lot of the migration in the Indian context is seasonal in nature, where people work in the cities during the agricultural slack season but on the farms in the peak season. For a city wage W_c that lies between the peak season w_p and slack season w_s wages, after taking into account the cost of migrating c_m, a laborer word work in the city during the agricultural slack season and return to his fields in the peak season. If a worker doesn't migrate in the slack season (he lives in a region where c_m is high), then he would engage in non-agricultural economic activity instead, but if he does migrate (c_m is low) then there is no need for non-agricultural enterprises to exist in these regions since in the slack season workers would easily migrate to the city. A lowering of migration costs by upgrading the highways would then facilitate seasonal migration and thereby reduce any non-agricultural enterprises from setting up in rural areas near the highways.

BIBLIOGRAPHY

Abbott, B., Gallipoli, G., Meghir, C., and Violante, G. L. (2013). Education Policy and Intergenerational Transfers in Equilibrium. *NBER Working Papers 18782*.

Acemoglu, D. (2010). Theory, General Equilibrium, and Political Economy in Development Economics. *Journal of Economic Perspectives*, 24(3):17–32.

Adler, S. (2016). Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development. *Mimeo*.

Afridi, F. (2010). Child Welfare Programs and Child Nutrition: Evidence from a Mandated School Meal Program in India. *Journal of Development Economics*, 92(2):152–165.

Aggarwal, Y. (2000). Monitoring and Evaluation Under DPEP: Measurement of Social Impact. *NIEPA*, New Delhi.

Aguero, J. M. and Bharadwaj, P. (2014). Do the More Educated Know More about Health? Evidence from Schooling and HIV Knowledge in Zimbabwe. *Economic Development and Cultural Change*, 62(3):489–517.

Albrecht, J., van den Berg, G. J., and Vroman, S. (2009). The Aggregate Labor Market Effects of The Swedish Knowledge Lift Program. *Review of Economic Dynamics*, 12(1):129–146.

Anderson, S. (2005). Caste as an Impediment to Trade. *American Economic Journal: Applied Economics*, 3(1):239–63.

Andrabi, T., Das, J., and Khwaja, A. (April 2013). Students Today, Teachers Tomorrow: Do Public Investments Alleviate Constraints on the Provision of Private Education? *Journal of Public Economics*, 100(1):1–14.

Angrist, J. (1995). The Economic Returns to Schooling in the West Bank and Gaza Strip. *The American Economic Review*, 85(5):1065–1087.

Angrist, J., Bettinger, E., and Kremer, M. (2006). Long-Term Consequences of Secondary School Vouchers: Evidence from an Administrative Records in Colombia. *American Economic Review*, 96(3):847–72.

Arcidiacono, P., Aucejo, E., Coate, P., and Hotz, V. J. (2014). Affirmative Action and University Fit: Evidence from Proposition 209. *IZA Journal of Labor Economics*, 3(7).

Arcidiacono, P., Aucejo, E., Fang, H., and Spenner, K. (2011). Does Affirmative Action Lead to Mismatch? A New Test and Evidence. *Quantitative Economics*, 2(3):303–333.

Asher, S. and Novosad, P. (2016). Market Access and Structural Transformation: Evidence from Rural Roads in India. *Mimeo*.

Ashraf, N., Bau, N., Nunn, N., and Voena, A. (2015). Bride Price and the Returns to Education for Women. *Mimeo*, Harvard.

Assuncao, J. and Ferman, B. (2015). Does Affirmative Action Enhance or Undercut Investment Incentives? Evidence from Quotas in Brazilian Public Universities. *Mimeo*.

Asturias, J., Garcia-Santana, M., and Ramos, R. (2015). Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India. *Mimeo*.

Atack, J., Haines, M., and Margo, R. (2008). Railroads and the Rise of the Factory: Evidence for the United States, 1850-70. *NBER Working Paper 14410*.

Ayyar, R. (2008). Country Agency Relationship in Development Cooperation: An Indian Experience. *Global Monitoring Report*, UNESCO.

Bagde, S., Epple, D., and Taylor, L. (2016). Does Affirmative Action Work? Caste, Gender, College Quality, and Academic Success in India. *American Economic Review*, 106(6):1495–1521.

Banerjee, A., Banerji, R., Duflo, E., Glennerster, R., and Khemani, S. (2010). Pitfalls of Participatory Programs: Evidence From a Randomized Evaluation in Education in India. *American Economic Journal: Economic Policy*, 2(1):1–30.

Banerjee, A., Cole, S., Duflo, E., and Linden, L. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *Quarterly Journal of Economics*, 22(3):1235–1264.

Banerjee, A. and Duflo, E. (2005). Growth Theory Through the Lens of Development Economics. *Handbook of Economic Growth*, 1A:473–552.

Banerjee, A., Duflo, E., and Qian, N. (2012). On the Road: Access to Transportation Infrastructure and Economic Growth in China. *NBER Working Paper 17897*.

Banerjee, A., Iyer, L., and Somanathan, R. (2008). Public Action for Public Goods. *Handbook of Development Economics*, 4:3117–3154.

Barro, R. J. and Sala-i Martin, X. (1992). Convergence. *Journal of Political Economy*, (100):223–51.

Barro, R. J. and Sala-i Martin, X. (2004). Economic Growth. *The MIT Press*, 1.

Bartalotti, O. and Brummet, Q. (2015). Estimation and Inference in Regression Discontinuity Designs with Clustered Sampling. *Mimeo*, US Census Bureau.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M., and Zhang, Q. (2014). Roads, Railroads and Decentralization of Chinese Cities. *Mimeo*.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M., and Zhang, Q. (2015). Transport Infrastructure, Urban Growth and Market Access in China. *Mimeo*.

Becker, G. S. (1967). Human Capital and the Personal Distribution of Income: An Analytical Approach. *Woytinsky Lecture, Ann Arbor: University of Michigan, Institute of Public Administration*.

Bedard, K. (2001). Human Capital Versus Signaling Models: University Access and High School Drop-outs. *Journal of Political Economy*, 190(4):749–775.

Behrman, J. (1999). Schooling in Asia: Selected Microevidence on Determinants, Effects, and Policy Implications. *Journal of Asian Economics*, 10:147–194.

Behrman, J. R., Birdsall, N., and Kaplan, R. (1996). The Quality of Schooling and Labor Market Outcomes. *Nancy Birdsall and Richard H. Sabot (eds.) Opportunity Foregone: Education in Brazil. Johns Hopkins University Press.*, 8.

Bertrand, M., Hanna, R., and Mullainathan, S. (2010). Affirmative Action: Evidence from College Admissions in India. *Journal of Public Economics*, 94(1-2):16–29.

Bianchi, N. (2016). The Effects of Educational Expansions: Evidence from a Large Enrollment Increase in STEM Majors. *Mimeo*, Northwestern.

Birdsall, N. (1982). Child Schooling and the Measurement of Living Standards. *World Bank Living Standards Measurement Study Working Paper*, (14).

Birdsall, N. (1985). Public Inputs and Child Schooling in Brazil. *Journal of Development Economics*, 18(1):67–86.

Bobonis, G. and Finan, F. (2009). Neighborhood Peer Effects in Secondary School Enrollment Decisions. *The Review of Economics and Statistics*, 91(4):695–716.

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., and Sandefur, J. (2013a). Scaling-up What Works: Experimental Evidence on External Validity in Kenyan Education. *CSAE Working Paper Series 2013-04, Centre for the Study of African Economies, University of Oxford*.

Bold, T., Kimenyi, M., and Sandefur, J. (2013b). Public and Private Provision of Education in Kenya. *Journal of African Economies*, 22(2):39–56.

Borkum, E., He, F., and Linden, L. (2010). School Libraries and Language Skills in Indian Primary Schools: A Randomized Evaluation of the Akshara Library Program. *Mimeo*, Columbia.

Bound, J. and Johnson, G. (1992). Changes in the Structure of Wages During the 1980s: An Evaluation of Alternative Explanations. *American Economic Review*, 82(3):371–392.

Breierova, L. and Duflo, E. (2003). The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter less than Mothers? *NBER Working Papers 10513*.

Bryan, G., Chowdhury, S., and Mobarak, A. M. (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica*, 82(5):1671–1748.

Calonico, S., Cattaneo, M., and Titiunik, R. (2014a). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*, 14(4):909–946.

Calonico, S., Cattaneo, M., and Titiunik, R. (2014b). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326.

Cameron, A. C., Gelbach, J. B., and Miller, D. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, 90(3):414–427.

Cannonier, C. and Mocan, N. (2012). Empowering Women Through Education: Evidence from Sierra Leone,. *NBER Working Papers 18016*.

Card, D. (1999). TheCausal Effect of Education on Earnings. *Handbook of Labor Economics*, 5:1801–1863.

Card, D. and Krueger, A. (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy*, 100:1–40.

Card, D. and Krueger, A. B. (2005). Would the Elimination of Affirmative Action Affect Highly Qualified Minority Applicants? Evidence from California and Texas. *Industrial and Labor Relations Review*, 58(3).

Card, D. and Lemieux, T. (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics*, 116(2):705–746.

Card, D. and Payne, A. A. (2002). School finance reform, the distribution of school spending, and the distribution of student test scores. *Journal of Public Economics*, 83:49–82.

Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating Marginal Returns to Education. *American Economic Review*, 101(6):2754–2781.

Cattaneo, M., Jansson, M., and Ma, X. (2015). Simple Local Regression Distribution Estimators with an Application to Manipulation Testing. *Mimeo*, Michigan.

Chandra, A. and Thompson, E. (2000). Does Public Infrastructure Affect Economic Activity? Evidence from the Rural Interstate Highway System. *Regional Science and Urban Economics*, 30(4):457–90.

Chin, A. (2005). Can Redistributing Teachers Across Schools Raise Educational Attainment? Evidence from Operation Blackboard in India. *Journal of Development Economics*, 78:384–405.

Coate, S. and Loury, G. C. (1993). Will Affirmative Action Eliminate Negative Stereotypes? *American Economic Review*, 83(5):1220–1240.

Colclough, C. and De, A. (2010). The Impact of Aid on Education Policy in India. *Research Consortium on Educational Outcomes and Poverty*, 30(5):497–507.

Crepon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. (2013). Do Labor Market Policies Have Displacement Effects: Evidence from a Clustered Randomized Experiment. *Quarterly Journal of Economics*, 128(2):531–580.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., and Sundararaman, V. (2013a). School Inputs, Household Substitution, and Test Scores. *American Economic Journal: Applied Economics*, 5(2):29–57.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., and Sundararaman, V. (2013b). School Inputs, Household Substitution, and Test Scores. *American Economic Journal: Applied Economics*, 5(2):29–57.

Das, J., Holla, A., Kremer, M., Mohpal, A., and Muralidharan, K. (2013c). The Fiscal Costs of Weak Governance: Evidence from Teacher Absence in India. *Mimeo*, UC San Diego.

DasGupta, M. (1987). Informal Security Mechanisms and Population Retention in Rural India. *Economic Development and Cultural Change*, 36(1):101–120.

de Brauw, A. and Giles, J. (2008). Migrant Opportunity and the Educational Attainment of Youth in Rural China. *World Bank Working Paper Series*, 1.

Deaton, A. (2010). Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, 48:425–455.

Dee, T. S. (2004). Teachers, Race, and Student Achievement in a Randomized Experiment. *The Review of Economics and Statistics*, 86(1):195–210.

Deininger, K. (2003). Does Cost of Schooling Affect Enrollment by the Poor? Universal Primary Education in Uganda. *Economics of Education Review*, 22(3):291–305.

Desai, S. and Kulkarni, V. (2008). Changing Educational Inequalities in India in the Context of Affirmative Action. *Demography*, 45(2).

Deshingkar, P. and Anderson, E. (2004). People on the Move: New Policy Challenges for Increasingly Mobile Populations. *Natural Resources Perspectives*, 92. Overseas Development Institute.

DiNardo, J. and Lee, D. S. (2011). Program Evaluation and Research Designs. *Handbook of Labor Economics*, 4(A)(5):463–536.

Dinkelman, T. and Martinez, C. (2014). Investing in Schooling in Chile: The Role of Information about Financial Aid for Higher Education. *Review of Economics and Statistics*, 96(2):244–257.

Domina, T. (2007). Higher Education Policy as Secondary School Reform: Texas Public High Schools After Hopwood. *Educational Evaluation and Policy Analysis*, 29(3):200–217.

Donald, S. and Lang, K. (2007). Inference with Difference-in-Differences and Other Panel Data. *The Review of Economics and Statistics*, 89(2):221–233.

Donaldson, D. (2014). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review (forthcoming)*.

Donaldson, D. and Hornbeck, R. (2015). Railroads and American Economic Growth: New Data and Theory. *NBER Working Paper 19213*.

Donohue, John J., H. J. J. (1991). Continuous Versus Episodic Change: The Impact of Civil Rights Policy on Economic Status of Blacks. *Journal of Economic Literature*, 29(4):1603–1643.

Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review*, 1(4):795–813.

Duflo, E. (2004). Medium Run Effects of Educational Expansion: Evidence from a Large School Construction Program in Indonesia. *Journal of Development Economics*, 74(1):163–197.

Duflo, E., Hanna, R., and Ryan, S. (2012). Incentives Work: Getting Teachers to Come to School. *American Economic Review*, 102(4):1241–78.

Duflo, E. and Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, 122(2):601–646.

Evans, P. (1997). How Fast Do Economics Converge? *The Review of Economics and Statistics*, 79(219-225):2.

Faber, B. (2014). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies*, 81(3):1046–1070.

Fogel, R. (1964). Railroads and American Economic Growth: Essays in Econometric History. *Baltimore: Johns Hopkins Press*.

Foster, A. and Rosenzweig, M. (1996). Technical Change and Human Capital Returns and Investments: Evidence from the Green Revolution. *American Economic Review*, 86(4):931–953.

Freeman, R. (1975). Legal Cobwebs: A Recursive Model of the Market for New Lawyers. *Review of Economics and Statistics*, (57):171–179.

Freeman, R. (1976). The Overeducated American. *New York: Academic Press*.

Fryer, R. and Loury, G. (2005). Affirmative Action and Its Mythology. *Journal of Economic Perspectives*, 19(3):147–162.

Fryer, R. and Torelli, P. (2010). An Empirical Analysis of 'Acting White'. *Journal of Public Economics*, 95(5-6):380–396.

Ghani, E., Goswami, A., and Kerr, W. (2015). Highway to Success in India: The Impact of the Golden Quadrilateral Project for the Location and Performance of Manufacturing. *The Economic Journal*.

GOI (1994). *DPEP Principles*, volume New Delhi. Government of India Report.

GOI (2000). *DPEP Moves on ... Towards Universal Primary Education*, volume New Delhi. Government of India Report.

Grembi, V., Nannicini, T., and Troiano, U. (2016). Do Fiscal Rules Matter? *American Economic Journal: Applied Economics*, (forthcoming).

Griliches, Z. (1977). Estimating the Returns to Schooling: Some Econometric Problems. *Econometrica*, 45:1–22.

Griliches, Z. (1997). Education, Human Capital, and Growth: A Personal Perspective. *Journal of Labor Economics*, 15:330–344.

Grogger, J. (1996). School Expenditures and Post-Schooling Earnings: Evidence from High School and Beyond. *The Review of Economics and Statistics*, 78(4):653–664.

Gruber, J. (1994). The Incidence of Mandated Maternity Benefits. *American Economic Review*, 84:622–641.

Hahn, J., Todd, P., and van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression Discontinuity Design. *Econometrica*, 69(1):201–209.

Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, 24(3):1141–1147.

Hanushek, E. A. (1997). Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis*, 19(2):141–164.

Hanushek, E. A. (2003). The Failure of Input-Based Schooling Policies. *The Economic Journal*, 113:64–98.

Hanushek, E. A. (2006). School Resources. *Handbook of the Economics of Education*, 2:865–908.

Hanushek, E. A. (2008). Education Production Functions. *The New Palgrave Dictionary of Economics*, Basingstoke.

Heckman, J., LaLonde, R., and Smith, J. (1999). The Economics and Econometrics of Active Labor Market Programs. *Handbook of Labor Economics*, 3A:1856 – 2097. Orley Ashenfelter and David Card (eds.).

Heckman, J. and Robb, R. (1985). Alternative Methods for Evaluating the Impact of Interventions. *Journal of Econometrics*, 30:239–267.

Heckman, J. J., Lochner, L. J., and Taber, C. (1998a). Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents. *Review of Economic Dynamics*, 1(1):1–58.

Heckman, J. J., Lochner, L. J., and Taber, C. (1998b). General Equilibrium Treatment Effects: A Study of Tuition Policy. *The American Economic Review: Papers and Proceedings*, 88(2):381–386.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *American Economic Review*, 102(2):994–1028.

Hickman, B. (2013). Pre-College Human Capital Investment and Affirmative Action: A Structural Policy Analysis of US College Admissions. *Mimeo*.

Higgins, M. J., Levy, D., and Young, A. T. (2006). Growth and Convergence Across the U.S.: Evidence from County-Level Data. *Review of Economics and Statistics*, 88:671–81.

Hirschman, A. O. (1969). The Strategy of Economic Development. *in Agarwal, A.N. and Singh, S.P.(eds), Accelerating Investment in Developing Economies (London Oxford Press).*

Hoxby, C. M. (2001). All School Finance Equalizations Are Not Created Equal. *Quarterly Journal of Economics*, 116(4):1189–1231.

Imbens, G. and Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–75.

Imbens, G. W. and Kalyanaraman, K. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies*, 79(3):933–959.

Jacob, V., Kochar, A., and Reddy, S. (2008). School Size and Schooling Inequalities. *Mimeo*, Stanford.

Jagnani, M. and Khanna, G. (2016). Does Public Investment in Higher Education Promote Private Provision of Education? Evidence from India. *Mimeo*.

Jalan, J. and Glinskaya, E. (2013). Small Bang for Big Bucks? An Evaluation of a Primary School Intervention in India. *Centre for Training and Research in Public Finance And Policy*, Calcutta.

Jensen, R. (2010). The (Perceived) Returns to Education and the Demand for Schooling. *Quarterly Journal of Economics*, 125(2):515–548.

Jensen, R. (2012). Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India. *Quarterly Journal of Economics*, 127(2):753–792.

Jensen, R. and Miller, N. (2015). Keepin em Down on the Farm: Old Age Security and Strategic Underinvestment in Children. *Mimeo*.

Kane, T. (1994). College Entry by Blacks since 1970: The Role of College Costs, Family Background and the Returns to Education. *Journal of Political Economy*, 102(5):878–911.

Katz, L. F. and Murphy, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107.(1):35–78.

Kazianga, H., Levy, D., Linden, L., and Sloan, M. (2013). The Effects of Girl-Friendly Schools: Evidence from the BRIGHT School Construction Program in Burkina Faso. *American Economic Journal: Applied Economics*, 5(3):41–62.

Khanna, G. (2014). That's Affirmative: Incentivizing Standards or Standardizing Incentives. *Mimeo*, Michigan.

Khanna, G. (2015). The Effectiveness of Education Programs: A Multi-Dimensional Regression Discontinuity Approach. *Mimeo*, Michigan.

Khanna, G. (2016). Large-scale Education Reform in General Equilibrium: Regression Discontinuity Evidence from India. *Mimeo*.

Khanna, G. and Morales, N. (2015). The IT Boom and Other Unintended Consequences of Chasing the American Dream. *Mimeo*, Michigan.

King, E. and Orazem, P. (2008). Schooling in Developing Countries: The Roles of Supply, Demand and Government Policy. *Handbook of Development Economics*, 4.

Kingdon, G. and Teal, F. (2010). Teacher Unions, Teacher Pay and Student Performance in India: A Pupil Fixed Effects Approach. *Journal of Development Economics*, 91(2):278–288.

Kochar, A. (2004). Urban Influences on Rural Schooling in India. *Journal of Development Economics*, 74:113–136.

Kohli, A. (2001). The Success of India's Democracy. *Cambridge University Press*.

Kremer, M. and Muralidharan, K. (2007). Public and Private Schools in Rural India. *School Choice International*, MIT Press.

Krishna, K. and Robles, V. F. (2015). Affirmative Action in Higher Education in India: Targeting, Catch Up, and Mismatch. *Higher Education*, 71(5):611–649.

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, 114(2):497–532.

Krueger, A. B. (2003). Economic Considerations and Class Size. *The Economic Journal*, 113:34–63.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, 99:483–499.

Lee, D. (2005). An Estimable Dynamic General Equilibrium Model of Work, Schooling, and Occupational Choice. *International Economic Review*, 46(1):1–34.

Leonard, J. S. (1984). The Impact of Affirmative Action on Employment. *ournal of Labor Economics*, 2(4):439–463.

Linden, L. (2008). Complement or Substitute? The Effect of Technology on Student Achievement in India. *InfoDev*, Columbia University, MIT Jameel Poverty Action Lab, IZA.

Loeb, S. and Bound, J. (1996). The Effect of Measured School Inputs on Academic Achievement: Evidence from the 1920s, 1930s and 1940s Birth Cohorts. *The Review of Economics and Statistics*, 78(4):653–664.

Loury, G. (1992). The Incentive Effects of Affirmative Action. *Annals of the American Association of Political and Social Science*, 523:19–29.

Mandal, B. (1980). Mandal Commission Report. *Government of India*.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics*, 107:407–438.

McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*, 142(2):698–714.

Menon, P. (2001). Content Analysis of training Modules for Village Education Committees: A Study of seven DPEP states (Phase I). *Mimeo*, NIEPA, New Delhi.

Michaels, G. (2008). The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. *The Review of Economics and Statistics*, 90(4):683–701.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial Ethnic Institutions and Contemporary African Development. *Econometrica*, 81(1):113–152.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4):281–302.

Munshi, K. and Rosenzweig, M. (2009). Why is Mobility in India so Low? Social Insurance, Inequality and Growth. *NBER working paper*, 14850.

Munshi, K. and Rosenzweig, M. (2015). Netoworks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap. *Mimeo*, Yale.

Munshi, K. and Rosenzweig, M. (2016). Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap. *American Economic Review*, 106(1):46–98.

Muralidharan, K. (2013). *Priorities for Primary Education Policy in Indias 12th Five-year Plan*. NCAER-Brookings India Policy Forum.

Muralidharan, K. and Sundararaman, V. (2010). The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India. *Economic Journal*, 120(546):F187–F203.

Muralidharan, K. and Sundararaman, V. (2015). The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India. *Quarterly Journal of Economics*, 130(3):1011–66.

Ogbu, J. (2003). Black American Students in an Affluent Suburb: A Study of Academic Disengagement. *Routledge*.

Oreopoulos, P. (2006). Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter. *American Economic Review*, 96:152–75.

Osili, U. O. and Long, B. T. (2008). Does Female Schooling Reduce Fertility? Evidence from Nigeria. *Journal of Development Economics*, 87(1):57–75.

Oster, E. and Steinberg, M. B. (2013). Do IT Service Centers Promote School Enrollment? Evidence from India. *Journal of Development Economics*, 104(15922).

Pal, S. (2010). Public Infrastructure, Location of Private Schools and Primary School Attainment in an Emerging Economy. *Economics of Education Review*, 29:783–794.

Pandey, R. S. (2000). Going to Scale with Education Reform: India's district Primary Education Program, 1995-99. *Washington DC: The World Bank*, 1(4).

Prakash, N. (2010). Improving the Labor Market Outcomes of Minorities: The Role of Employment Quota. *Mimeo*.

Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to Investment in Education: A Further Update. *Education Economics*, 12(2):111–134.

Rao, G. (2016). Familiarity Does Not Breed Contempt: Diversity, Discrimination and Generosity in Delhi Schools. *Mimeo*.

Redding, S. and Turner, M. (2015). Transportation Costs and the Spatial Organization of Economic Activity. *Handbook of Regional and Urban Economics*, 5:1–1653.

Roy, A. (1950). The Distribution of Earnings and of Individual Output. *Economic Journal*, 60(239):489–505.

Ryoo, J. and Rosen, S. (2004). The Engineering Labor Market. *Journal of Political Economy*, 112(S1):S110–S140.

Sala-i Martin, X. (1996). Regional Cohesion: Evidence and Theories of Regional Growth and Convergence. *European Economic Review*, 40:1325–52.

Sen, A. (1999). *Development as Freedom*. Random House Inc., United States of America.

Shreshta, S. (2016). No Man Left Behind: Effects of Emigration Prospects on Educational and Labor Outcomes of Non-migrants. *The Economic Journal*.

Sifuna, D. N. (2007). The Challenge of Increasing Access and Improving Quality: An Analysis of Universal Primary Education Interventions in Kenya and Tanzania Since the 1970s. *International Review of Education*, 53(5).

Singh, G. (1990). State of Haryana: Backward Classes Commission Report. *Government of Haryana*.

Smith, A. (1775). *An Inquiry into the Nature and Causes of the Wealth of Nations*. The Glasgow Edition, pp 29.

Smith, J. P. and Welch, F. (1989). Black Economic Progress After Myrdal. *Journal of Economic Literature*, 27(2):519–564.

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics*, 70(1):65–94.

Strauss, J. and Thomas, D. (1995). Human Resources: Empirical Modeling of Households and Family Decisions. *Handbook of Development Economics*, 3(34).

Topalova, P. (2005). Factor Immobility and Regional Impact of Trade Liberalization: Evidence on Poverty and Inequality from India. *Mimeo*, MIT.

van der Klaauw, W. (2008). Regression-Discontinuity Analysis: A Survey of Recent Developments in Economics. *Labour*, 22(2).

Varghese, N. (1994). District Primary Education Programme: the Logic and the Logistics. *Journal of Educational Planning and Administration*, 2(1):69–106.

Willis, R. J. (1986). Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions. *Handbook of Labor Economics*, 1:525–602.

WorldBank (1997). *Primary Education in India*. The World Bank, Washington DC.