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

Gaurav Khanna *
University of California – San Diego

Abstract

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 general equilibrium model and derive sufficient statistics that capture the consequences of such massive countrywide schooling initiatives. I provide unbiased estimates of the sufficient statistics using a Regression Discontinuity design generated by Indian government policy. The earnings returns to a year of education are 13.4%, and the general equilibrium labor market effects are substantial: they depress the returns to skill by 6.5 percentage points. 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.

JEL: I25, I26, O15, I28

Keywords: Returns to education, school building, general equilibrium

*gauravkhanna.info. Assistant Professor of Economics, School of Global Policy and Strategy, University of California – San Diego. I thank John Bound, Jeff Smith, Dean Yang and Ach Adhvaryu for guidance. I thank Raj Arunachalam, Hoyt Bleakley, Charles Brown, Sebastian Calonico, Matias Cattaneo, James Fenske, Gordon Hanson, Brian Jacob, David Lam, Karthik Muralidharan, Anant Nyshadham, Tommaso Porzio, Isaac Sorkin, Kevin Stange, and Mel Stephens for valuable comments. I am grateful to seminar participants at UMichigan, NEUDC (Brown), Urban Institute, McGill, UC Davis, UCSD, William and Mary, Georgia State, Mathematica, PAA, Society of Labor Economists, Stockholm School of Economics, Stockholm University, IFPRI, UReno, The World Bank, UToronto, and the Center for Global Development for valuable comments. I thank the Alfred P Sloan Foundation & NBER, and the Rackham Fellowship for generous research support.

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. I causally estimate and take into account these GE effects in determining the overall economic consequences and benefits of nationwide education programs.

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 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. The allocation rule under which Indian districts receive the funding allows me to 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—these effects depress the returns to skill by 32% and dampen the increase in labor market benefits by 23%. By expanding the skilled workforce, the policy makes skilled workers worse off and unskilled workers better off, and leads to the adoption of skill-biased capital. 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 models of education choice (Becker, 1967) with frameworks that determine the skill-premium (Card and Lemieux, 2001) 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 for welfare. 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). If education levels rise, 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 the earning skill premium. At the same time, as more skilled workers join the labor force, skill-biased capital may be adopted by firms in these regions, raising the premium. Last, as workers switch to more productive skill groups, overall output may increase to the benefit of all workers. I, therefore, estimate all such components of the GE effects to better quantify the distributional impacts and the overall increase in labor market benefits.

The policy I study was India's flagship scheme in the 1990s and early 2000s, the District Primary Education Program (DPEP), which expanded public schooling in half the country by targeting low-literacy regions. 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 would induce some students who have potentially high returns to schooling but could not previously afford it to get more education.

Under the allocation rule, districts that had a female literacy rate below the national average were more likely to receive the program. I compare regions on either side of the cutoff to estimate causal impacts. The RD design allows me to tackle biases that arise when estimating the individual returns to education, and when comparing earnings in two different local economies. I 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 skilled industries, 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, test score surveys and the Indian Census. I assemble a 10 year long panel of districts to track long-run outcomes, allowing me to follow local labor and education markets over time.

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. 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 – both young and old. At the same time, since the young and old are not perfect substitutes (and in some contexts may be complements), 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. Using the estimated parameters, I measure the overall impact of the policy on welfare for the different types of workers and cohorts.

Given evidence from other contexts, it is important for researchers to address these labor market 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 GE effects.[1] 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, estimate crucial economic parameters, and schooling returns both in the presence and absence of GE effects.

Duflo (2004) shows that Indonesia school-building depressed average wages for older untreated cohorts. Average wages for untreated cohorts, however, may have limited information as skilled wages fall and unskilled wages rise. In contrast, my method estimates the GE effects on each skill-group separately, and for *all* cohorts (including treated cohorts), thus allowing me to separately identify the returns to education in both partial and general equilibrium.[2] As the young and the old are not perfect substitutes, and may, in some cases, be complements in production, it is important to identify the GE effects for each cohort separately. Indeed, I find that the GE effects on the old are negligible in comparison to the GE effects on the (treated) young. I corroborate my results with a Difference-in-Differences (DID) analysis similar to Duflo (2001), where I compare treated to untreated districts and the younger to older cohorts. Using a DID design, however, it is difficult to recover the GE effects as the portion that affect both the young and old 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.

There are already a substantial number of micro-interventions that can help guide policy-makers in supply-side interventions.[3] These micro-interventions are, however, inherently different from large school expansion programs, and my analysis is informative of how to measure the effects of scaled-up versions of such interventions. 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.[4] A

[1] There are other types of labor market GE effects in other contexts: job assistance programs (Crepon et al., 2013), wage depression in segmented labor-markets (Angrist, 1995), calibrated macro models (Albrecht et al., 2009), effects on demographics and house prices (Epple and Ferreyra, 2008), and more recently Bianchi (2016) shows how college major choice in Italy can affect returns.

[2] Given elasticities of substitution across skill and cohorts, the GE effects may indeed raise average wages for older cohorts, as the unskilled wage rises.

[3] In India 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 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., 2013; Duflo et al., 2012; Muralidharan, 2013).

[4] 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., 2013), and India (Afridi, 2010; Chin, 2005).

concern with an expansion in public schooling is that it may crowd out private supply negating the effects of public funds that could have been spent elsewhere. 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. I model and estimate this change in private supply, and in line with other work (Andrabi et al., 2013), I find an influx of private schools when public schooling grows.

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 household 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 percentage 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 younger cohorts. High-skill workers who did not change their educational levels under the policy are adversely affected, whereas low-skill workers benefit.

My analysis allows for both the mobility of workers across regions and the adoption of skill-biased capital or technology. Importantly, the adoption 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.

These results have four significant implications for research and policy. First, it speaks to our attempts to causally estimate the private returns to education using macro-level variation from tuition reductions, changes in compulsory schooling laws, schooling expansions or other large-scale policy reforms. Such large-scale variation identifies a different parameter, as it is no longer one person being treated with a year of education as envisioned by Becker (1967); Mincer (1958); Willis (1986). Rather it involves treating an entire cohort of students, leading to GE effects, which I find to be important. Second, I show that the returns to education is not a fixed parameter, but rather an endogenous quantity which depends on the local labor market, and I derive meaningful relationships between labor market changes and the returns.

Third, GE effects will either undermine or amplify the effectiveness of micro-interventions when they are scaled up (Acemoglu, 2010; Deaton, 2010), and this is particularly important for researchers in development performing micro-experiments, even as the point is also stressed outside the realm of Development Economics (Heckman et al., 1999). Last, the methodology developed in this paper can be applied in other similar contexts, allowing researchers and policy-makers to estimate the welfare consequences of such interventions around the world (for e.g. school-building, compulsory schooling, fee reductions or job training programs).

1 The District Primary Education Project (DPEP)

I use plausibly exogenous variation generated by a large schooling expansion policy (the District Primary Education Project (DPEP)) in India. Any district below the national average female literacy rate was eligible to receive funds. This allows for an RD design, isolating the impact of the policy from other changes. To causally estimate the parameters of the model I compare districts that should have received the policy to those that should not have, on either side of the RD cutoff. There are two additional advantages of using districts around the national average. The first is that there is a large density of districts at the RD cutoff, and the second is that this analysis is representative of the district with average literacy induced into receiving the policy. In this section I discuss the program; additional details on the history, funding and secondary objectives can be found in Appendix C.

In 1994, the District Primary Education Project (DPEP) was introduced, eventually serving 271 of approximately 600 districts in the country. DPEP grew with funding from international agencies, making it one of the largest donor assisted programs in the world (Jalan and Glinskaya, 2013). States 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.[5] In 2002-3 alone, $345 mn of foreign funds was spent – concentrated in less than half the districts in the country, allowing for a valuable policy experiment. But the project was rapidly discontinued in 2006, when only $24mn was spent.[6]

The program claims to have covered about 271 low literacy districts, and served approximately 51.3 million children and 1.1 million teachers (Jalan and Glinskaya, 2013). These districts were geographically dispersed all over the country (map in Appendix Figure A.1). It created about 160,000 new schools (Azam and Saing, 2016), and trained about 1 million teachers and 3 million community members. In the project states, it increased the average allocation of funds for primary school education by between 17-20% (Jalan and Glinskaya, 2013).[7]

The primary objective of the program was to improve student access to and retention in primary and upper primary education by building schools, hiring teachers, supporting school and community organizations, constructing new classrooms and improving existing school facilities. 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

[5] 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. Taxes were not raised to directly fund DPEP.

[6] Source: Parliamentary Questions. Lok Sabha Unstarred Question Numbers: 1807, 552, 55, 267, 1320, 2018, and Rajya Sabha Unstarred Question No. 2855

[7] In this period, DPEP was the flagship education program, despite being restricted to less than half the country. 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 in the country, like the Mid-day Meal Scheme ($232 mn), and Operation Blackboard ($130 mn).

of private and government-aided schools.

There are numerous World Bank and Government of India briefs and media reports that refer to the program's success.[8] Current work uses difference-in-differences methods to show that DPEP increased the years of education in treated districts (Azam and Saing, 2016). Similarly, a working paper by Jalan and Glinskaya (2013) uses a difference-in-differences methodology to study enrollment in the first 42 districts to receive funds. They find that five years after the program started, enrollment and grade progression of minority groups improved only in some specific states. 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 allowed to receive about $8 million, which came to approximately $9.1 per student. Jalan and Glinskaya (2013) estimate that this intervention lowered 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 first few districts in 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 districts received funding.[9]

2 The Model, Comparative Statics and Welfare

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 derive estimation equations and identify sufficient statistics that determine the effect of schooling expansion policies on economic benefits, as discussed in Section 5.[10]

The demand for education (skill) is determined by students' optimization decisions, that also depend on the labor market returns and the general equilibrium effects on earnings. The supply depends on the choices made by both public and private schools. Building new schools and increasing access to schools will reduce the marginal cost of schooling (Behrman et al., 1996; Birdsall, 1985); directly raising household welfare, and inducing more education.[11]

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

[9] Other descriptive studies examine enrollment outcomes for DPEP districts, and by and large consider the program to be a success (Aggarwal, 2000; Menon, 2001; Pandey, 2000). However, they do not compare DPEP districts to others, and hence cannot distinguish between the changes in overall education taking place all across the country.

[10] As explained in Section 5.1.1, the advantage of sufficient statistics is that the estimation procedure and measured GE effects do not depend on the specific functional form of the production functions. Indeed, the estimated wage benefits will hold true even under many alternative formulations, like signaling models. However, couching it in a canonical labor economics model allows us to understand parameters driving the different effects, drawing links to important elasticities estimated in other work, and determinants of the returns to education.

[11] 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.

2.1 Economic Production and the Labor Market

Aggregate output Y_d in district d depends on L_d (effective labor) and K_d (capital).[12] Capital is perfectly elastically supplied across districts at rental rate R^*.[13] Effective labor supply L_d depends on the labor aggregate L_{sd} at each skill level s.

$$Y_d = L_d^{\varrho} K_d^{(1-\varrho)} \quad where \quad L_d = \left(\sum_s \theta_{sd} L_{sd}^{\frac{\sigma_E - 1}{\sigma_E}} \right)^{\frac{\sigma_E}{\sigma_E - 1}} \tag{1}$$

$0 < \varrho < 1$ 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 $\theta'_{sd}(k_{sd}) > 0$.[14] 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:

$$L_{sd} = \left(\sum_a \psi_a \ell_{asd}^{\frac{\sigma_A - 1}{\sigma_A}} \right)^{\frac{\sigma_A}{\sigma_A - 1}} \tag{2}$$

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.[15] A worker gets paid their marginal product. The average log earnings are therefore:[16]

$$\log w_{asd} = \log \left(\frac{\partial Y_d}{\partial \ell_{asd}} \right) = \log \widetilde{\varrho} + \log \theta_{sd} + \log \psi_a + \frac{1}{\sigma_E} \log Y_d + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E} \right) \log L_{sd} - \frac{1}{\sigma_A} \log \ell_{asd} \ , \tag{3}$$

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.[17]

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 (4):

[12] Adding non-tradables like land into the aggregate production function Equation (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 examining the returns to owners of capital and land.

[13] The perfectly elastic capital assumption is not essential. The results are unaffected by assuming a fixed capital stock (see Appendix B.III).

[14] For completeness, in Appendix B.IV 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.

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

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

[17] For tractability, I have ignored the role played by changes in prices. It is straightforward to include a $\log P_d$ that will be associated with the $\frac{1}{\sigma_E} \log Y_d$ term, and not affect the returns to skill.

$$\log\left(\frac{w_{asd}}{w_{a's'd'}}\right) = \underbrace{\log\left(\frac{\theta_{sd}}{\theta_{s'd'}}\right)}_{\text{productivity}} + \underbrace{\log\left(\frac{\psi_a}{\psi_{a'}}\right)}_{\text{cohort}}$$
$$+ \underbrace{\frac{1}{\sigma_E}\log\left(\frac{Y_d}{Y_{d'}}\right)}_{\text{output}} + \underbrace{\left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right)\log\frac{L_{sd}}{L_{s'd'}}}_{\text{skill distribution}} \underbrace{-\frac{1}{\sigma_A}\log\frac{\ell_{asd}}{\ell_{a's'd'}}}_{\text{skill-cohort distribution}} \tag{4}$$

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 affects earnings. The second component – 'cohort' – captures age-specific productivities and returns to experience ψ_a. The third – 'output' – is the difference across labor markets related to differences in the size of the economy. 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 GE effects that affect all cohorts. Last – 'skill-cohort distribution' – affects 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 distribution by age will have important GE effects on earnings.

The model highlights the importance of elasticities: how much the skill distribution affects the difference in earnings depends on the elasticities of substitution σ_E and σ_A. For instance, 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}:

$$log\frac{w_{asd}}{w_{aud}} = log\frac{\theta_{sd}}{\theta_{ud}} + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right) log\frac{L_{sd}}{L_{ud}} - \frac{1}{\sigma_A}log\frac{\ell_{asd}}{\ell_{aud}} \equiv \beta_{asd} \tag{5}$$

This highlights an important fact: the returns to education are not an exogenous parameter, but rather an endogenous variable that depends on local labor market conditions: 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}. 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.

Importantly, migration flows are directly incorporated in this set up. If migration is not skill-biased it will not affect the relative quantities in Equation (5). However, skill-biased migration will change these quantities and affect skill-premia. Theory suggests that such flows will be in the direction of equalizing wages, attenuating the negative GE effects. I, therefore, also estimate changes in migration patterns across the cutoff to see how they effect changes in the skill distribution.

2.2 Students' Decisions

Students, 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 2.1, these choices will also eventually affect earnings, and lifetime utility.[18] 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 $\kappa(s_{id})$:[19]

$$\max_{s_{id}} \ \log \ w_{aid}(s_{id}) - \left(\log \ r_{id} + r_{id}s_{id} + \frac{1}{2}\Gamma s_{id}^2\right) \ , \tag{6}$$

where Γ is the quadratic cost parameter. Equations (4) and (5) 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:

$$\log \ w_{aid}(s_{id}) = \gamma_d + \nu_a + \beta_{asd}s_{id} + \log \ \epsilon_i \ , \tag{7}$$

where ϵ_i is the ability of the worker (not observable to researchers) the distribution of which is the same across districts. This ability is correlated with the marginal costs of schooling r_{id} and leads to biases in standard OLS regressions ($corr(\epsilon_i, r_{id}) \neq 0$) – high-ability workers earn high wages but have lower costs of schooling. 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 (7), 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 captured in Equation (4). Given this setup, from the first order conditions one can obtain the optimal years of education for person i:

$$s_{id}^* = \frac{\beta_{asd} - r_{id}}{\Gamma} \tag{8}$$

The variation in s_{id}^* within a district d is driven entirely by the variation in the marginal cost parameter r_{id}.[20] 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

[18] Appendix B.I derives this problem from a conventional utility function.

[19] Since the linear form of $\kappa(s_{id})$ only captures the opportunity costs, Card (1999) suggests a more general formulation of the cost function, to capture credit and other monetary constraints – 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).

[20] Notice, however, that the distribution of earnings in district d is driven both by the costs of education r_{id}, and by ϵ_i abilities. 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 much 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."

heterogeneity in the costs of schooling. The district-level costs depend on the access to schooling A_d (distance to the nearest school) and the monetary price of school p_d (like school fees).[21]

$$r_{id} \equiv -\Psi A_d + p_d + \eta_i \ , \tag{9}$$

where Ψ represents how aggregate access to education affects each i individual. 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.[22] These education decisions are a nested portion of the problem where individuals maximize their consumption-driven lifetime utility (see Appendix B.I).

2.3 Schools

While public schools aim to increase access to schooling for citizens, private schools care about profits (Kremer and Muralidharan, 2007). Both can have heterogeneous costs or efficiency, but provide the same output. Students merely choose the school that is less costly, where costs not only depend on school fees p_d, but also transportation and non-monetary costs A_d.

2.3.1 District Level Public School Administrator's Decisions

Public school administrators for district d maximize the access to schooling A_d for students by investing in inputs $\mathbf{x_m}$, such as schools, teachers and infrastructure. As access to schooling is increased, this reduces the marginal costs of going to school for students. By building more schools, public officials reduce distances to the nearest school and increase access to schools. They do, however, have a budget constraint that restricts their spending. The district d receives R_d from the government, and spends p_m for each input x_m into the schooling production function. Funds received under government schemes will increase the value of R_d:[23]

$$\max_{\mathbf{x_m}} \ A_d(\mathbf{x_m}) \ \text{ s.t. } \sum_{\mathbf{m=1}}^{\mathbf{M}} \mathbf{p_m x_m} \leq \mathbf{R_d} \ , \tag{10}$$

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.[24]

[21] 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.

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

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

[24] See Appendix B.II for a parametric example of this set-up.

2.3.2 Private schools

Building 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 profit maximizers and price takers in the competitive market charging a fee p_d. Muralidharan and Sundararaman (2015) are among the first to provide causal evidence that students in private schools have similar test scores as public school students for subjects taught in both. They are, however, more cost-effective. Private schools, in my model, therefore, have the same output as publics, but may do so at a different cost; and there is heterogeneity in these costs (Kremer and Muralidharan, 2007).[25]

Total educational output (in student-years) Q_{jd} by school j is a function of X_{jd} its aggregate inputs: $Q_{jd} = \overline{\theta_d} X_j$, and the average skill level of the district $\overline{\theta_d}$. This captures demand externalities (Birdsall, 1982), peer effects in school participation (if students are encouraged to go to school, then demand from neighbors may rise (Bobonis and Finan, 2009)), and self-segregation motives (as low income students enter public schools high income students demand more private schools (Lucas and Mbiti, 2012)). The school chooses inputs to maximize profits:

$$\max_{X_j} \; p_d \overline{\theta_d} X_j - Z(X_j) \tag{11}$$

The costs $Z(X_j) = z_{1j} X_j + \frac{1}{2} z_{2d} X_j^2$ have a simple quadratic formulation.[26] There is a heterogeneity in costs z_{1j} across schools (some schools are more cost effective) and a heterogeneity in costs z_{2d} across districts, where certain districts have better infrastructure for setting up a school and access to more teachers. The supply curve and profits are:

$$Q_{jd} = \overline{\theta_d} X_j^* = \overline{\theta_d} \frac{p_d \overline{\theta_d} - z_{1j}}{z_{2d}} \quad and \quad \pi_{jd} = \frac{(p_d \overline{\theta_d} - z_{1j})^2}{2 z_{2d}} \tag{12}$$

Since there is free entry of private schools into these regions, schools will enter until $\pi_{jd} = 0$. 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 the strongest driving force is that an increase in the supply of public schooling drives down the equilibrium price and crowds-out private schools.

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, θ_d, drive up the equilibrium price and induce private schools to enter. The second is that infrastructure upgrades and the presence of more teachers lowers the operating costs, z_{2d}, lead

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

[26] 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.

to more private school entry and further lower the equilibrium price. The price is, therefore, informative in distinguishing between these channels and pin down the mechanism.

The best evidence for how private schools respond comes from Andrabi et al. (2013), who show that an expansion in public schooling increased education for girls, and these girls became teachers in Pakistani districts. This allowed private schools to enter the market soon after. Similarly, Jagnani and Khanna (2018) and Pal (2010) find that physical infrastructure upgrades induce private-school entry in India.[27]

2.4 Definition of an Equilibrium

The exogenous elements are the utility, cost and production functions, and the amount of government spending on schooling. What is endogenous is the amount of and returns to schooling, the optimal inputs in schools and schooling supply, firm output, and the equilibrium price and quantity of schooling.[28] Appendix B.II characterizes and derives the education-sector equilibrium. For product markets to clear, the amount of consumption must equal the amount of output Y_{td}. For labor markets to clear, the demand for workers ℓ_{asd} with education level s (Equation (3)) must equal the supply from the equilibrium amount of schooling.

Proposition 1 (Equilibrium) *Given the following dimensions of the model: A student utility function $U(C)$, earnings 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, 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 price of going to school p^*_d; optimal inputs into the access function $x^*_m(R_d, \alpha_{\mathbf{m}}, \mathbf{p_m})$; optimal private school inputs $X^*_j(p_d, z_{1j})$; equilibrium earnings w_{asd} and quantities of each type of worker ℓ_{asd}.*

As the model derived equations are directly used in the estimation of GE effects, we return to these equations in Section 5.

[27] See Appendix B.II for a derivation for the overall supply of private schooling in the region – in the parametric formulation, it is easy to see that $\frac{\partial p^*_d}{\partial z_{2d}} > 0$.

[28] In line with the literature, so far I have assumed perfect foresight. Students know the earnings-returns *including* the GE effects. If expectations were adaptive, the equilibrium is approached slowly as each cohort revises its expectations. For a cobweb style model see Freeman (1975).

3 Data

For a comprehensive analysis I combine a number of large datasets. The data are merged at the district level as 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 education, migration decisions and schooling expenditures, labor market data on earnings and occupations, and firm-level data on types of manufacturing in the different regions.

For certain variables I can study the dynamic consequences of the policy. 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. Data details can be found in Data Appendix D.

Data for inputs into schools comes from the District Information System for Education (DISE), established by the Indian government to have annual data on the education sector, covering about 1.45 million schools across the country by 2012. I compile data at the school-level for all available waves between 2005 and 2012, including data on the 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. Table 1 summarizes the variables of interest in the year 2005, at the end of DPEP. The top panel 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.

To study educational outcomes, I use household surveys and Census data. The Census has detailed tables at all three of the administrative levels - states, districts and sub-districts, but has a limited number of outcome variables, and no information on earnings or the detailed level of education by age. I create a panel of districts using the 1991 Census as a baseline. Districts splits-and-merges are well documented by official Census crosswalks, so I normalize any changes to districts to be at the 1991 level. The 1991 Census female literacy rate is the running variable for the RD.

I use three different rounds of the National Sample Survey (NSS) to study the impact on education, earnings, expenditures, migration and other labor-market characteristics. It is the largest nationally representative 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. It also includes a detailed consumption module that asks questions about expenditures, including educational expenditures. Summary statistics for the 2009 NSS round are presented in Table 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.

The three rounds of the NSS data I use are the 2004-5, 2007-8 and 2009-10 rounds. The 2004-5

"thick" round on consumption expenditures 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. Yet, it does not ask the precise location of the household before they migrated. My 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 labor-force 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 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. Post 2009, the only available labor force survey is the 2011 round, but I avoid using both the 2011 Census and NSS survey as the next flagship education law, the Right to Education (RTE) act, may affect outcomes once it was passed in 2010.

To study firm behavior, I use the Annual Survey of Industries (ASI), 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, and number of employees. I use these data to study whether changes in the skill level of the population can affect production technique and skill biased capital. In particular, I use information on mechanization in production, and the average compensation for employees.

I use the Annual Status of Education Report (ASER) to study impacts on test scores. This data is collected annually by an NGO (Pratham), and surveys children between the ages of 3 and 16. These surveys are done at home, so as to capture school dropouts as well. These surveys did not exist prior to DPEP, and I use all available surveys between 2007-8 and 2011-12. Last, I use District (Gross) Domestic Product data that is compiled by each state's statistical office. The data are reported in real terms, with a base year of 2000.

4 Estimation: A Regression Discontinuity Design

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. 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. As the districts were around the national average, there is a large density at the cutoff (Figure 1), allowing for more robust identification using the RD, and less concerns about outliers driving results. Furthermore, my estimates should be thought of being representative of the average-literacy district induced into taking up treatment, as the lack of perfect compliance implies that I am estimating an unbiased LATE.

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. Indeed, at the cutoff, we should expect no discontinuity in pre-policy labor market characteristics, skill biased capital 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.

Since more able workers are also more capable students, OLS estimates are biased, and the variation generated by the policy overcomes 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 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.

The first stage is presented in Figure 2. It is clear that the more literate among the eligible districts (i.e. among 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. An 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 (Imbens and Angrist, 1994).

The parameters, like the estimated returns to education, are for the students who were induced into getting more schooling and lived in districts near the cutoff that took up the policy. The GE effects as well depend on what type of students get induced into getting more skill, as this may affect the amount of skill-biased capital adopted by 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 as the cutoff chosen was the national average of the female literacy rate from the previous 1991 Census. Furthermore, McCrary (2008) tests indicate that the density of districts and of households around the cutoff is not discontinuous (Figure 1), since the p-value of the change in density at the cutoff is 0.71. Importantly, Figure 1 also

shows that there is a high density of districts at the national average, allowing me to have enough data for RD estimation. Other falsification tests will be discussed below that solidify the RD assumptions that there were no other discontinuities at the same cutoff.[29]

While RD results will be presented graphically, the coefficients of interest will also be calculated where the optimal bandwidth is calculated using two different methods. I use the Calonico et al. (2014b) and the Imbens and Kalyanaraman (2012) methods 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.[30] Results using both optimal bandwidth procedures are presented, and are robust to using parametric approaches like local linear and quadratic control-function approaches as suggested by Hahn et al. (2001) and Imbens and Lemieux (2008).[31]

5 Identification of Economic Benefits

5.1 Using Policy Changes to Estimate Parameters

The variation in schooling, s^*_{id}, is driven entirely by the variation in the marginal costs of schooling 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. Comparisons in the cross-section across different labor markets will provide biased estimates due to underlying baseline differences in the skill distribution and skill-biased capital across these markets (Equation (4)). From Section 2 we derived that the equilibrium amount of schooling is affected by the expansion of public schooling:[32]

$$S^*_d = \phi_1 \beta_{asd} + \phi_2 R_d - \frac{\eta_d}{\Gamma} \tag{13}$$

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 B.II). The term $\phi_1 \beta_{asd}$, captures how changes in the returns to education will affect equilibrium

[29] 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.

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

[31] The results are robust to using various alternative procedures (Appendix Table A.8). The first, described in Bartalotti and Brummet (2017) allows for computing the standard errors at an aggregated level. The second method allows for different-sized optimal bandwidths on either side of the cutoff and for nearest neighbor standard errors at an aggregate level (Calonico et al. (2017)).

[32] See Appendix B.II for a parameterization of ϕ_1 and ϕ_2, where $\phi_1 \equiv \left(\frac{\overline{\theta^2_d}}{\Gamma\overline{\theta^2_d}+z_{2d}}\right) > 0$ and $\phi_2 \equiv \left(\frac{\left(z_{2d}+\Psi\overline{\theta^2_d}\right)\left(\prod_m \frac{\alpha_m}{p_m}^{\alpha_m}\right)}{\Gamma\overline{\theta^2_d}+z_{2d}}\right) > 0$, and $\eta_d = \mathbb{E}[\eta_i | i \in d]$..

schooling. If, for instance, the labor-market general equilibrium effects substantially lower the returns to education β_{asd}, then there may be no 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 unobserved district-level characteristics causing biased estimates in OLS regressions. On the other hand, we should expect that η_d is not different for districts that just fall on either side of the cutoff, allowing the RD to recover unbiased estimates.

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:

$$S_d = \phi D_d + \frac{\eta_d}{\Gamma} \quad and \quad \mathbb{E}[\eta_d | D_d = 1] = \mathbb{E}[\eta_d | D_d = 0] \tag{14}$$

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 $\phi > 0$.

5.1.1 Returns to Education and Disentangling Earnings

An important role that the model plays is in deriving an equation for the returns to education as a function of quantities of labor and skill-biased capital. This can be strictly linked to the policy which changes the distribution of earnings across the RD cutoff. In Equation (3), 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 GE effect that affects all cohorts:[33]

$$\log w_{asd} = \log \widetilde{\varrho} + \log \theta_{sd} + \log \psi_a + \frac{1}{\sigma_E} \log Y_d + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right) \log L_{sd} - \frac{1}{\sigma_A} \log \ell_{asd} \tag{3}$$

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, in treated districts, that were too old to change their years of education when the policy was implemented will be affected by some part of the labor-market general equilibrium effects. The general equilibrium effects that affect all cohorts can thus be isolated by looking at the impact on the skill-premium of older 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. As

[33] 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. The functional form is to better understand the role played by underlying economic parameters.

the young and old are not necessarily perfect substitutes, and may indeed be complements in production, it is important to estimate effects for each cohort separately.[34]

For ease of exposition I restrict the analysis to two skill levels – skilled s and unskilled u workers. For instance, 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, the GE effects would be infinitesimally small. If the person lives in the untreated region $D = 0$, then that person's earnings on acquiring skill would increase by:

$$log\frac{w_{as,D=0}}{w_{au,D=0}} = \underbrace{log\frac{\theta_{s,D=0}}{\theta_{u,D=0}}}_{\text{Productivity}} + \underbrace{\left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right) log\frac{L_{s,D=0}}{L_{u,D=0}}}_{\text{Aggregate skill distribution}} \underbrace{-\frac{1}{\sigma_A} log\frac{\ell_{as,D=0}}{\ell_{au,D=0}}}_{\text{Cohort specific skill distribution}} \equiv \beta_{as,D=0} \ , \quad (15)$$

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

$$log\frac{w_{as,D=1}}{w_{au,D=1}} = log\frac{\theta_{s,D=1}}{\theta_{u,D=1}} + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right) log\frac{L_{s,D=1}}{L_{u,D=1}} - \frac{1}{\sigma_A} log\frac{\ell_{as,D=1}}{\ell_{au,D=1}} \equiv \beta_{as,D=1} \ , \quad (16)$$

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 across regions because of the differences in the amount of skill-biased capital $\left(log\frac{\theta_{s,D=1}}{\theta_{u,D=1}} - log\frac{\theta_{s,D=0}}{\theta_{u,D=0}}\right)$, in the size of the skilled workforce $\left(log\frac{L_{s,D=1}}{L_{u,D=1}} - log\frac{L_{s,D=0}}{L_{u,D=0}}\right)$, and also the size of the young skilled $\left(log\frac{\ell_{as,D=1}}{\ell_{au,D=1}} - log\frac{\ell_{as,D=0}}{\ell_{au,D=0}}\right)$.

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 adoption of skill biased capital. These are the GE effects on the returns to education:

$$\Delta\beta_{as} = \underbrace{\left(log\frac{\theta_{s,D=1}}{\theta_{u,D=1}} - log\frac{\theta_{s,D=0}}{\theta_{u,D=0}}\right) + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right)\left[log\frac{L_{s,D=1}}{L_{u,D=1}} - log\frac{L_{s,D=0}}{L_{u,D=0}}\right]}_{\text{GE effects on all cohorts}}$$
$$\underbrace{-\frac{1}{\sigma_A}\left[log\frac{\ell_{as,D=1}}{\ell_{au,D=1}} - log\frac{\ell_{as,D=0}}{\ell_{au,D=0}}\right]}_{\text{Additional GE on young}} \quad (17)$$

In order to disentangle the GE 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

[34] 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.

RD cutoff are only due to differences in the skill distribution and the amount of skill-biased capital.

The change in returns in Equation (17) 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:[35]

$$\underbrace{log\frac{w_{so,D=1}}{w_{so,D=0}} - log\frac{w_{uo,D=1}}{w_{uo,D=0}}}_{\text{GE effects on all cohorts}} = \underbrace{\left(log\frac{\theta_{s,D=1}}{\theta_{u,D=1}} - log\frac{\theta_{s,D=0}}{\theta_{u,D=0}}\right)}_{\text{Skill biased capital}} + \underbrace{\left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right)\left[log\frac{L_{s,D=1}}{L_{u,D=1}} - log\frac{L_{s,D=0}}{L_{u,D=0}}\right]}_{\text{Aggregate skill distribution}} \tag{18}$$

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 leads to the adoption of skill-biased capital and raises the skill premium. On the other hand, increasing the relative supply of skilled workers makes them less valuable, lowering the skill premium.

Migration directly affects these quantities. 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 by altering the size of the skilled workforce in treated districts. The model, therefore, explicitly incorporates migration in determining what the GE effects will be.

The second component of the GE effects from Equation (17) is the additional GE effect on the young y, driven solely by the change in the age-specific skill distribution. This component can be measured by estimating the earnings differential between the skilled young and unskilled young, and differencing out the earnings differential between the skilled unskilled old:[36]

$$\underbrace{\left[log\frac{w_{sy,D=1}}{w_{sy,D=0}} - log\frac{w_{uy,D=1}}{w_{uy,D=0}}\right] - \left[log\frac{w_{so,D=1}}{w_{so,D=0}} - log\frac{w_{uo,D=1}}{w_{uo,D=0}}\right]}_{\text{Additional GE on young}} = \underbrace{-\frac{1}{\sigma_A}\left[log\frac{\ell_{ys,D=1}}{\ell_{yu,D=1}} - log\frac{\ell_{ys,D=0}}{\ell_{yu,D=0}}\right]}_{\text{Age specific skill distribution}} \tag{19}$$

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:[37]

[35] Regardless of the specific formulation of the production function, the change in the skill premium for older cohorts will be the GE effects on all cohorts, and estimates of the returns and cohort-specific GE effects will empirically rely on the left hand sides of equations (18) and (19). For instance, this is true even if wage returns are determined in a purely signaling model. The right hand sides merely helps us understand the underlying economic parameters.

[36] Notice that if $\sigma_A < \sigma_E$ then the two components may be of opposite signs.

[37] See Appendix B.V for detailed derivations of these equations.

$$log\frac{w_{y,D=1}}{w_{y,D=0}} = \ell_{sy,D=1}log\frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=1}log\frac{w_{uy,D=1}}{w_{uy,D=0}} + \Delta\ell_{sy}\underbrace{log\frac{w_{sy,D=0}}{w_{uy,D=0}}}_{\beta_{ys,D=0}} \tag{20}$$

$$log\frac{w_{y,D=1}}{w_{y,D=0}} = \ell_{sy,D=0}log\frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=0}log\frac{w_{uy,D=1}}{w_{uy,D=0}} + \Delta\ell_{sy}\underbrace{log\frac{w_{sy,D=1}}{w_{uy,D=1}}}_{\beta_{ys,D=1}} \tag{21}$$

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 times 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:

$$\begin{aligned}\Delta S &= (\ell_{sy,D=1}s_1 + \ell_{uy,D=1}s_0) - (\ell_{sy,D=0}s_1 + \ell_{sy,D=0}s_0) \\ &= \Delta\ell_{sy}s_1 + \Delta\ell_{uy}s_0 = \Delta\ell_{sy}(s_1 - s_0)\ ,\end{aligned} \tag{22}$$

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 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:[38]

$$log\frac{w_{s,D=1}}{w_{s,D=0}} = \frac{1}{\sigma_E}log\frac{Y_{D=1}}{Y_{D=0}} + log\frac{\theta_{s,D=1}}{\theta_{s,D=0}} + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right)log\frac{L_{s,D=1}}{L_{s,D=0}} - \frac{1}{\sigma_A}log\frac{\ell_{as,D=1}}{\ell_{as,D=0}} \tag{23}$$

5.2 Outcomes and Economic Benefits

The economic benefits depend on the changes in the wage distribution across the cutoff. Indeed, the benefits for the workers induced into getting more skill are directly equal to the sum of the partial equilibrium returns to education and the GE effect on skilled wages. This highlights the importance of estimating *both* parameters in order to measure welfare changes.

The model predicts that when a district receives more funds, the following happens: First, public administrators build more schools, increasing the access to schooling (Section 2.3.1). This lowers the marginal cost of schooling for households, and induces certain students to get more education (Section 2.2). Private schools decide whether to enter or exit the education sector, leading to either a crowd-in or crowd-out of schools (Section 2.3.2). When the newly skilled workforce joins the labor market they earn the higher skilled wage (Section 2.1). There is, however, a distributional impact on the earnings of skilled and unskilled workers (Section

[38] See Appendix B.III for details on modeling skill biased capital. If aggregate output prices change, then the skill-premium is unaffected since both the skilled and unskilled within a district face the same price change.

5.1). If skilled workers are more productive and firms adopt more skill-biased capital, then there is an increase in overall output, productivity and consumption (Section 2.1).

The first determinant of the changes in overall benefits is the reduction in costs of schooling for younger cohorts. This can have significant benefits for infra-marginal students that were always going to attend school, even if it does not induce marginal students to get more education. Studies that focus on only the enrollment response in education interventions may miss this large component of welfare that affects all infra-marginal students.

The labor market benefits depend on the increase in overall output due to skill adoption, 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 GE effects will have distributional consequences. 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.[39]

$\beta_{as,D=0}$ is the returns to education in untreated districts, and $\beta_{as,D=1}$ the returns including the general equilibrium effects. $\Delta\beta_{as}$ is thus the change in the returns due to GE effects. 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. Labor market welfare for them is $log\frac{w_{as,D=1}}{w_{as,D=0}}$. Similarly, for those 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 younger cohorts, who are induced into getting more skill, the labor market welfare change depends on the skilled wage in treated districts and the unskilled wage in untreated districts: $log\frac{w_{ys,D=1}}{w_{yu,D=0}}$. This welfare, depends on the returns $\beta_{ys,D=0}$:

$$log\frac{w_{ys,D=1}}{w_{yu,D=0}} = \quad log\frac{w_{ys,D=0}}{w_{yu,D=0}} + log\frac{w_{ys,D=1}}{w_{ys,D=0}} = \quad \beta_{ys,D=0} + log\frac{w_{ys,D=1}}{w_{ys,D=0}} \tag{24}$$

The change in labor market benefits for those induced into getting more skill therefore consists of two components – the partial equilibrium returns to skill and the GE change in economic benefits to the 'always skilled.' This is why it is important to causally estimate both the partial equilibrium returns and the GE effects. In the absence of any GE effects, the change in earnings for a person induced into getting more education would simply be $\beta_{ys,D=0}$. 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. The benefits, however, include the present discounted value of a skilled worker's earnings stream.

[39] 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 concentrates on the earnings of workers and costs of schooling.

6 Results and Discussion

The household level analysis is split up by age groups that should and should not have been directly affected by the DPEP program. In Appendix Figure A.2 one can see a sharp drop in schooling enrollment at the age of 19. By that age students have usually finished schooling, and child-labor laws prevent many workplaces from hiring children below eighteen.[40] Since the 2009 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. The results are robust to using alternative age cutoffs: I present appendix tables with multiple age groupings, widening age restrictions, and in Difference-in-Differences specifications I show impacts on each age cohort separately.[41]

For school-level outcomes I present RD figures showing the discontinuity along the running variable for the year 2005, which was the last year before the phase-out of funds begin, and the first year that the school-level data is available. The 2005 figures can be thought of as capturing the cumulative effects of the last twelve years of funding increases, and alongside I present the fuzzy RD 2SLS coefficient over time to highlight the dynamics once the funding is cut in 2006, and stopped entirely in 2007. As I use multiple data sources, at the bottom of each table and figure I mention the data source.

6.1 School Building

The primary objective of the DPEP program was to build new schools. Figure 3 shows the effect of the program on schools built once the program was underway in 1994 (top two panels), and before the program started (bottom panel). While the top panel shows the fraction of all schools built post 1993, the middle panel shows the discontinuity in government schools and private schools separately. Both panels show that DPEP districts had a substantially larger fraction of new schools than non-DPEP districts. The ITT estimates indicate a 4.9 percentage point increase in the fraction of government schools that were new.[42]

I trace out the longer terms effects by studying how the coefficient in the Figure 3 top panel changes over time. 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 districts on either side of the cutoff. While large amounts of funding were still being received by these districts in 2005, more schools were being built. However, once funding was

[40] The Factories Act of 1948 and the Mines Act of 1952.

[41] An alternative is to use lower age groups as many students graduate upper primary by age 16, and the results are robust to using lower cutoffs (Appendix Table A.2). Yet, as students may stick around and finish high school once incentivized to finish upper primary school, and because the age 19 cutoff is institutionally driven by laws, 19 is the preferred cutoff.

[42] In Appendix Figure A.3 I show alternative versions where I plot the total schools per capita, and the dynamic trends for old private schools.

whittled down so did its impact. In the absence of funds, regions on the other side of the cutoff catch up over time by building schools a relatively more rapid rate.

As a falsification test, I 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). The bottom panel of Figure 3 shows the lack of a discontinuity in older schools.

6.2 Private Schooling Response

How private schools respond to such interventions is crucial for determining the overall benefits. 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, self-segregation motives, and changes to the local economy and infrastructure with such a large-scale program. In Figure 3, there is no evidence of crowd-out, and if anything, there is mild evidence of crowd-in.

What drives the crowd in? On the one hand, the demand externality could raise the equilibrium price and draws in private schools; on the other, the fall in operating costs may induce private school entry and lower the equilibrium price. In Section 2.3.2 I discuss how we can determine which of these mechanisms is stronger by seeing how the price changes. Late, in Section 6.5 – specifically Table 7 – I show that household expenditure on schooling falls suggesting that the cost-reduction mechanism is stronger. In Section 6.7 I find that the initial increase in the supply of college educated teachers, and infrastructure upgrades may help drive cost reductions.

6.3 Education, Earnings and Gender

The DPEP program was specifically targeted towards the primary and upper primary levels, and we may expect the largest impacts at those levels. Any student who was past school going age should be unaffected by the program.

I check for a lack of a discontinuity in schooling attainment at the cutoff for older individuals, using the same RD methods. The left panel of Figure 4 shows how the older populations do not have any discontinuities in literacy, probability of finishing primary school, or upper primary school. The top panel of Table 3 shows there is no economically or statistically significant discontinuity in educational outcomes for older populations.

For the younger population in the right panel of Figure 4, we see discontinuities in different levels of education. The ITT estimates show that the young have 0.217 more years of schooling in regions that were just eligible for the program. The treatment of the treated (TOT) scaled up by probability of treatment can be found in Table A.1. In all tables, there is no statistically significant discontinuity in education for the older population. The policy, therefore, directly

affected cohorts young enough to change their schooling attainment, and had no impact on the education of older cohorts.

Table 4 looks at the fraction of people who have completed at least a given level of education. As the program was targeted to the primary and upper-primary sections, the biggest increases are seen here. Literacy rates are higher by 3 percentage points, and the likelihood of finishing primary school by 5.8 percentage points. Table 5 and Appendix Figure A.4 display the analogous results for the full sample, rather than the sub-sample of those reporting earnings.

Figure 5 and Table 3 show the RD impacts on education and log earnings for those who reported earnings, across the different bandwidth selection procedures and age groups. The ITT estimates on the years of education is 0.7 years, and on earnings 0.11 log points for younger cohorts (Table 3). Older populations have no discontinuity in education and earnings. This does not indicate that older cohorts had no GE effects, as *average* earnings conflate both the fall in skilled wages and the rise in unskilled wages.[43]

Comparing Tables 3 and 5, it is clear that, as in Duflo (2001), the impact on education is higher for the sub-sample that reported earnings. Yet, as the top panel of Appendix Table A.3 shows, there is no discontinuity in the probability that earnings are reported at the cutoff, suggesting that DPEP did not lead to differential selection into who reported their earnings.

The difference in the educational impacts between those that reported earnings and the full sample can be tied to the difference in labor market returns by gender. In the NSS, the probability that earnings are reported is uncorrelated with working in agriculture or being self-employed. 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. Since men are more likely to be in occupations that report earnings, and gain from education, while women are more likely to be engaged in domestic work, we should expect men to be more responsive to these interventions (Dreze and Sen, 2002; Kingdon, 1998).

Indeed, in Appendix Tables A.4 and A.5, I find that the effects are concentrated among males, which is similar to most of the related literature (Ashraf et al., 2015; Breierova and Duflo, 2003), and other work on this program Jalan and Glinskaya (2013). In the full sample, men increase their years of education by about 0.3 years, whereas women increase theirs only by about 0.09 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.

[43] I find some negative effects on average wages for closest substitutes: cohorts that were close to the treated cohorts. In Appendix Table A.2, the sample is sliced thin into more age groups, and even though they are imprecisely estimated, there do seem to be negative earnings effects on the 36 to 45-year-olds which is the closest age group to the treated cohorts.

6.3.1 Alternative Specifications and Robustness

I perform a difference-in-differences (DID) analysis that estimates a different parameter – the average effect among all treated districts. In Appendix Table A.6, 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 0.39 years, and for the sub-sample of those who reported earnings education is higher by 0.46 years and log earnings increase by 0.065. These TOT education increases are smaller than the RD estimates, suggesting that districts further away from the cutoff did a relatively poor job of implementing the program.

In the appendix I conduct a number of robustness checks. I collapse all the household data into district-age cells, and re-run the regressions. Even as collapsing the data loses valuable information used in estimating the optimal bandwidth, the results do not change, as can be seen in Appendix Table A.7. I try more in-progress RD bandwidth selection procedures and standard error estimation methods in Table A.8, including methods that allow for different bandwidths on either side of the cutoff, and nearest-neighbor variance estimation at district clusters. I test the sensitivity to age cutoffs by binning age groups into finer categories in Table A.2, and to age restrictions by including a larger sample of ages in Table A.9.

6.4 Returns to Education

6.4.1 Conventional Methods

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, 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 returns into getting more education.[44] Another possibility is that OLS estimates suffer from measurement error (Griliches, 1977).

The canonical IV-Wald method to estimate the returns to education is to simply use the RD cutoff to first estimate the change in the years of education 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. Yet, as my model stresses, the policy should simultaneously affect both the skill premium and the overall output in the district. Since the change in the

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

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 individual returns would be conflated with the changes in output and the skill premium.

The estimates in Table 3 can be used to calculate the returns using the conventional method of taking the ratio of the change in log earnings and the change in years of education. The ratio of 0.112 log earnings and 0.72 years gives us a return of about 15.5%. The bottom panel of Table 3 shows the 2SLS-LATE version of this exercise. However, due to the size of the confidence intervals, this number is not statistically indistinguishable from numbers as low as 7% nor the OLS Mincerian return of 10% estimated in this sample. This estimate, therefore, lies reasonably within the range of comparable estimates found in the literature (Banerjee and Duflo, 2005; Psacharopoulos and Patrinos, 2004). One of the most recent experimental estimates of returns in a developing country is 13% (Duflo et al., 2017).[45]

Another IV method is to use a difference-in-differences (DID) strategy. In Appendix Table A.10 and Figure A.5, I compare DPEP districts to non-DPEP districts, and the older cohorts to the younger cohorts.[46] 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 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 A.10).

The 2SLS return of 15.5% using conventional IV methods, however, is neither the partial

[45] 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%.

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

$$y_{ida} = \beta_{DiD} T_{da} + \mu_d + \varpi_a + \epsilon_{ida} \ , \tag{25}$$

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.

equilibrium returns to education, nor the returns with the entire GE effects. As I elaborate upon below, it is merely a weighted average of both and lies in between them.

6.4.2 Returns to Education and the Labor Market GE Effects

The model allows me to estimate meaningful equations to calculate the GE effects, and highlights an important point: the conventional method 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 GE effects in the local economy. From Equation (20) we know:

$$log\frac{w_{y,D=1}}{w_{y,D=0}} = \ell_{sy,D=1}\underbrace{log\frac{w_{sy,D=1}}{w_{sy,D=0}}}_{\neq 0} + \ell_{uy,D=1}\underbrace{log\frac{w_{uy,D=1}}{w_{uy,D=0}}}_{\neq 0} + \Delta\ell_{sy}\underbrace{log\frac{w_{sy,D=0}}{w_{uy,D=0}}}_{\beta_{ys,D=0}} \tag{20}$$

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 for the compliers $\Delta\ell_{sy}$, since under these assumptions $\beta_{ys,D=0} = log\frac{w_{y,D=1}}{w_{y,D=0}}/\Delta\ell_{sy}$.

Sometimes referred to as the LATE theorem, this is often used to estimate the 2SLS-Wald returns to education. Yet, when there are GE effects $log\frac{w_{uy,D=1}}{w_{uy,D=0}} \neq 0$ and $log\frac{w_{sy,D=1}}{w_{sy,D=0}} \neq 0$, and these confound the estimates. $log\frac{w_{uy,D=1}}{w_{uy,D=0}}$ and $log\frac{w_{sy,D=1}}{w_{sy,D=0}}$, however, are measurable quantities, and so all components of Equation (20) are estimable, allowing me to recover $\beta_{ys,D=0}$. Ignoring the GE effects produces an estimate that lies between the partial and general equilibrium returns to education.

The 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 adoption of skill-biased capital affects the skilled and unskilled differently, as captured by Equation (17). 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 GE effects it is necessary to use the method outlined in Section 5.1.1 and specifically, Equations (20) and (21) to derive the returns to education with and without the labor market general equilibrium effects. Table 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 targeted getting students through this level, and because the largest earnings increase in OLS regressions on untreated districts comes when a student finishes upper primary school.[47] After scaling up by the treatment probability, there was a 17 percentage point increase in skilled workers across the cutoff (see middle panel of Table A.1).

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

As the returns depend on the shares of skilled and unskilled, as in Equation (20), I bootstrap one thousand draws with replacement, creating a null by jointly permutating the RD running variable, treatment status and probability of treatment. The estimated returns to shifting into the skilled group in the absence of GE effects are 19.9% per year. 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. In order to do this, I use Equations (18) and (19) discussed in the identification section. Table 6 implies 91.87% of the 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 as the two components that determine this effect may counteract each other – an increase in the relative supply of skilled workers $\left(log\frac{L_{s,D=1}}{L_{u,D=1}}\right)$ will lower the skill premium, but adoption of skill biased capital $\left(log\frac{\theta_{s,D=1}}{\theta_{u,D=1}}\right)$ may increase this skill premium.

Furthermore, the additional impacts on the young term is high implying that the young and the old are not close substitutes in production. This indicates that looking only at the GE effects on older cohorts grossly understates the GE effects on younger cohorts, and my paper is among the first to estimate the GE effects on all cohorts.

Even as the specifics of the model do not determine the GE effects, the advantage is that it allows one to speak to relatable elasticities. The elasticity of substitution across age groups $\sigma_A = 5$ is similar to Card and Lemieux (2001). In the absence of the adoption of skill-biased capital, the elasticity of substitutions across education groups would be $\sigma_E = 4.24$, yet the differential adoption of skill-biased capital inflates this figure.

6.5 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 is adopted by the region, overall productivity and output in the region may increase. In the top panel of Appendix Figure A.7, one can see the impact on total output (the District Domestic Product). These regressions are underpowered, and the standard errors are quite large. The (imprecisely estimated) point estimates indicate that between 2000 and 2004, the increase in GDP associated with the policy was between 0.068 and 0.11 log points (Appendix Table A.11).

The change in total output will lead to a change in total consumption. In the top panel of Table 7, I show that the change in consumption expenditure in the last year of the policy (2004-5) was about 0.06 log points. At the same time, in 2004 the money spent for educational purposes

(tuition, fees, books and stationery) falls by between 0.085 and 0.21 log points.[48] 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 7), since expenditures on books and stationery can rise when households gain more education – these are complements in consumption. In the bottom panel of Appendix Figure A.7 one can see discontinuously lower education expenditures and tuition.

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 households with 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 for the infra-marginal students even as they do not change the years of education.

6.6 Migration, Productivity, and the Adoption of Capital

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, and this migration would dampen any GE effects that depend on the the skill distribution. Second, existing firms may have switched the composition of their workforce by hiring more skilled 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 adoption of skill-biased capital, therefore, will increase the returns to skill in treated districts and be a crucial determinant of the GE effects.[49]

Importantly, my analysis explicitly allows for migration to change the quantities of labor in Equation (17). Worker mobility will tend to equalize earnings across regions and mitigate any negative GE effects on skilled wages or positive GE effects on unskilled wages. Therefore, 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. Permanent worker migration is extremely low in the Indian context (DasGupta, 1987; Munshi and Rosenzweig, 2009; Topalova, 2010).[50]

[48] Jalan and Glinskaya (2013) measure a 20-40% fall in household educational expenditure.

[49] 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.

[50] Many studies on India are explicit about ignoring migration in the main analysis as the numbers are low (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

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. Panel B of Appendix Table A.3 shows that the policy did not impact migration. There are no economically meaningful changes to either any migration or migration specifically for work-related reasons.[51]

On the other hand, firm capital is 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 A.6, where one can see that even at the manufacturing establishment level, the average wage paid to workers increases as more and more educated workers start joining the labor market 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 findings are suggestive evidence in support of the adoption of skill-biased capital into these regions.

One relevant question is whether this capital was previously being utilized in other forms or is flowing from other regions, 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. Policy regions are geographically dispersed all over the country (Figure A.1) rather than being neighbors of districts just on the other side of the cutoff. In the bottom panel of Figure A.6, I look at the density of capital-intensive firms in the early period and 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, supporting anecdotal evidence of people residing in major cities investing in villages that they originate from.

In general, there are some clear changes to the labor market for the workers in these regions. The bottom half of Appendix Table A.3 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, possible in this context (Foster and Rosenzweig, 1996).

(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.

[51] RD estimates by finer skill groups or age cohorts are not possible as almost nobody is migrating in the data.

6.7 School Quality, Teachers, Infrastructure, and Other Grants

While the primary focus of the program was to increase educational attainment by building schools, there may have been improvements in quality given such a large amount of funding. Such improvements may have increased the returns to schooling in the labor market, attenuating the negative GE effects on returns. In Appendix Table A.12 I use 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 statistically 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 better test scores that may attenuate the estimated negative 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.

In 2005, when program funding was still high, the number of college-educated teachers in DPEP districts was higher. However, once the funding is no longer targeted to DPEP districts, this discontinuity dissipated over time (top panel of Appendix Figure A.8). A lack of targeted funds may have led to a relative slowdown in the hiring of teachers.[52] Tangible infrastructure in schools seems to last even when the DPEP funding is reduced. Drinking water and electricity are consistently higher in regions that received the DPEP (Appendix Figure and A.9). Other inputs, such as medical checkups, are also consistently higher for DPEP districts.

The condition of the classrooms deteriorated over time once funding was stopped. While, in 2005, schools in DPEP regions had a lower number of classrooms needing repair, over time more of these classrooms broke down (Appendix Figure A.8). These results indicate that a constant source of funding may be needed to keep rooms in good condition. In Appendix E, I discuss other changes like the crowd out of other funds, the construction of pre-primary sections, the establishment of resource centers, school inspections, and the medium of instruction.

6.8 Overall Economic Benefits

Increases in overall output and reductions in the total cost of schooling will benefit households. The change in labor market earnings depend on the returns to skill and the GE effects on these returns. Table 6 shows the returns by skill group, which helps back out the parameters and the changes in yearly labor market benefits shown in Table 8. These estimates depend on the TOT effects on earnings, scaled up by the probability of treatment. For these calculations, the average real interest rate during that period was used (5%). A gap of 10 years is assumed between the

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

time the costs of education are borne and the labor market returns are realized.[53]

In the top panel of Table 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.[54]

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 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 from the reductions in the total costs of schooling. Even young students who were always going to attend school even in the absence of the program benefit from the reductions in schooling costs.[55]

7 Conclusion

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 an RD approach. The estimates imply that the return to acquiring

[53]The average real interest rate comes from the World Bank WDI. 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.

[54]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.

[55]These results do not necessarily indicate that the policy was cost effective. I have shown that the direct 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 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.

skill is 13.4%, but that it is 6.5 percentage points lower than it would be in the absence of general equilibrium effects. These changes imply elasticities of substitution across skill groups and cohorts that are in line with previous literature in other contexts (Card and Lemieux, 2001). 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 the local economy output.

These findings have two important implications. First, we can apply this methodology when discussing the benefits demonstrated by small-scale interventions, as scaled up versions of such interventions may have GE effects. Second, it speaks to the large body of literature that uses large-scale variation to estimate the returns to education (such as tuition subsidies, compulsory schooling laws, distance based measures, and school building). Macro-level variation estimates a different parameter and may conflate the individual returns and general equilibrium effects. This is because an experiment where a single individual receives more education is inherently different from the variation induced by changes that affect entire cohorts of students.

The methodology I develop can be applied to other similar settings across the world. My estimates, however, are about local labor markets and not for the entire country. Furthermore, they are not generalizable to regions further away from the RD cutoff in the absence of stronger assumptions. Indeed, as my difference-in-differences results indicate, the policy was poorly implemented at the districts with the lowest literacy rates. Yet, the great advantage of the RD cutoff is that it was for districts around the national average of female literacy; therefore, we should think of these results as pertaining to the average district.

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 importantly 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 instance, 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; Deaton, 2010). The methods in this paper can be used to quantify the expected impacts of scaled-up micro-interventions. 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.

References

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.

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.

Ashraf, N., Bau, N., Nunn, N., and Voena, A. (2015). Bride Price and the Returns to Education for Women. *Mimeo*, Harvard.

Ayyar, R. (2008). Country Agency Relationship in Development Cooperation: An Indian Experience. *Global Monitoring Report*, UNESCO.

Azam, M. and Saing, C. H. (2016). Assessing the Impact of District Primary Education Program in India. *Review of Development Economics*, (1467-9361).

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., Iyer, L., and Somanathan, R. (2008). Public Action for Public Goods. *Handbook of Development Economics*, 4:3117–3154.

Bartalotti, O. and Brummet, Q. (2017). Regression Discontinuity Designs with Clustered Data. *Regression Discontinuity Designs (Advances in Econometrics)*. forthcoming.

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.*

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.

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., and Sandefur, J. (2013). 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.

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., Farrell, M., and Titiunik, R. (2017). Software for Regression Discontinuity Designs. *Stata Journal*, 17(2):372–404.

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.

Cannonier, C. and Mocan, N. (2012). Empowering Women Through Education: Evidence from Sierra Leone,. *NBER Working Papers 18016*.

Card, D. (1999). The Causal Effect of Education on Earnings. *Handbook of Labor Economics*, 5:1801–1863.

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.

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.

Chin, A. (2005). Can Redistributing Teachers Across Schools Raise Educational Attainment? Evidence from Operation Blackboard in India. *Journal of Development Economics*, 78(2):384–405.

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., Holla, A., Kremer, M., Mohpal, A., and Muralidharan, K. (2013). 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.

Deaton, A. (2010). Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, 48:425–455.

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.

Deshingkar, P. and Anderson, E. (2004). People on the Move: New Policy Challenges for Increasingly Mobile Populations. *Natural Resources Perspectives*, 92. Overseas Development Institute.

Dreze, J. and Sen, A. (2002). India. Development and Participation. *Oxford University Press*. 2nd Edition.

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., Dupas, P., and Kremer, M. (2017). The Impact of Free Secondary Education: Experimental Evidence from Ghana. *Working paper*.

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.

Epple, D. and Ferreyra, M. (2008). School Finance Reform: Assessing General Equilibrium Effects. *Journal of Public Economics*, 92:1326–1351.

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.

Ghani, E., Goswami, A. G., and Kerr, W. R. (2015). Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing. *Economic Journal*, doi:10.1111/ecoj.12207.

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.

Griliches, Z. (1977). Estimating the Returns to Schooling: Some Econometric Problems. *Econometrica*, 45:1–22.

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.

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. 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.

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.

Imbens, G. W. and Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2):615–635.

Jacob, V., Kochar, A., and Reddy, S. (2008). School Size and Schooling Inequalities. *Mimeo*, Stanford.

Jagnani, M. and Khanna, G. (2018). The Effects of Elite Public Colleges on Primary and Secondary Schooling Markets in 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.

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.

Kingdon, G. (1998). Does the Labour Market Explain Lower Female Schooling in India. *Journal of Development Studies*, 35(1).

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.

Kremer, M. and Muralidharan, K. (2007). Public and Private Schools in Rural India. *School Choice International*, MIT Press.

Lee, D. (2005). An Estimable Dynamic General Equilibrium Model of Work, Schooling, and Occupational Choice. *International Economic Review*, 46(1):1–34.

Linden, L. (2008). Complement or Substitute? The Effect of Technology on Student Achievement in India. *InfoDev*, Columbia University, MIT Jameel Poverty Action Lab, IZA.

Lucas, A. and Mbiti, I. (2012). Access, sorting, and achievement: The short-run effects of free primary education in kenya. *American Economic Journal: Applied Economics*, 4(4):226–53.

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.

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.

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.

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.

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).

Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to Investment in Education: A Further Update. *Education Economics*, 12(2):111–134.

Roy, A. (1950). The Distribution of Earnings and of Individual Output. *Economic Journal*, 60(239):489–505.

Sen, A. (1999). *Development as Freedom.* Random House Inc., United States of America.

Smith, A. (1775). *An Inquiry into the Nature and Causes of the Wealth of Nations.* The Glasgow Edition, pp 29.

Topalova, P. (2010). Factor Immobility and Regional Impact of Trade Liberalization: Evidence on Poverty and Inequality from India. *American Economic Journal: Applied Economics*, MIT.

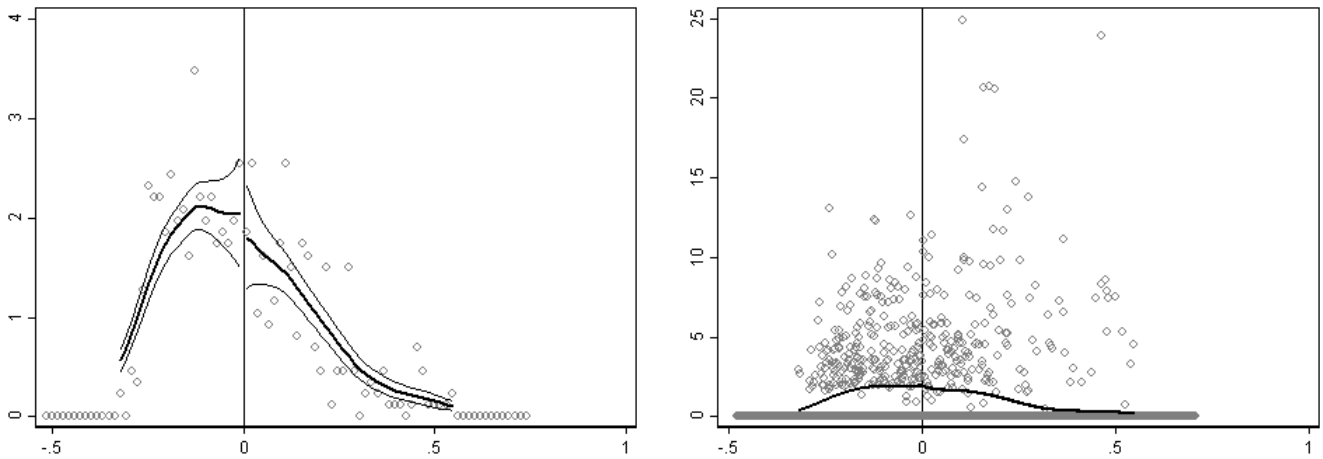
Varghese, N. (1994). District Primary Education Programme: the Logic and the Logistics. *Journal of Educational Planning and Administration*, 2(1):69–106.

Willis, R. J. (1986). Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions. *Handbook of Labor Economics*, 1:525–602.

World Bank (1997). *Primary Education in India.* The World Bank, Washington DC.

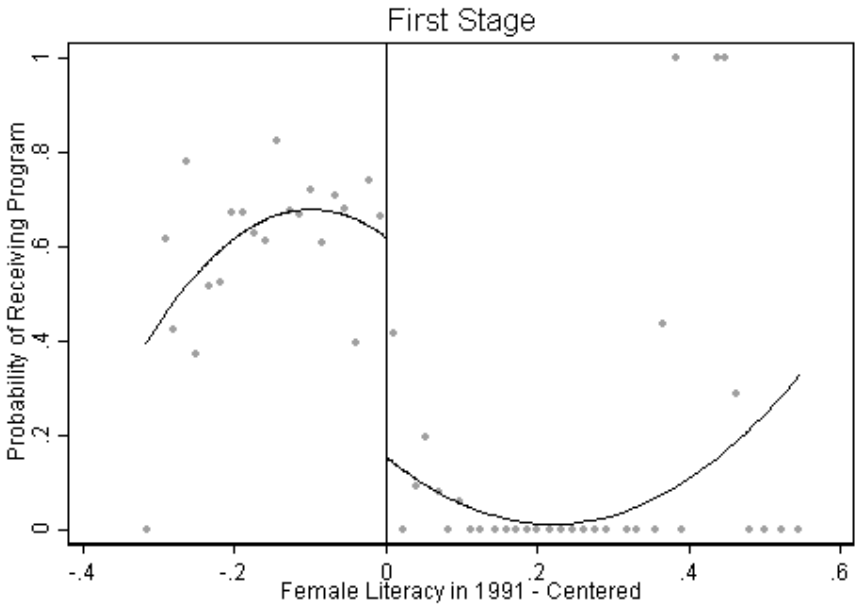
I Figures

Figure 1: McCrary (2008) Density Tests



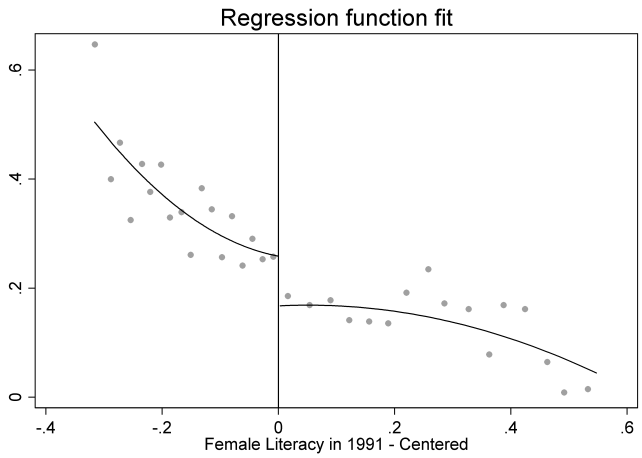
McCrary (2008) tests for discontinuity in density at the cutoff. These tests look for evidence of one-sided manipulation of the running variable by testing the discontinuity in the density at the RD cutoff.

Figure 2: First Stage of DPEP

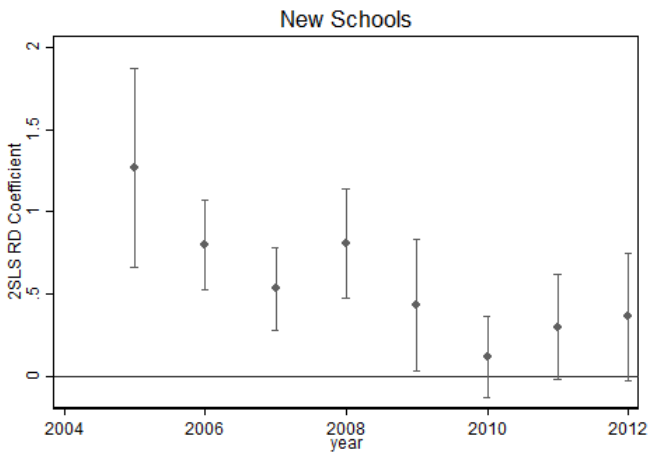


First stage graphs showing probability that a district received DPEP funds. Optimal bin sizes calculated using Calonico et al. (2014b) method.

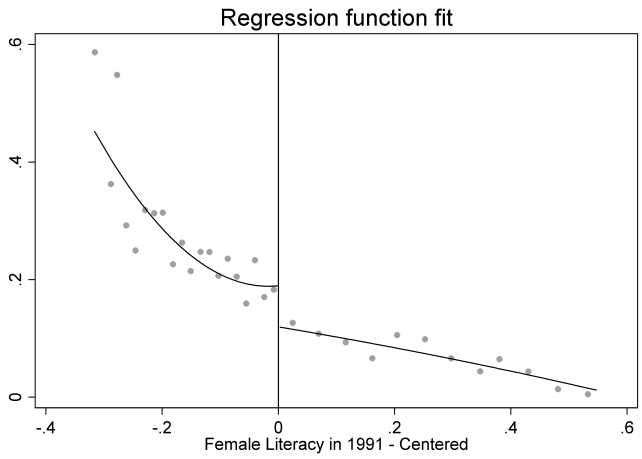
Figure 3: School Building Before and After 1993



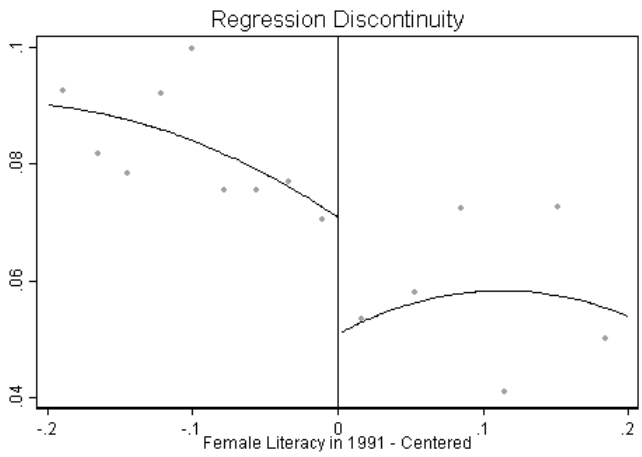
Fraction of All Schools Built Post 1993



Frac New Schools (2SLS) Over Time

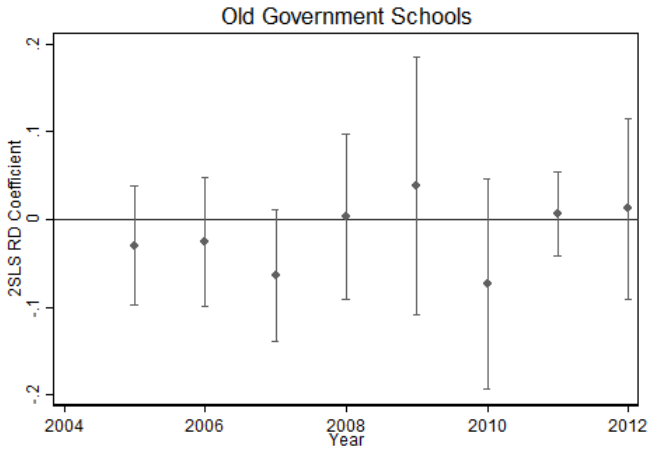


Fraction of Government schools Built Post 1993



Fraction of Private schools built post 1993

Total Number of Old Gov Schools (pre 1993)



Fraction Old Government Schools (1973-93)

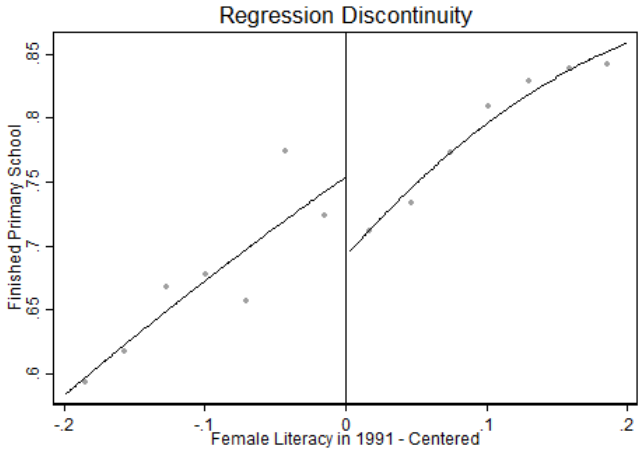
Source: DISE (District Information System for Education) data. Scatter plots use the 2005 data. 'New schools' are schools built post 1993. 'Old schools' are schools built between 1973-93. RD graph optimal binning and 2SLS (scaled up by the probability of treatment) RD coefficients calculated using Calonico et al. (2014b) procedure.

Figure 4: RD figures - Levels of Education

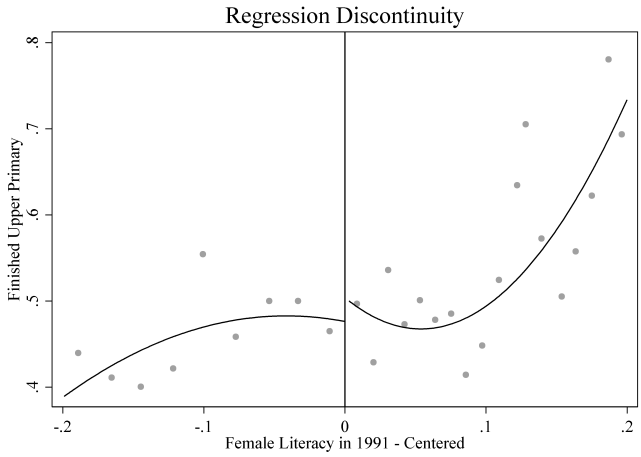
Literate - Older (Placebo)

Literate - Younger

Finished Primary - Older (Placebo)



Finished Primary - Younger



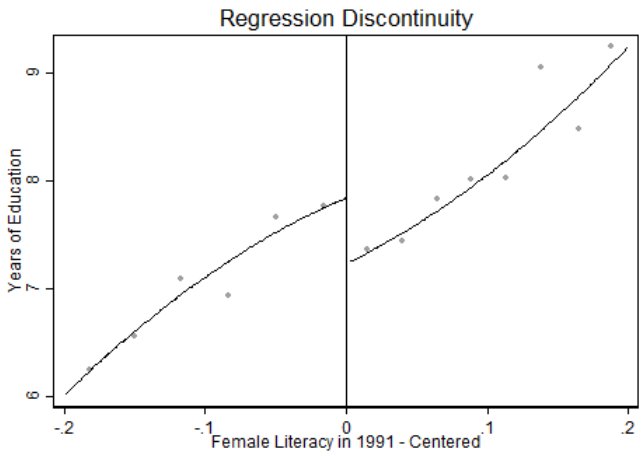
Upper Primary - Older (Placebo)

Upper Primary - Younger

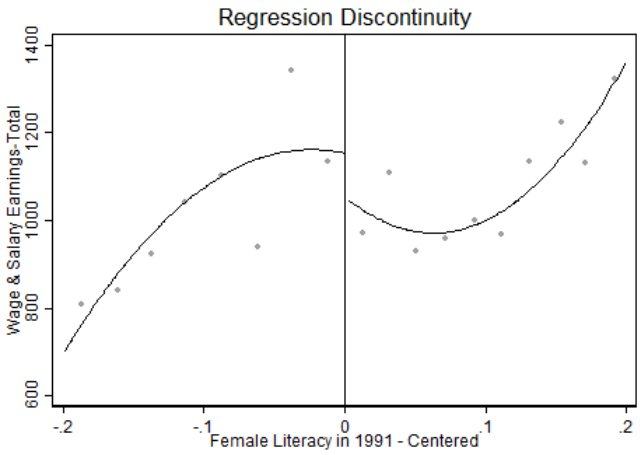
National Sample Survey 2009 for persons who report earnings in primary occupation. Appendix Figure A.4 shows the analogous graphs for the full sample of persons. Figures made using Calonico et al. (2014b) method of using regression curves to approximate the conditional means on either side of the cutoffs and the equally spaced sample means, and optimally spaced bins.

Figure 5: RD Years of Education and Earnings

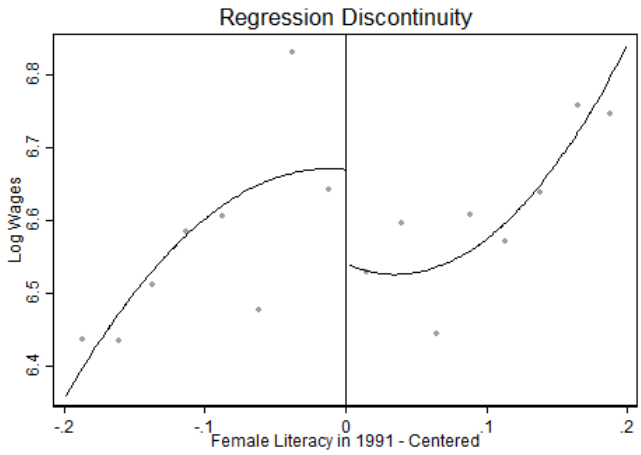
Years of Education - Older (Placebo)



Years of Education - Younger



Earnings (Rupees) - Younger



Log(Earnings) - Younger

National Sample Survey 2009 for those who reported earnings. 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.

II Tables

Table 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 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 3: Education and Earnings for those with Reported Earnings

Panel A: Reduced Form Years of Education	Young	Old	Young	Old
RD Estimate	0.720	-0.0856	0.698	0.100
	(0.199)***	(0.218)	(0.173)***	(0.188)
Observations	10,175	11,293	14,277	16,007
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log Earnings	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.112	-0.0114	0.145	0.0432
	(0.0312)***	(0.0372)	(0.0269)***	(0.0318)
Observations	10,175	11,293	14,277	16,007
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Panel B: 2SLS IV-LATE Conventional Method Returns **Log Earnings**	**Young**	**Old**	**Young**	**Old**
Years of Education	0.155	0.129	0.208	0.442
	(0.0427)***	(0.303)	(0.0460)***	(0.666)
Observations	10,175	7,994	14,277	8,627
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 'Young' are of school going age during the policy, whereas those 'Old' 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. Panel A report Intent to Treat (ITT) Effects.
Panel B shows 2SLS regressions which treats the first stage as 'change in years of education'. This is the ratio of the top two panels and is similar to conventional IV-LATE methods of computing the returns to education.

Table 4: Fraction of People that Have Finished At Least a Given Level of Education

Literate	Young	Old	Young	Old
RD Estimate	0.0328	-0.0121	0.0291	0.00711
	(0.0143)**	(0.0169)	(0.0124)**	(0.0146)
Observations	9,003	7,413	14,277	11,088
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Primary	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.0588	-0.0117	0.0574	0.00401
	(0.0174)***	(0.0180)	(0.0150)***	(0.0154)
Observations	9,273	7,869	11,972	9,920
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper-primary	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.0743	-0.0169	0.0733	0.000212
	(0.0197)***	(0.0184)	(0.0169)***	(0.0158)
Observations	9,045	7,729	10,175	9,920
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 'Young' are of school going age during the policy, whereas those 'Old' 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. Coefficients measure the change in the dependent variable on crossing the RD cutoff.

Table 5: Education Changes - Full Sample

Years of Education	Young	Old	Young	Old
RD Estimate	0.217	0.122	0.244	0.167
	(0.0891)**	(0.120)	(0.0767)***	(0.114)
Observations	45,208	51,037	45,208	51,037
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper-primary	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.0266	0.0107	0.0313	0.0166
	(0.00893)***	(0.0112)	(0.00762)***	(0.0107)
Observations	45,208	51,037	45,208	51,037
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). Coefficients measure the change in the dependent variable on crossing the RD cutoff. The sample of 'Young' are of school going age during the policy, whereas those 'Old' 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.

Table 6: Returns, and Wage Parameters

	Fraction Switched	Change in Returns $\Delta\beta$	
Estimate	0.171	-0.065	
SE	(0.045)	(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
$\Delta\beta$	-0.0053	-0.0594	91.87%

National Sample Survey 2009-10. The estimation follows the procedures described in the Model section 2, and detailed in Appendix B.V, specifically Equations (17), (20) and (21).
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 adoption of additional skill-biased capital the elasticity of substitution across skill groups would be $\sigma_E = 4.24$.

Table 7: Household Expenditures

	Log(Consumption Expenditure)			
	2004-5		2009-10	
RD Estimate	0.0664	0.0659	0.0589	0.0575
	(0.0125)***	(0.0120)***	(0.0177)***	(0.0171)***
Observations	27,372	33,758	12,563	26,420
Bandwidth selection procedure	CCT	I and K	CCT	I and K
	Log(Total Educational Expenditure)			
	2004-5		2009-10	
RD Estimate	-0.0857	-0.216	-0.0296	-0.0333
	(0.0581)	(0.0492)***	(0.0538)	(0.0545)
Observations	8,922	11,388	8,205	9,937
Bandwidth selection procedure	CCT	I and K	CCT	I and K
	Log(School Fees and Tutoring)			
	2004-5		2009-10	
RD Estimate	-0.205	-0.389	-0.215	-0.229
	(0.0806)**	(0.0679)***	(0.0927)**	(0.0722)***
Observations	8,308	12,034	7,608	10,219
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	0.0675	-0.0250	0.175	0.197
	(0.0550)	(0.0416)	(0.0379)***	(0.0311)***
Observations	8,783	14,068	12,614	14,207
Bandwidth selection procedure	CCT	I and K	CCT	I and K

Household Expenditure Sources: National Sample Survey 2004-5 and 2009-10. Bandwidths: 'CCT' is the Calonico et al. (2014b) method. 'I and K' is the Imbens and Kalyanaraman (2012) method. Coefficients measure the change in the dependent variable on crossing the RD cutoff.

Table 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

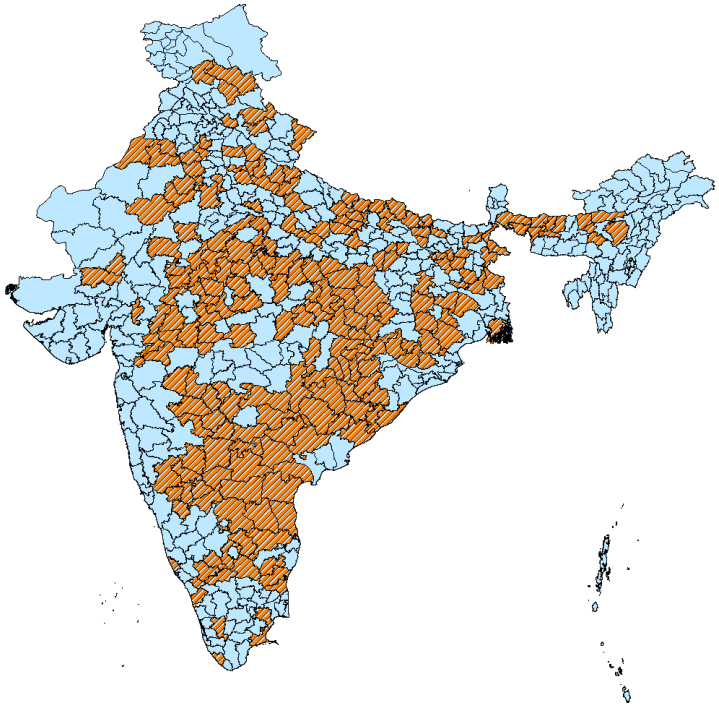
Costs	Benefits	Net	% Change (due to GE)
5.153	6.596	1.443	-23.3%

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%.

ONLINE APPENDIX

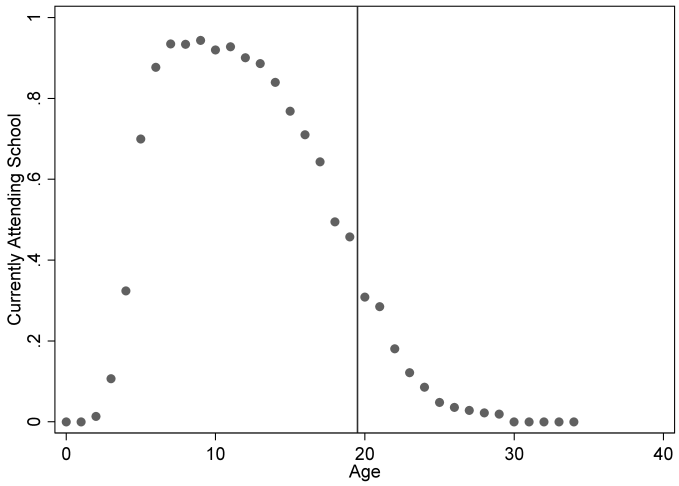
A Additional Tables and Figures

Figure A.1: Map of DPEP Districts



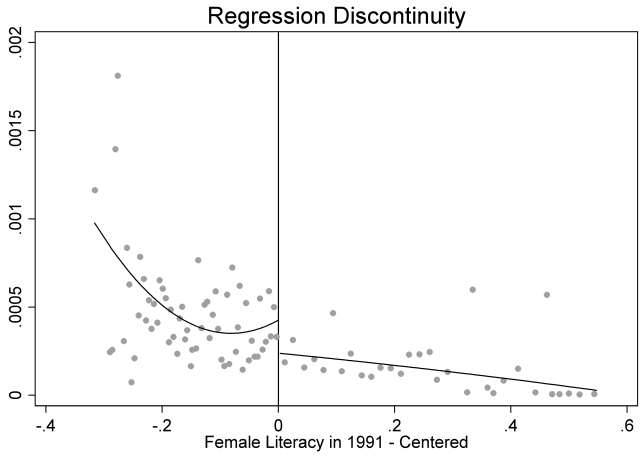
Orange and shaded districts received DPEP, whereas blue-unshaded districts did not.

Figure A.2: Enrollment Rates by Age

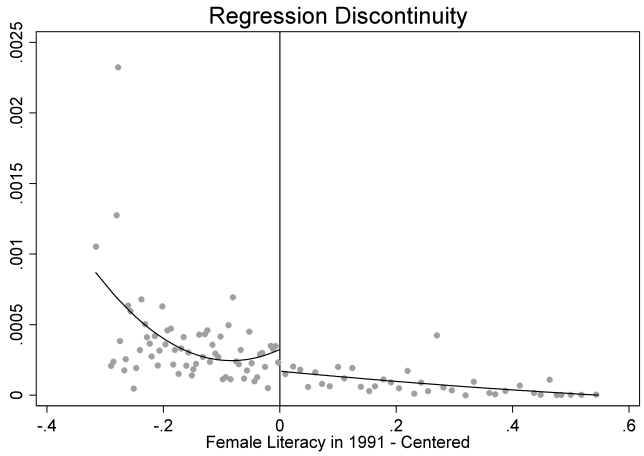


National Sample Survey 2009. The largest drop in school enrollment occurs between the ages of 19 and 20 - representing a 15 percentage point fall.

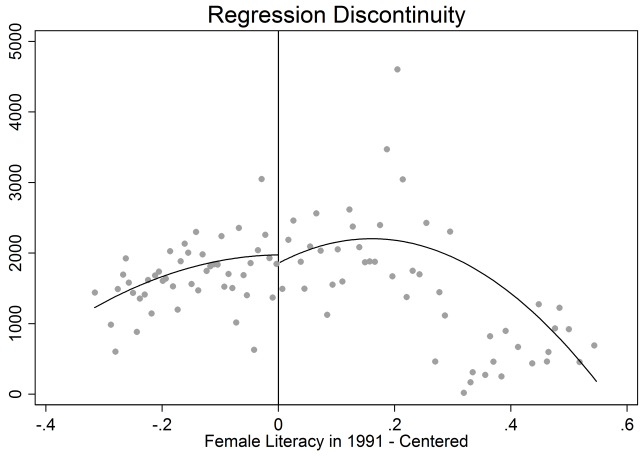
Figure A.3: Old and New Schools



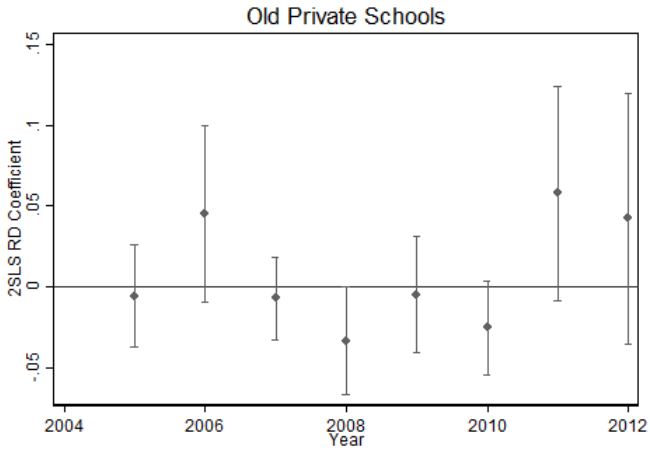
Total Schools (per cap) Built Post 1993



Total Government Schools (per cap) post 1993



Total Number of Old Schools (built pre-1993)

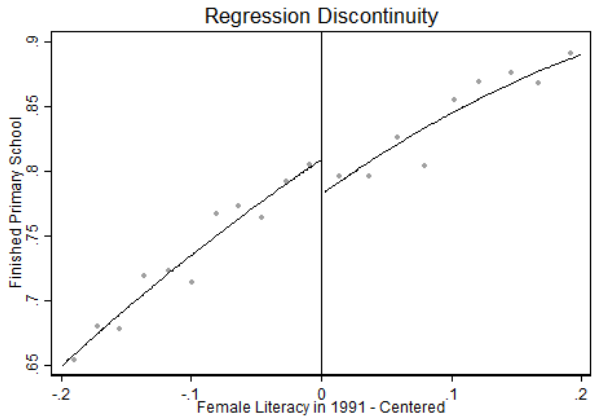


Frac of Private Schools that are 'old' (1973-93)

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. 'per cap' figures normalized by total population in district.

Figure A.4: RD Figures - Levels of Education - Full Sample

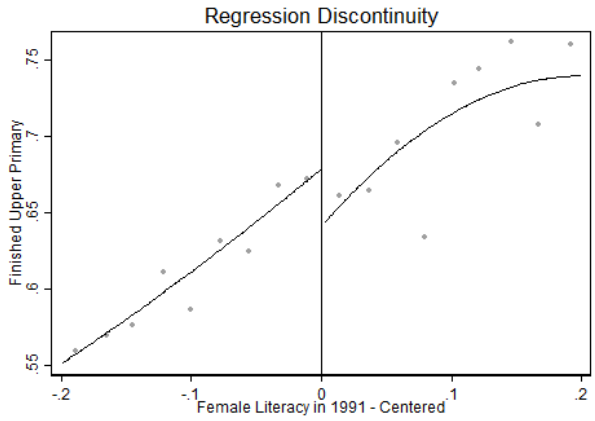
Finished Primary School - Old



Finished Primary - Young

.

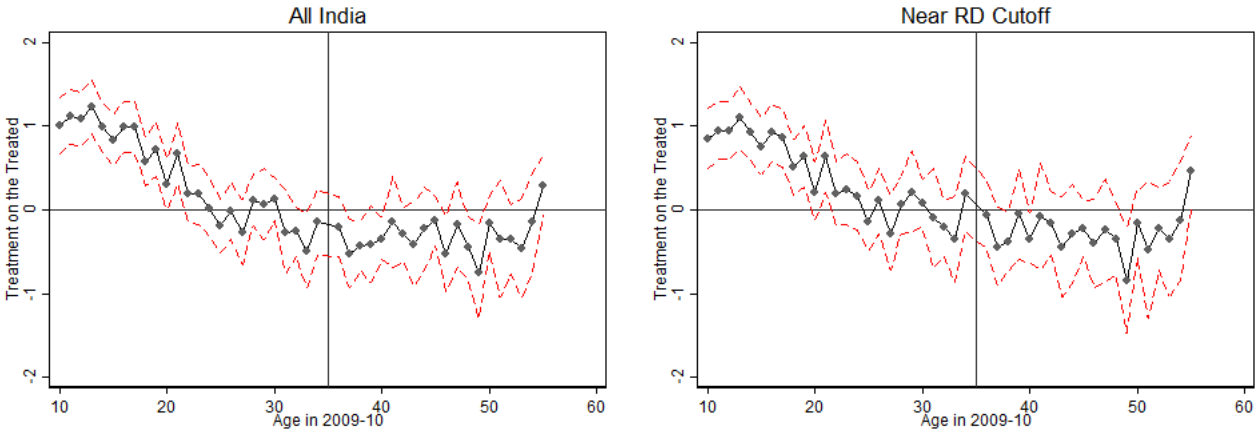
Years of Education - Old



Finished Upper Primary - Young

National Sample Survey 2009 for all 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 optimal number of bins.

Figure A.5: Difference-in-Differences: Years of Education

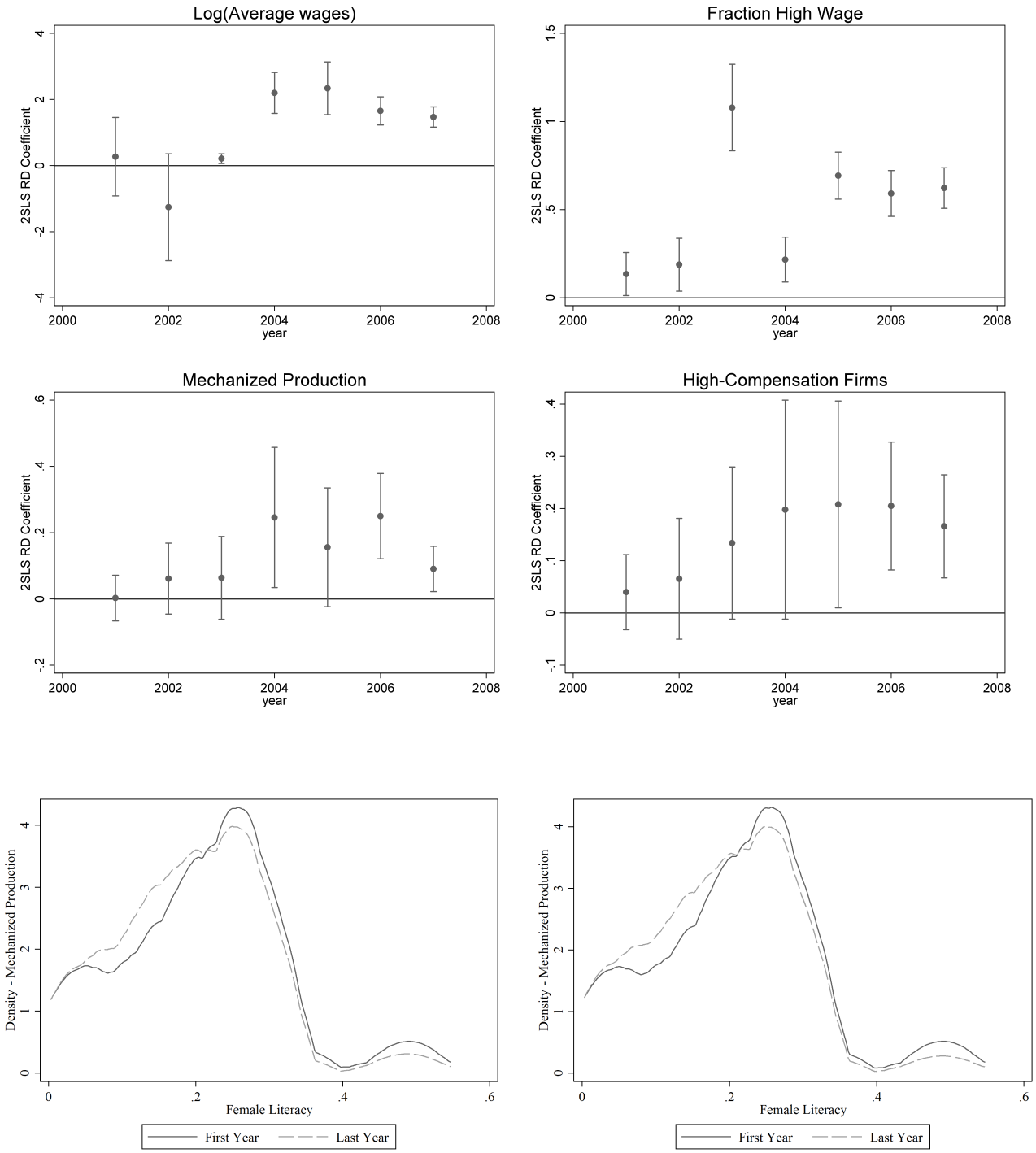


Full Sample

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 A.6: Adoption of Skill Biased Capital: Firm-Level Data

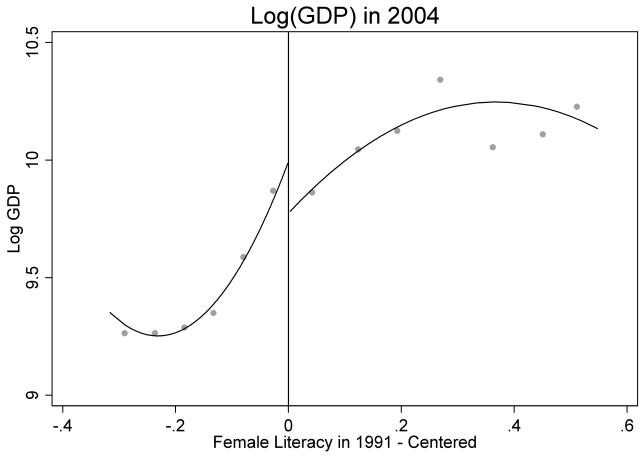


Density Above Cutoff: Mechanized Production

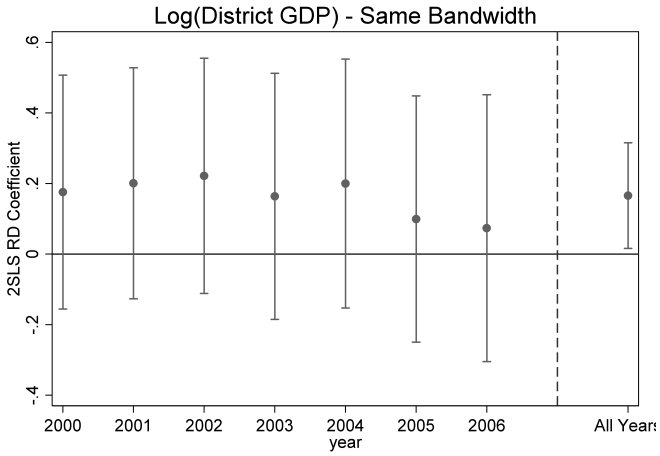
Density Above Cutoff: High Compensation

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. In the bottom panel, 2001 is the first year of data and 2007 is the last year of data.

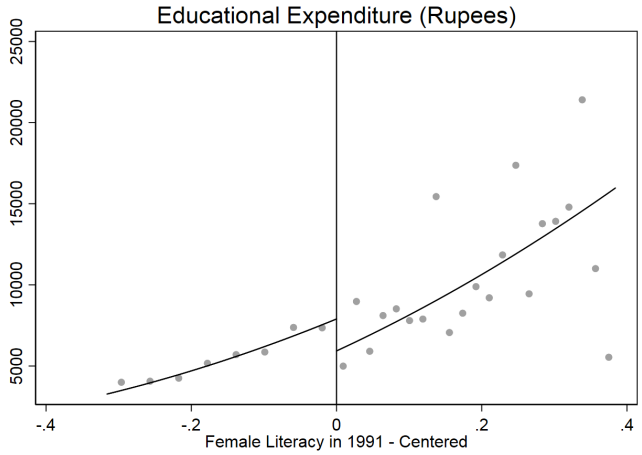
Figure A.7: Change in Overall Output and Household Expenditure on Education



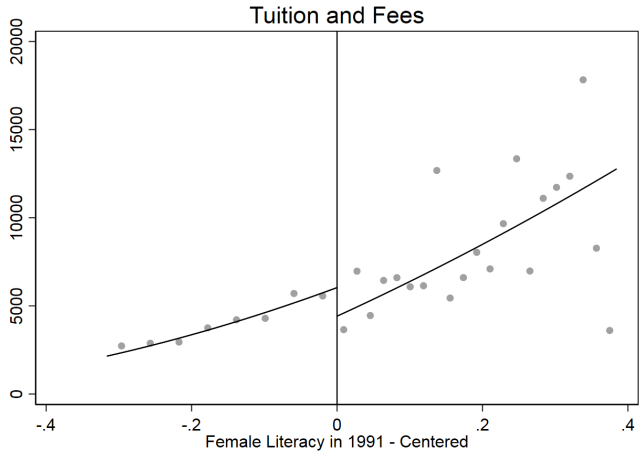
Log District GDP in 2004



RD 2SLS Coefficients - Same Bandwidth



Total Educational Expenditure

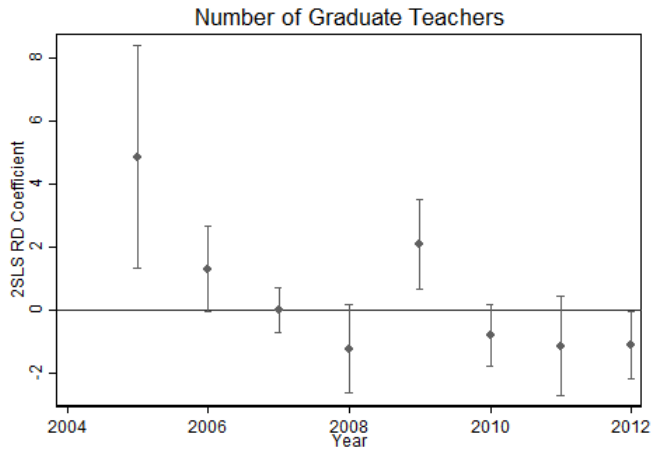


Expenditure on Tuition and Fees

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.

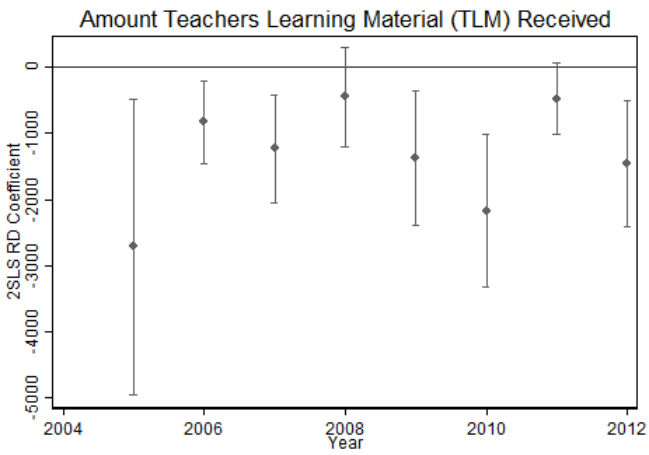
Figure A.8: Crowd-out of Grants, Classroom Repair, and Teachers

Teachers (per school) with College Degrees



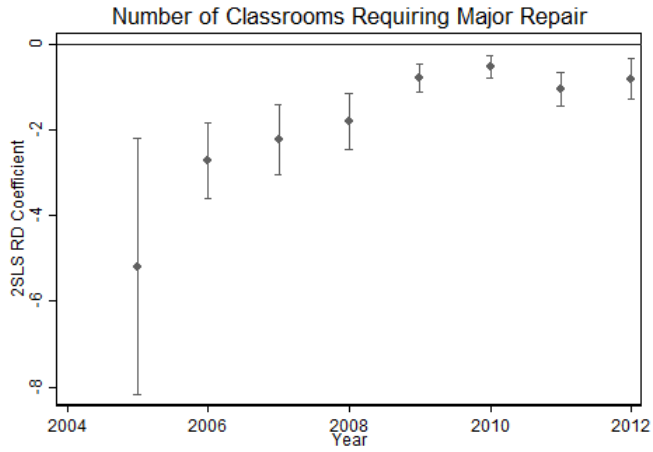
Teachers (per school) with College Degrees

TLM grants Spent (2005)



RD Coefficient Over Time: TLM grants Received

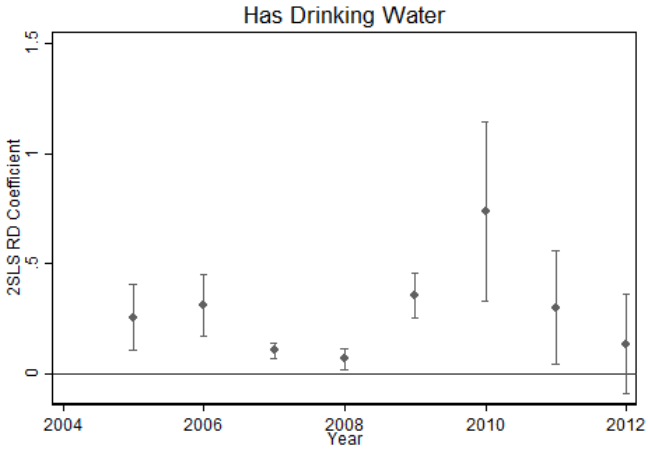
Classrooms Needing Major Repair (2005)



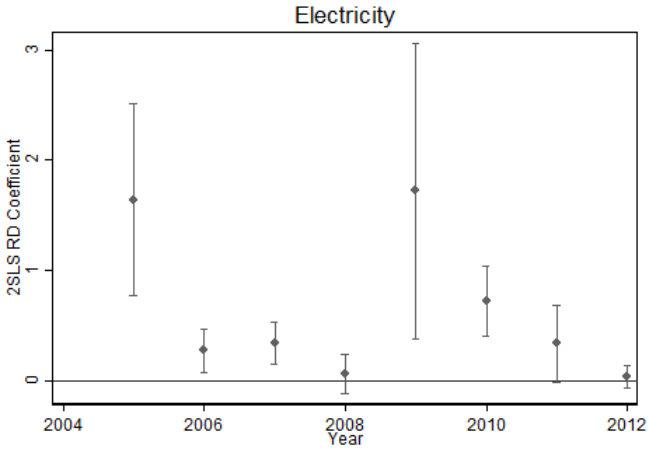
RD Coefficients: Classrooms Need Major Repair

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. All schools, regardless of their which district they are in, are eligible to receive the Teacher Learning Materials (TLM) grant.

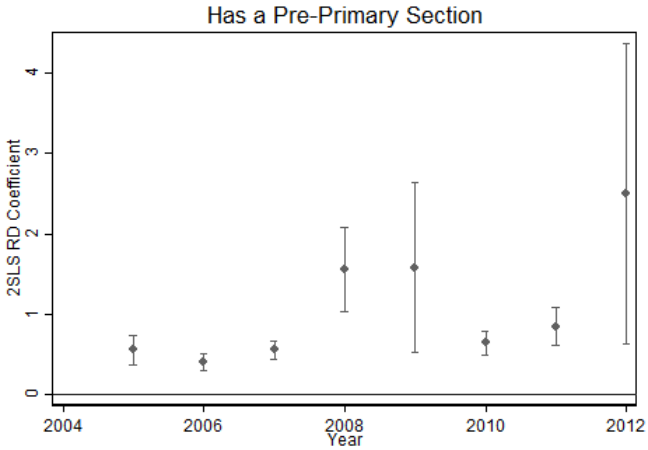
Figure A.9: Infrastructure, Distance to Resource Centers and Pre-Primary Sections, Medical Checkups



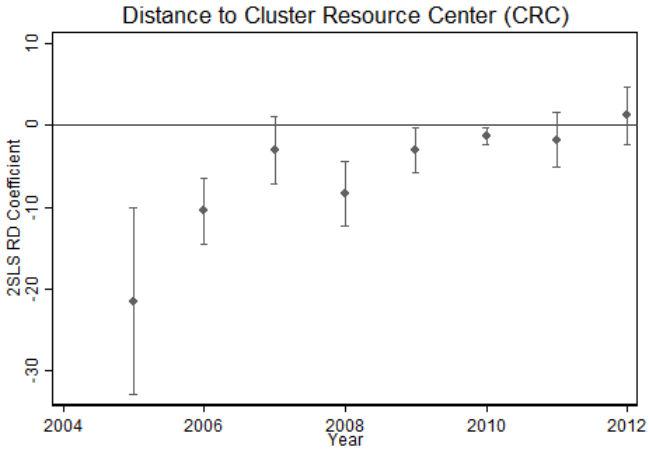
Drinking Water



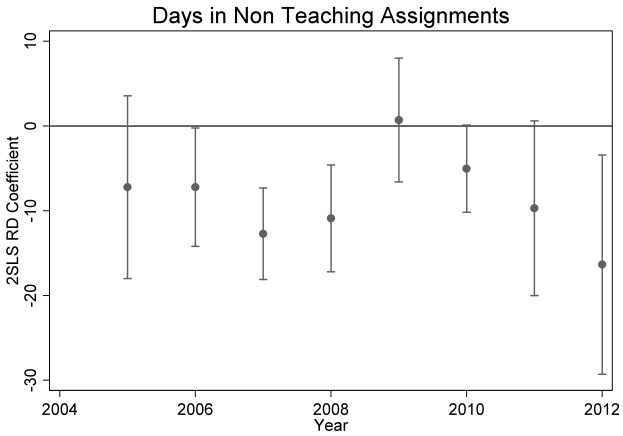
Coefficient Over Time: Electricity



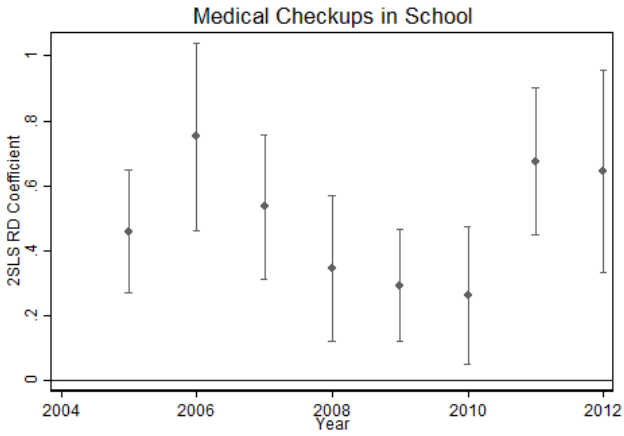
Coefficient Over Time: Pre-Primary Schools



Coefficient Over Time: Distance to CRC



Days Involved in Non-Teaching Assignments



Medical Checkups

Source: DISE data. RD graphs (Regression Function Fit) uses the 2005 data. RD graph optimal binning and 2SLS RD coefficients calculated using Calonico et al. (2014b) procedure. Cluster Resource Centers (CRCs) provide facilities and training to teachers.

Table A.1: Treatment on the Treated using Two-Staged Least Squares Fuzzy RD

Panel A: Full Sample Years of Education	Young	Old	Young	Old
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
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Panel B: Reported Earnings Years of Education	**Young**	**Old**	**Young**	**Old**
RD Estimate	1.660	-0.177	1.569	0.211
	(0.458)***	(0.451)	(0.390)***	(0.396)
Observations	10,175	11,293	14,277	16,007
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper-primary	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.171	-0.0350	0.165	0.000448
	(0.0454)***	(0.0381)	(0.0380)***	(0.0333)
Observations	9,045	7,729	10,175	9,920
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log Earnings	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.258	-0.0235	0.326	0.0910
	(0.0720)***	(0.0769)	(0.0605)***	(0.0671)
Observations	10,175	11,293	14,277	16,007
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 'P(receiving DPEP)'. The sample of 'Young' are of school going age during the policy, whereas those 'Old' 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.

Table A.2: Education and Earnings by Age Groups

Years of Education - Young	16 to 25	26 to 35	16 to 25	26 to 35
RD Estimate	1.038	0.519	0.963	0.548
	(0.262)***	(0.282)*	(0.206)***	(0.234)**
Observations	4,071	5,747	7,301	8,874
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Years of Education - Old	**36 to 45**	**46 to 55**	**36 to 45**	**46 to 55**
RD Estimate	-0.397	0.396	-0.221	0.301
	(0.363)	(0.429)	(0.316)	(0.358)
Observations	4,502	3,158	5,508	4,285
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log(Wages) - Young	**16 to 25**	**26 to 35**	**16 to 25**	**26 to 35**
RD Estimate	0.152	0.0607	0.195	0.123
	(0.0420)***	(0.0441)	(0.0325)***	(0.0358)***
Observations	4,072	5,747	7,302	8,874
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log(Wages) - Old	**36 to 45**	**46 to 55**	**36 to 45**	**46 to 55**
RD Estimate	-0.0890	0.0163	-0.0287	0.0880
	(0.0589)	(0.0750)	(0.0509)	(0.0623)
Observations	4,501	3,157	5,507	4,284
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. Coefficients measure the change in the dependent variable on crossing the RD cutoff. Bandwidths: 'CCT' is the Calonico et al. (2014b) method. 'I and K' is the Imbens and Kalyanaraman (2012) method.

Table A.3: Migration, Earnings Reported, Paid Monthly, and Unemployment

Panel A: Work Structure P(Wages Reported)	Young	Old	Young	Old
RD Estimate	-0.00366	-0.0105	-0.00366	-0.0105
	(0.00488)	(0.00865)	(0.00513)	(0.00844)
Observations	37,201	42,316	32,742	39,823
Bandwidth selection procedure	CCT	CCT	I and K	I and K
P(Unemployed)	Young	Old	Young	Old
RD Estimate	-0.0125	-0.00421	-0.0157	-0.00394
	(0.00225)***	(0.00186)**	(0.00272)***	(0.00160)**
Observations	82,936	38,060	62,393	50,887
Bandwidth selection procedure	CCT	CCT	I and K	I and K
P(Paid monthly)	Young	Old	Young	Old
RD Estimate	0.0828	0.00998	0.0874	0.0170
	(0.0211)***	(0.0264)	(0.0198)***	(0.0198)
Observations	7,962	7,680	10,395	9,869
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Panel B: Migration Fraction Migrated	Total	Total	Economic	Economic
RD Estimate	-0.000987	-0.00293	-0.000416	0.000628
	(0.0106)	(0.0101)	(0.00124)	(0.000913)
Observations	4,808	5,295	5,762	13,405
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Panel A studies the work structure using National Sample Survey 2009-10. 'P(Earnings Reported)' is probability that earnings are reported (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.'
The sample of 'Young' are of school going age during the policy, whereas those 'Old' are too old to change their schooling in response to the policy.
Panel B on migration uses the small-sample National Sample Survey 2007-8 (64th Round) that asks questions on migration. 'Fraction of household migrated' is the share of household members that have migrated out. 'Total Migrants' are people who may have ever left the village for any reason - the most common reasons are marriage (54%). 'Economic Migrants' (less than 30% of migration) is for work-related reasons.
Bandwidths: 'CCT' is the Calonico et al. (2014b) method. 'I and K' is the Imbens and Kalyanaraman (2012) method.

Table A.4: Education and Earnings - Men

Panel A: Full Sample Years of Education	Young	Old	Young	Old
RD Estimate	0.391	0.188	0.285	0.188
	(0.142)***	(0.130)	(0.0978)***	(0.130)
Observations	16,197	29,622	34,248	29,622
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Panel B: Reported Earnings Years of Education	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.674	0.0582	0.681	0.207
	(0.211)***	(0.283)	(0.196)***	(0.209)
Observations	8,047	6,767	9,638	12,517
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper-primary	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.0697	-0.00119	0.0720	0.0151
	(0.0217)***	(0.0246)	(0.0200)***	(0.0179)
Observations	6,947	6,589	9,841	13,236
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log Earnings	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.146	-0.0332	0.154	0.0632
	(0.0324)***	(0.0460)	(0.0299)***	(0.0334)*
Observations	8,047	6,766	9,638	12,516
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 'Young' are of school going age during the policy, whereas those 'Old' 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. Coefficients measure the change in the dependent variable on crossing the RD cutoff.

Table A.5: Education and Earnings - Women

Panel A: Full Sample Years of Education	Young	Old	Young	Old
RD Estimate	0.0933	-0.0271	0.0675	-0.0227
	(0.152)	(0.143)	(0.157)	(0.131)
Observations	17,244	16,834	16,486	19,809
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Panel B: Reported Earnings Years of Education	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.806	-0.0653	0.782	-0.0780
	(0.479)*	(0.443)	(0.418)*	(0.457)
Observations	2,213	2,128	2,945	2,026
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper-primary	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.0762	-0.0295	0.0882	-0.0297
	(0.0422)*	(0.0349)	(0.0364)**	(0.0360)
Observations	2,620	2,157	2,250	1,998
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log Earnings	**Young**	**Old**	**Young**	**Old**
RD Estimate	-0.0455	-0.0595	0.0351	-0.0686
	(0.0745)	(0.0769)	(0.0643)	(0.0794)
Observations	2,213	2,126	2,945	2,024
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 'Young' are of school going age during the policy, whereas those 'Old' 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. Coefficients measure the change in the dependent variable on crossing the RD cutoff.

Table A.6: 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
	(0.139)	(0.146)	(0.202)
Old	7.91	6.08	-1.83
	(0.153)	(0.147)	(0.212)
Difference	0.66	1.12	0.458**
	(0.127)	(0.127)	(0.179)
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. Tables report means (top) and standard errors (below) clustered at the district level.
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 A.7: District-Age Cells and the Parametric RD

Panel A: District - Age Cells Years of Education	Young	Old	Young	Old
RD Estimate	0.603	0.0458	0.644	0.524
	(0.262)**	(0.321)	(0.232)***	(0.326)
Observations	4,055	4,117	5,614	2,709
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper-primary	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.0765	0.000168	0.0755	0.0480
	(0.0224)***	(0.0265)	(0.0229)***	(0.0331)
Observations	5,433	4,011	4,991	2,561
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Panel B: Parametric RD Years of Education	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.872***	-0.146	0.884***	-0.124
	(0.224)	(0.279)	(0.225)	(0.283)
Observations	10,038	11,088	10,038	11,088
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper-primary	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.0882***	-0.0249	0.0875***	-0.0219
	(0.0219)	(0.0215)	(0.0220)	(0.0217)
Observations	10,038	11,088	10,038	11,088
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10. Sample of persons that reported earnings, ages between 16 and 75 years. The sample of 'Young' are of school going age during the policy, whereas those 'Old' 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. Coefficients measure the change in the dependent variable on crossing the RD cutoff.
Panel A: Data collapsed to the district-age-gender cell level. Panel B: Parametric RDs using local linear and quadratic functions. Bandwidth restricted to twenty percentage points. Sample of persons between 16 and 75 years that reported earnings.

Table A.8: Robustness: In-Progress RD Methods for Bandwidths and Standard Errors

Panel A: Bartalotti and Brummet (2017) cluster-robust variance estimation				
Years of Education	Young	Old	Young	Old
RD Estimate	0.720	-0.0881	0.698	0.0902
	(0.336)**	(0.406)	(0.301)**	(0.378)
Observations	10,175	11,293	14,277	16,007
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper-primary	Young	Old	Young	Old
RD Estimate	0.0743	-0.0175	0.0733	-0.00107
	(0.0328)**	(0.0301)	(0.0284)***	(0.0273)
Observations	9,045	7,729	10,175	9,920
Bandwidth selection procedure	CCT	CCT	I and K	I and K

Panel B: Calonico et al. (2017) 2-sided bandwidth; district cluster-robust nearest neighbor SEs				
Years of Education	Young	Old	Young	Old
RD Estimate	0.463	-0.191	0.563	-0.312
	(0.189)**	(0.304)	(0.205)***	(0.369)
Observations	9,876	8,346	6,709	6,271
Bandwidth selection procedure	MSE-2	MSE-2	CER-2	CER-2
Finished Upper-primary	Young	Old	Young	Old
RD Estimate	0.0814	-0.0230	0.0896	-0.0371
	(0.0166)***	(0.0230)	(0.0188)***	(0.0277)
Observations	9,732	8,323	6,882	6,502
Bandwidth selection procedure	MSE-2	MSE-2	CER-2	CER-2

National Sample Survey 2009-10. The sample of 'Young' are of school going age during the policy, whereas those 'Old' 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.
Panel A: Uses the Bartalotti and Brummet (2017) method to compute standard errors at the district-age group level. I thank the authors for sharing their code. The optimal bandwidths are chosen using the Calonico et al. (2014b) and Imbens and Kalyanaraman (2012) methods.
Panel B: Uses an in-progress method developed by Calonico et al. (2017) that allows for a separate optimal bandwidth on either side of the cutoff and cluster-robust standard errors at the district level. MSE-2 is mean squared error optimal two-sided bandwidth, and CER-2 is the coverage error rate two sided bandwidth.

Table A.9: Robustness: Widening Age Restrictions

Panel A: Full Sample Years of Education	Young	Old	Young	Old
RD Estimate	0.257	0.104	0.259	0.140
	(0.0680)***	(0.116)	(0.0734)***	(0.108)
Observations	74,342	35,064	63,388	39,456
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Panel B: Reported Earnings Years of Education	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.736	-0.180	0.732	-0.0557
	(0.194)***	(0.269)	(0.180)***	(0.247)
Observations	10,559	8,002	12,814	9,057
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper-primary	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.0728	-0.0262	0.0727	-0.0162
	(0.0204)***	(0.0231)	(0.0201)***	(0.0174)
Observations	9,662	7,734	10,117	13,441
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log Earnings	**Young**	**Old**	**Young**	**Old**
RD Estimate	0.116	-0.0116	0.130	-0.00263
	(0.0308)***	(0.0372)	(0.0283)***	(0.0340)
Observations	10,560	11,302	12,815	13,823
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 'Young' are of school going age during the policy, whereas those 'Old' 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. Coefficients measure the change in the dependent variable on crossing the RD cutoff.

Table A.10: 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 A.11: District GDP 2000 and 2004

Log(District GDP)	Y 2000	Y 2000	Y 2004	Y 2004
RD Estimate	0.0684	0.0890	0.110	0.144
	(0.141)	(0.139)	(0.114)	(0.116)
Observations	103	105	173	178
Bandwidth selection procedure	CCT	I and K	CCT	I and K
District GDP (Rupees)	**Y 2000**	**Y 2000**	**Y 2004**	**Y 2004**
RD Estimate	2,072	2,415	2,834	3,449
	(2,951)	(3,222)	(2,369)	(2,690)
Observations	112	112	195	210
Mean dependent variable	15731	15731	17459	17459
Bandwidth selection procedure	CCT	I and K	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.
'Y 2000' indicates calendar year 2000, whereas 'Y 2004' is calendar year 2004.
Bandwidths: 'CCT' is the Calonico et al. (2014b) method. 'I and K' is the Imbens and Kalyanaraman (2012) method.

Table A.12: 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).
Bandwidths: 'CCT' is the Calonico et al. (2014b) method. Results show Treatment on the Treated (TOT) scaled up by probability of treatment.
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.

B Derivations in the Model

B.I Student Lifetime Utility

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:

$$\max_{\mathbf{C_{it}},\mathbf{s_{id}}} \int_0^\infty u(C_{it})e^{-\delta t}dt \quad s.t. \quad \int_0^\infty w_{aid}(s_{id})e^{-r_{id}(\kappa(s_{id})+t)}dt \geq \int_0^\infty C_{it}e^{-r_{id}t}dt \ , \tag{26}$$

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

$$\log \left(\int_0^\infty w_{aid}(s_{id})e^{-r_{id}(s_{id}+t)}dt \right) = \log \ w_{aid}(s_{id}) - (\log \ r_{id} + r_{id}s_{id}) \tag{27}$$

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.

After the students choose their optimal years of education, the consumption side can be solved. 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.

B.II Education Sector

B.II.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.

$$\max_{\mathbf{x_m}} \ A_d(\mathbf{x_m}) \tag{28}$$

$$s.t. \ \sum_{m=1}^{M} p_m x_m \leq R_d \ , \tag{29}$$

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:

$$A(\mathbf{x_m}) = \prod_{\mathbf{m}} \mathbf{x_m}^{\alpha_\mathbf{m}} , \tag{30}$$

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:

$$A_d(R_d, \mathbf{p_m}) = \mathbf{R_d} \prod_{\mathbf{m}} \left(\frac{\alpha_\mathbf{m}}{\mathbf{p_m}} \right)^{\alpha_\mathbf{m}} \tag{31}$$

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:

$$\max_{X_j} \; p_d \bar{\theta}_d X_j - Z(X_j) , \tag{32}$$

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:

$$Q_{jd} = \bar{\theta}_d X^*_j = \bar{\theta}_d \frac{p_d \bar{\theta}_d - z_{1j}}{z_{2d}} , \tag{33}$$

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:

$$S^{sy}_{pvt,d} = \int_0^{p_d \bar{\theta}_d} \bar{\theta}_d \frac{p_d \bar{\theta}_d - z_{1j}}{z_{2d}} f(\tilde{z_1}) dz_{1j} = \frac{\bar{\theta}_d}{z_{2d}} [p_d \bar{\theta}_d - \mathbb{E}_d(z_{1j} | z_{1j} < p_d \bar{\theta}_d)] , \tag{34}$$

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:

$$\Pi = \int_0^{\bar{\theta}_d p_d} \frac{(p_d \bar{\theta}_d - z_{1j})^2}{z} dF(z_{1j}) \tag{35}$$

B.II.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:

$$S^{Dd}_d = \int \frac{\beta_d + \Psi A_d - p_d - \eta_i}{\Gamma} dH(\eta) = \frac{\beta_d + \Psi A_d - p_d - \bar{\eta_d}}{\Gamma}, \tag{36}$$

where $\bar{\eta_d} = \mathbb{E}[\eta_i | i \in d]$. The overall supply of schooling comes from both public and private schools:[56]

$$S^{Sy}_d = \frac{\bar{\theta_d}}{z_{2d}} [p_d \bar{\theta_d} - \mathbb{E}_d(z_{1j} | z_{1j} < p_d \bar{\theta_d})] + A_d \tag{37}$$

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 (36) and (37) 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 $f(\tilde{a}) \sim U[0, p_d \bar{\theta_d}]$, then the equilibrium price and quantity is:[57]

$$p^*_d = \frac{\beta_d + (\Psi - \Gamma) A_d - \bar{\eta_d}}{\Gamma \left(\frac{\bar{\theta_d^2}}{z_{2d}} \right) + 1} \quad and \quad S^*_d = \frac{\bar{\theta_d^2} (\beta_d + \Psi A_d) + z_{2d} A_d}{\Gamma \bar{\theta_d^2} + z_{2d}} - \frac{\bar{\eta_d}}{\Gamma} \tag{38}$$

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.

$$r_{id} = -R_d \Psi \prod_m \frac{\alpha_m}{p_m}^{\alpha_m} + p^*_d(R_d) + \eta_i \tag{39}$$

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

$$S^*_d = \phi_1 \beta_d + \phi_2 R_d - \frac{\eta_d}{\Gamma}, \tag{40}$$

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 (40) the equilibrium

[56] 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.

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

amount of schooling is affected by the expansion of public schooling.

B.III 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:

$$\log w_{asd} = \log\left(\frac{\partial Y_d}{\partial \ell_{asd}}\right) = \log\theta_{sd} + \log\psi_a + \left(\left(\frac{1}{\sigma_E} - 1\right)\left(\frac{1}{\varrho}\right)\log Y_d - \left(\frac{1}{\sigma_E} - 1\right)\left(\frac{1-\varrho}{\varrho}\right)\log\bar{K}_d\right) + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right)\log L_{sd} - \frac{1}{\sigma_A}\log\ell_{asd}\ , \tag{41}$$

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.

B.IV Skill Biased Capital

In Model subsection 2.1 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 (2) has been modified into Equation (44), which includes an elasticity of substitution between labor ℓ_{sd} and skill biased capital k_{sd} represented by σ_s:

$$Y_d = L_d^{\varrho} K_d^{(1-\varrho)} \tag{42}$$

$$L_d = \left(\sum_s \theta_{sd} L_{sd}^{\frac{\sigma_E - 1}{\sigma_E}}\right)^{\frac{\sigma_E}{\sigma_E - 1}} \tag{43}$$

$$L_{sd} = \left(\Lambda_s k_{sd}^{\frac{\sigma_s - 1}{\sigma_s}} + (1-\Lambda_s)\ell_{sd}^{\frac{\sigma_s - 1}{\sigma_s}}\right)^{\frac{\sigma_s}{\sigma_s - 1}} \tag{44}$$

$$\ell_{sd} = \left(\sum_a \psi_a \ell_{asd}^{\frac{\sigma_A - 1}{\sigma_A}}\right)^{\frac{\sigma_A}{\sigma_A - 1}} \tag{45}$$

Given this new set up, earnings can be represented by Equation (46), instead of Equation (3):

$$\log w_{asd} = \log\widetilde{\varrho} + \log\psi_a + \frac{1}{\sigma_E}\log Y_d + \left(\frac{1}{\sigma_s} - \frac{1}{\sigma_E}\right)\log L_{sd} + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_s}\right)\log\ell_{sd} - \frac{1}{\sigma_A}\log\ell_{asd} \tag{46}$$

This new set up does not change the estimation or the interpretation of the estimates. In the following equation, that replaces Equation (18) to estimate the GE effects on all workers, the skill-biased capital term is captured by the term L_{sd}:

$$log\frac{w_{so,D=1}}{w_{so,D=0}} - log\frac{w_{uo,D=1}}{w_{uo,D=0}} = \left(\frac{1}{\sigma_s} - \frac{1}{\sigma_E}\right)\left[log\frac{L_{s,D=1}}{L_{u,D=1}} - log\frac{L_{s,D=0}}{L_{u,D=0}}\right] + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_s}\right)\left[log\frac{\ell_{s,D=1}}{\ell_{u,D=1}} - log\frac{\ell_{s,D=0}}{\ell_{u,D=0}}\right] \tag{47}$$

B.V Deriving Equations (20) and (21)

In Equations (20) and (21) 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:

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

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

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

$$= \ell_{sy,D=1}log\frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=1}log\frac{w_{uy,D=1}}{w_{uy,D=0}} + \Delta\ell_{sy}\underbrace{log\frac{w_{sy,D=0}}{w_{uy,D=0}}}_{\beta_{as,D=0}} \tag{48}$$

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

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

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

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

$$= \ell_{sy,D=0}log\frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=0}log\frac{w_{uy,D=1}}{w_{uy,D=0}} + \Delta\ell_{sy}\underbrace{log\frac{w_{sy,D=1}}{w_{uy,D=1}}}_{\beta_{as,D=1}} \tag{49}$$

C Details about DPEP Guidelines and Funding

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.[58]

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.[59]

The funding largely came from international donors like 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. In general, 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,

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

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

whereas UNICEF spent $ 153 mn in Bihar (GOI, 2000). World Bank (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).

Other than building schools and hiring teachers, an additional objective was to improve the access to primary and upper primary education by establishing district institutions to decentralize planning. 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. 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.

Of the approximately 160,000 new schools, more than 84,000 were 'alternative' or 'community schools.' 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. 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.

D 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 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 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.

Annual Status of Education Report (ASER): To study the impact on test scores, I use a geographically comprehensive data set that consists of a household survey done by an NGO (Pratham). The survey focuses on children in the age group 3-16. It surveys children at home – whether they went to government school, private school, religious schools and even dropouts. The focus of the testing is on the ability to read simple texts and do basic arithmetic.

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.

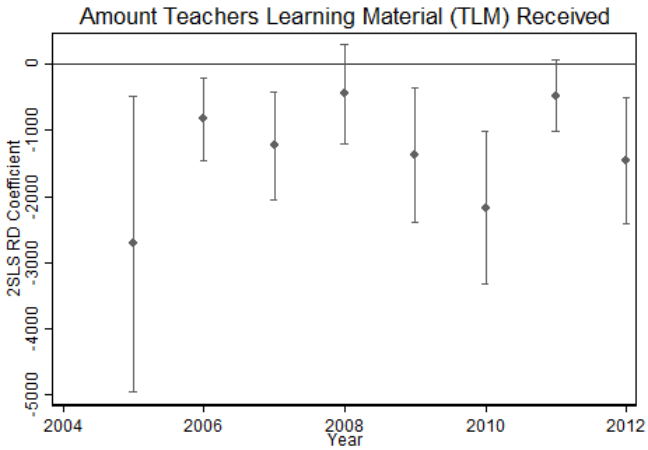
E Other Impacts on the Education Sector

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

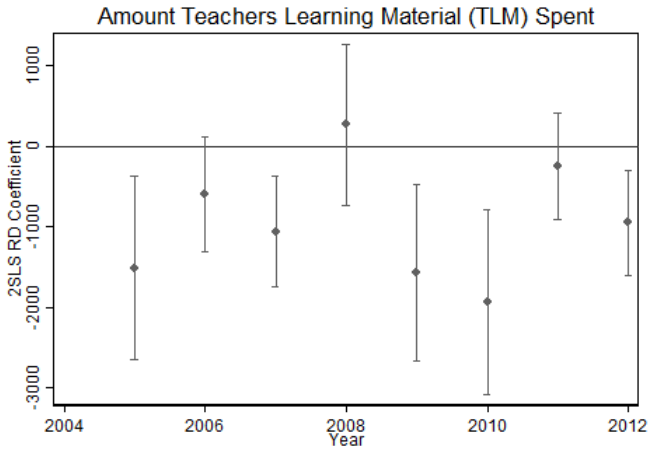
DPEP districts or not. In the middle panel of Appendix Figure A.8, one can see that regions that were eligible for DPEP systematically spent less TLM funds, showing the possibility that other funds were actually crowded out when DPEP funds were allocated. Appendix Figure E.1 shows that DPEP regions both received less and spent less TLM grant money. 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 (Appendix Figure E.3).

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 assist with teacher training. In Appendix Figures A.9 and E.2, 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 of Figure E.2, however, one can see that the number of academic inspections and visits by center officials was, over time, consistently higher in treated areas.

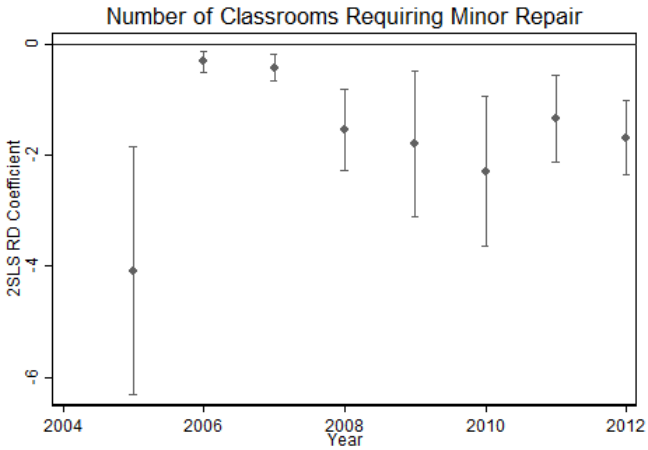
Figure E.1: Other Funds Spent, and Condition of Rooms



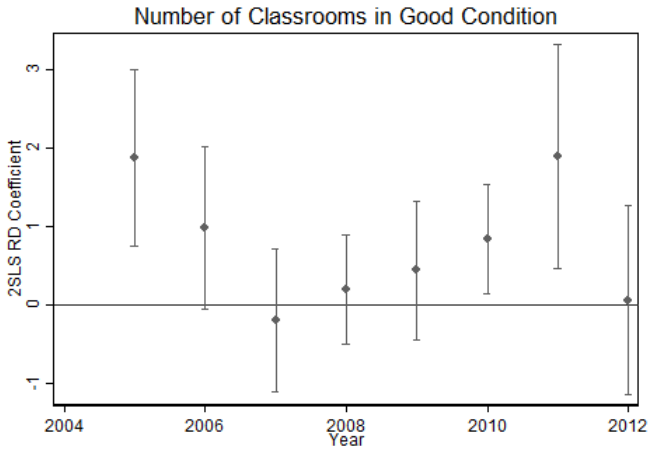
RD Coefficient Over Time: TLM grants Received



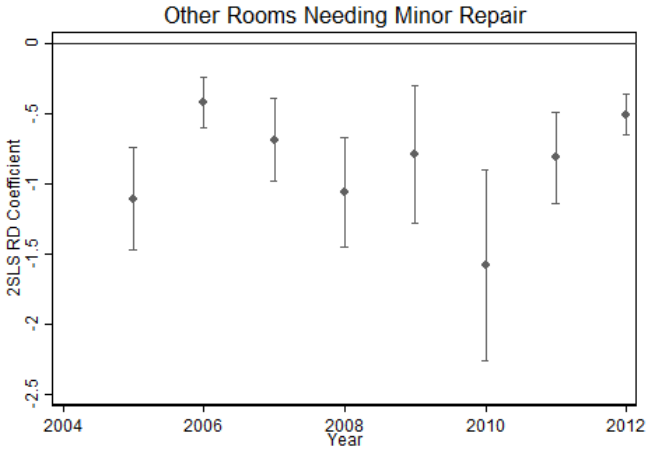
RD Coefficient Over Time: TLM grants Spent



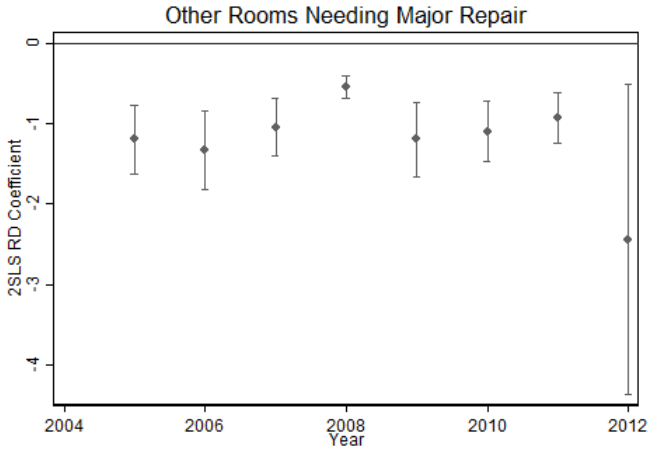
RD Coefficients: Classrooms Need Minor Repair



RD Coefficients: Classrooms in Good Condition



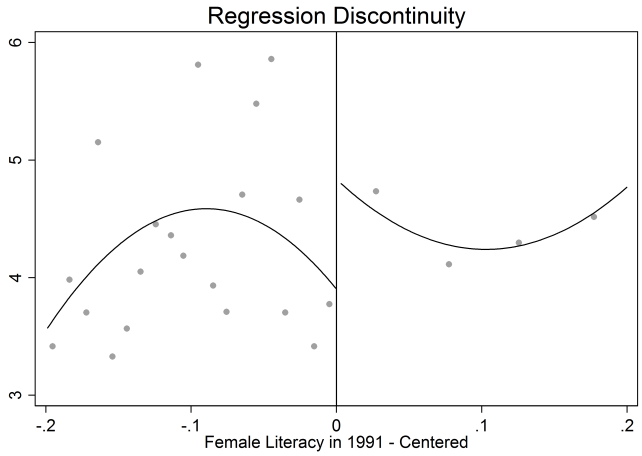
RD Coefficients: Other Rooms Need Minor Repair



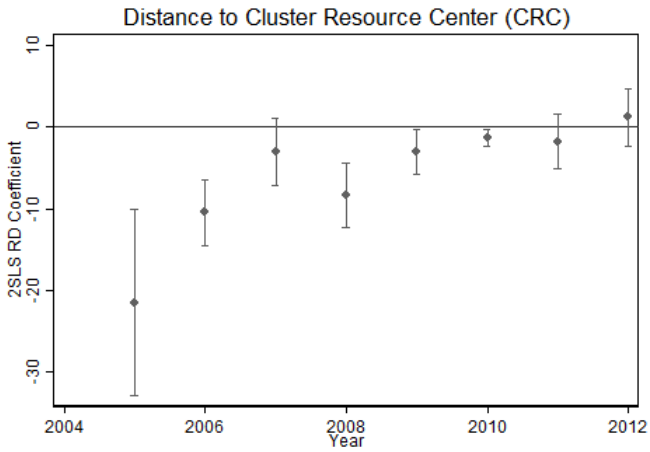
RD Coefficients: Other Rooms Need Major Repair

Source: DISE (District Information System for Education) data. RD graphs (Regression Function Fit) uses the 2005 data. RD graph optimal binning and 2SLS RD coefficients calculated using Calonico et al. (2014b) procedure. All schools, regardless of their which district they are in, are eligible to receive the Teacher Learning Materials (TLM) grant.

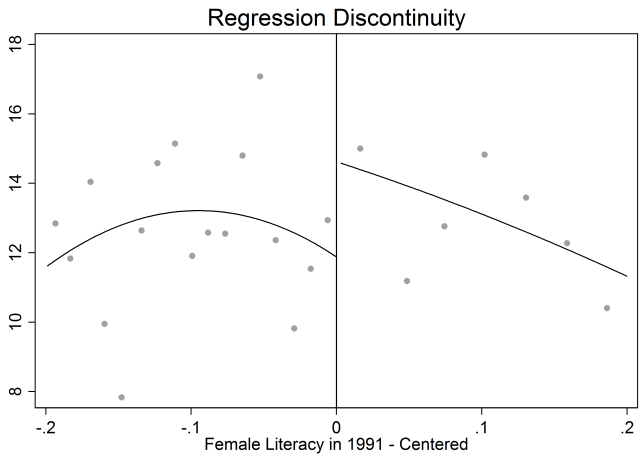
Figure E.2: Academic Inspections and Regional Resource Centers



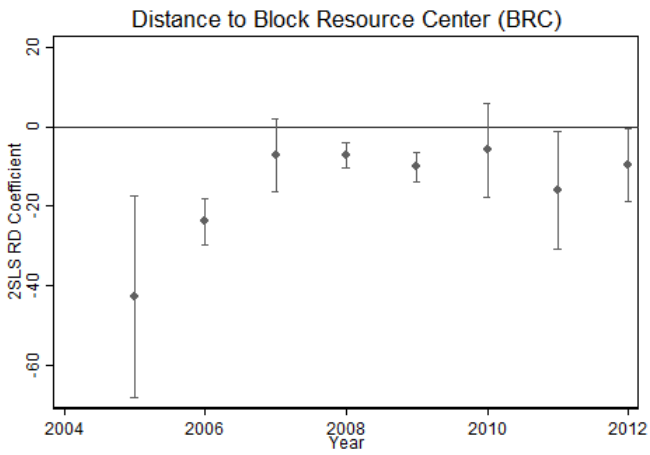
Distance to CRC (2005)



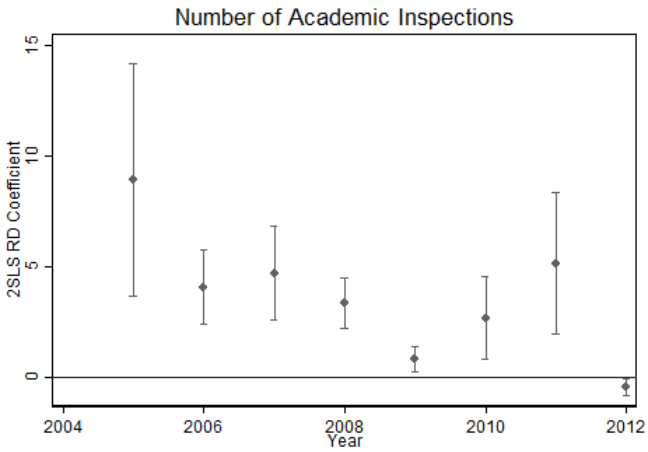
Coefficient Over Time: Distance to CRC



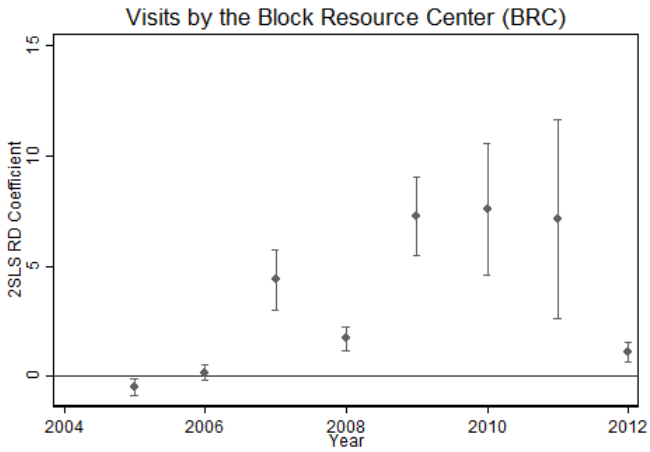
Distance to BRC (2005)



Coefficient Over Time: Distance to BRC



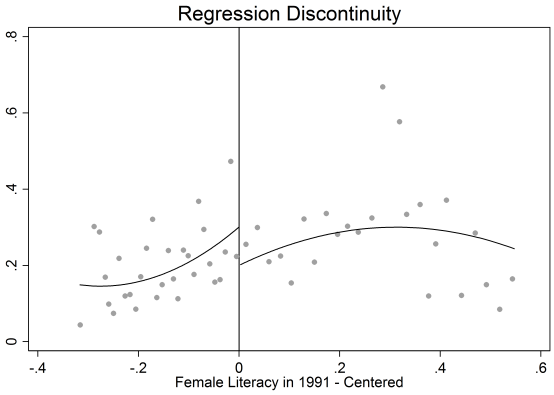
Number of Academic Inspections



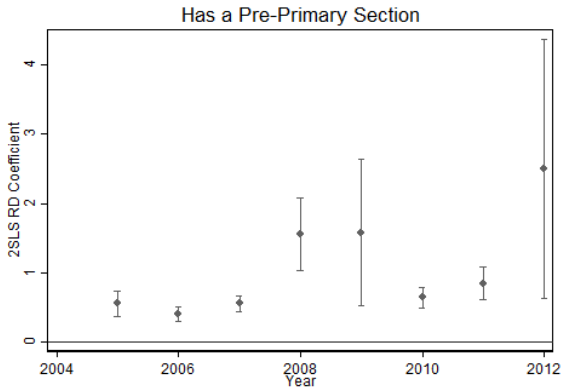
Visits by BRC Official

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 E.3: Pre-Primary Sections



Schools with Pre Primary Sections



Coefficient Over Time: Pre-Primary Schools

Number of Pre Primary Teachers

Number of 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.