

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 general equilibrium effects in both the labor market and education sector. I develop a general equilibrium model that captures the consequences of massive countrywide schooling initiatives. I provide unbiased estimates of the model's elasticities using a Regression Discontinuity design derived from 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 by 6.6 percentage points. These general equilibrium effects have distributional consequences across cohorts and skill groups: as a result of the policy, unskilled workers are better off and skilled workers worse off.

JEL: I25, I26, O15, I28

Keywords: Returns to education, school building, general equilibrium

*www.econgaurav.com. 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, Natalie Bau, Hoyt Bleakley, Charles Brown, Sebastian Calonico, Matias Cattaneo, Austin Davis, James Fenske, Rob Garlick, Gordon Hanson, Brian Jacob, David Lam, Margaret Lay, Katie Lim, Meera Mahadevan, Ben Meiselman, Karthik Muralidharan, Gareth Nellis, Tommaso Porzio, Heather Royer, Isaac Sorkin, Kevin Stange, Ben Thompson, and Mel Stephens for valuable comments. I am grateful to seminar participants at Michigan, Yale, NEUDC, Urban Institute, McGill, UC Davis, UCSD, William and Mary, Georgia State, Mathematica, PAA, Society of Labor Economists, Stockholm School of Economics, Stockholm University, IFPRI, UN Reno, World Bank, UToronto, and Center for Global Development for valuable comments.

Large-scale education expansions represent substantial investments of public resources and benefit households by increasing local productivity. 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, these estimates may not be sufficient when evaluating large-scale policies. Importantly, large-scale education programs may have sizable general equilibrium (GE) effects on both education and labor markets. I causally estimate and take into account these GE effects while determining the overall economic implications of nationwide education programs.

I build a 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 GE effects, which together determine welfare consequences. On the labor market side, I combine models of education choice (Becker, 1967) with determinants of the skill-premium (Card and Lemieux, 2001) to study how the distribution of earnings affects education choices, and consequently how changes in education affect the distribution of earnings.

The allocation rule under which Indian districts receive the funding allows me to estimate important elasticities using a Regression Discontinuity (RD) approach. The policy raised education levels for the young. As such, I further exploit variation between younger (young enough to change education) and older cohorts, and between high skill (educated) and low skill workers to identify GE effects, by estimating how the earnings skill-premium changes across local economies. I 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 33% and dampen the increase in labor market benefits by 23.8%. 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.

As education levels rise, we expect earnings and therefore the returns to be affected in a few ways. First, more educated workers are more productive and earn higher wages. Second, educated workers may reside in regions where there are fewer educated workers, making them 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 earnings skill premium. Yet, as more skilled workers join the labor force, skill-biased capital may be adopted by firms in these regions, raising the premium. Indeed, as workers switch to more productive skill groups, overall output may increase to the benefit of all workers. I estimate all of these components of the GE effects to better quantify the distributional impacts and the changes in labor market benefits.

The policy I study was India's flagship education 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 consequences would be similarly widespread (Jalan and Glinskaya, 2013). Such schooling expansions reduce the marginal cost of attaining education by improving access to schools (Behrman et al., 1996; Birdsall, 1985; Duflo, 2001), inducing 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 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 induced into getting more education to similarly competent students that were not. At the regional level, the RD tackles biases that arise due to differences in the local economy and labor market, as some regions may have more skilled workers or industries.

To support each piece of the general equilibrium model, I create a comprehensive dataset, 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 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. However, both the young and the old are differently 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 for all cohorts. At the same time, since the young and old are not perfect substitutes (and in some contexts may be complements), there is an often overlooked additional impact on young workers which I estimate by looking at the additional change in the skill-premium for the young. 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 us to address such labor market effects. In the US, Heckman et al. (1998a,b,c), Abbott et al. (2019), and Lee (2005) show how changes in taxes or tuition and financial aid may have large GE effects.¹ I flexibly model and causally estimate the GE effects on different cohorts and skill groups. This allows me to determine distributional consequences across both dimensions, estimate crucial economic parameters, and schooling returns both in the presence and absence of GE effects.

I corroborate my results with a Difference-in-Differences (DID) analysis, where I compare treated to untreated districts and younger to older cohorts. However, as I show, using a DID design, it is challenging to recover the GE effects, and accordingly adjust the returns to skill.

¹Labor market GE effects include job assistance programs (Crepon et al., 2013), calibrated macro models (Albrecht et al., 2009), effects on demographics (Epple and Ferreyra, 2008), and major choice (Bianchi, 2016).

The advantage of the RD is that I can estimate the entire extent of GE effects, and disentangle them into the portion that affects all cohorts and additional impacts on treated cohorts. [Duflo \(2004\)](#) shows that Indonesia’s school-building depressed average wages for older untreated cohorts. As the young and old differ in skill composition, and are not perfect substitutes (and in fact, may be complements), I show that the old may not be a reasonable counterfactual for the young when there are GE effects. I estimate 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 Indeed, I find that the GE effects on the old are negligible in comparison to the GE effects on the (treated) young. The impacts on older cohorts is an object of interest on its own, and such spillovers may be missed if we focus on treated cohorts alone.

The advantage of the RD (over difference-in-differences) is that I do not need to rely on the effects on the old to identify the effects on the young. Yet, simply comparing the change in average earnings of the young across the RD cutoff (for a unit increase in education) does not recover either the partial or general equilibrium returns to education, as one may do with a conventional IV-Wald estimator. However, by studying how the skilled wage, and the unskilled wage change at the RD cutoff allows us to recover the GE effects, and consequentially the partial equilibrium returns to education.

I find that the program increased both education and earnings for students in targeted regions. I find large overall economic benefits to households, driven by reductions in the costs of education and increases in the overall output of the region. However, GE 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 a year of education including GE effects 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 6.6 percentage points and dampen the increase in benefits to students by 23.8%.³ These GE effects have distributional consequences, with a transfer of labor-market benefits from skilled to unskilled workers, particularly among the young. High-skill workers who did not change their education 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, 2009, 2016](#)), I find no evidence of labor mobility.

These results have significant implications for research and policy. First, there are few contemporary estimates of skill returns from developing countries like India, which was undergoing

²Given elasticities of complementarities across skill and cohorts, the GE effects may also raise average wages for older cohorts as unskilled wages rise.

³As the policy targets the upper-primary level, and the largest education response is there, these returns are largely representative of a year of upper-primary schooling. Given the focus on primary schools (under the Millennium Development Goals, and national efforts like DPEP), this is a policy relevant parameter.

a sustained period of growth in both education and income. Second, my methods broadly speak to our attempts to causally estimate the private returns to education using macro-level variation from tuition reductions, compulsory schooling laws, schooling expansions, or 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 \(1958\)](#), and [Willis \(1986\)](#). Rather it involves treating an entire cohort of students, leading to GE effects, which I find to be substantial. Third, I show that returns to education are not fixed parameters, but rather endogenous quantities which depend on the local labor market, and I derive meaningful relationships between labor market changes and returns.

Fourth, a substantial number of micro-interventions help guide policy-makers.⁴ My analysis complements these with information on how to measure the effects of scaled-up versions of such interventions.⁵ GE effects will either undermine or amplify the effectiveness of micro-interventions when they are scaled up ([Acemoglu, 2010](#); [Deaton, 2010](#); [Egger et al., 2021](#); [Muralidharan and Niehaus, 2017](#)). This point is stressed even outside the realm of development economics ([Heckman et al., 1999, 1998a,b](#)).

Finally, these methodological insights can be applied in other contexts, allowing researchers and policy-makers to estimate the welfare consequences of such interventions around the world (for example, school-building, compulsory schooling, fee reductions or job programs).⁶

1 The District Primary Education Project (DPEP)

I derive plausibly exogenous variation from a large schooling expansion policy in India, where any district below the 1991 national average female literacy rate was eligible to receive funds. There are two 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 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](#)).⁷ 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. It was the

⁴In India these studies cover library programs ([Borkum et al., 2010](#)), teacher incentives ([Muralidharan and Sundararaman, 2010](#)), bicycles ([Muralidharan and Prakash, 2017](#)), computer-aided programs ([Linden, 2008](#)), remedial education ([Banerjee et al., 2007](#)), and teacher quality or absence ([Das et al., 2013](#); [Duflo et al., 2012](#)).

⁵While the evidence on smaller changes of inputs within schools is mixed ([Muralidharan, 2013](#)), large-scale investments in schooling like the one studied here, have been found to be relatively more successful across the world. Some examples are in Indonesia ([Duflo, 2001](#)), Burkina Faso ([Kazianga et al., 2013](#)), Zimbabwe ([Aguero and Bharadwaj, 2014](#)), Nigeria ([Osili and Long, 2008](#)), Uganda ([Deining, 2003](#)), Zambia ([Ashraf et al., 2020](#)), Kenya ([Bold et al., 2013](#)), and India ([Adukia, 2016](#); [Afridi, 2010](#); [Chin, 2005](#)).

⁶The RD is simply a special case of an instrumental variables (IV) method. The insights here can be applied to any IV that leverages variation from a large-scale policy.

⁷States maintained the level (in 1992 [Varghese \(1994\)](#)) and growth rate ([World Bank, 1997](#)) of expenditure that existed before the program, to ensure no crowd-out of funds. Taxes were not raised to directly fund DPEP.

flagship education program for more than a decade, but was phased out in 2006.⁸

The program served approximately 51.3 million children. These districts were geographically dispersed across the country (map in Figure A.1). It created about 160,000 new schools, trained 1.1 million teachers and 3 million community members, and increased funds for primary school education by between 17-20%. The primary objective was to improve student access to and retention in primary and upper primary education. Numerous policy briefs and media reports highlight the program’s success.⁹ Other work shows that DPEP increased education in treated districts (Azam and Saing, 2016; Jalan and Glinskaya, 2013). Over the period, districts could receive about \$8 million, or \$9.1 per student, and even in the short run this intervention lowered private household costs of schooling by 20-40% (Jalan and Glinskaya, 2013). Additional details on the history, funding and secondary objectives can be found in Appendix C.

2 The Model

My framework captures salient features of the local economy and market for education, including GE effects of policy reforms. I derive estimation equations and identify elasticities that determine the effect of schooling expansions on economic benefits in Section 4.¹⁰

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).¹¹ Capital is perfectly-elastically supplied across districts at rental rate R^* .¹² Effective labor supply L_d depends on the labor aggregate L_{sd} at each skill level s .

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

$0 < \varrho < 1$ is the share of output accruing to labor, $\theta_{sd} > 0$ captures worker productivity 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 $\theta'_{sd}(k_{sd}) > 0$.¹³

⁸Source: Parliamentary Questions: 1807, 552, 55, 267, 1320, 2018, and Upper House Question 2855.

⁹See World Bank (2003), “WB praises India for DPEP” Economic Times (2005) and Gov of India (2011).

¹⁰As I explain in Section 4.1.1, the advantage of relying on only a few elasticities is that the estimation procedure and measured GE effects do not directly depend on imposed functional forms. Indeed, the estimated wage benefits will hold true even under many alternative formulations. However, couching it in a canonical labor economics model allows us to understand parameters driving the effects, drawing links to important elasticities estimated in other work, and underlying determinants of the returns to education.

¹¹Including non-tradables like land, in the aggregate production function does not directly affect estimation. The policy will theoretically change the value of non-tradables; yet, I focus on the earnings of workers, and not returns to owners of capital and land.

¹²The perfectly elastic capital assumption is not essential. The results are unaffected by assuming a fixed capital stock (see Appendix B.II).

¹³The amount of ‘unbiased’ capital does not affect the returns to skill, yet skill-biased capital does. When there is an increase in the supply of skilled workers we may expect the skilled wage to fall. Yet, if firms start using more skill-biased capital (now that they have skilled labor), then that mitigates the fall in skilled wages. Allowing for such demand-side responses will allow for dampened GE effects. For completeness, in Appendix B.III, 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.

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 :

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

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} can depend on the ability of workers ϵ_i , and a worker gets paid their marginal product.¹⁴ The average log earnings are therefore:

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

where $\log \tilde{\varrho} \equiv \left[\left(1 - \frac{1}{\sigma_E} \right) \left(\frac{1-\varrho}{\varrho} \right) \log \left(\frac{1-\varrho}{R^*} \right) \right]$ is common across all districts and workers.¹⁵

There are a few components that drive differences in average earnings when comparing two different types of people in two different labor markets, as in Equation (4):

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

This equation is crucial in that it captures why earnings are systematically different across people and 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.

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 skill β_{asd} :

¹⁴In Section 5.3.6 we examine the implications of labor market distortions, factor misallocation and monopsonistic power.

¹⁵This is at the optimal value of K_d^* , where $Y_d = \left(\frac{1-\varrho}{R^*} \right)^{\frac{1-\varrho}{\varrho}} L_d$. For tractability, I ignore the role of changing prices of non-tradables. One can include a $\log P_d$ with the $\frac{1}{\sigma_E} \log Y_d$ term, not affecting the returns to skill.

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

This highlights an important fact: the returns to skill 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 are higher.

Importantly, as I show in Appendix B.VI, migration is consistent with 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 may be in the direction of equalizing wages, attenuating the negative GE effects.

2.2 Students' Decisions

Students choose the optimal level of education given their marginal costs of schooling and 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.¹⁶ Student i chooses their years of education s_{id} to maximize the present discounted value of expected log lifetime earnings $\mathbb{E}[\log w_{aid}(s_{id})]$ given schooling costs $\kappa(s_{id})$.¹⁷

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

where Γ is the quadratic cost parameter. Equations (4) and (5) determine the form of the individual earnings function. The return to an additional year of schooling $\tilde{\beta}_{asd}$ depends on returns to skills β_{asd} , and the difference in schooling years between the skilled (s_1 years) and unskilled (s_0), where $\tilde{\beta}_{asd} = \frac{\beta_{asd}}{(s_1 - s_0)}$. These returns differ across districts, cohorts and skill.¹⁸

$$\mathbb{E}[\log w_{aid}(s_{id})] = \mathbb{E}[\gamma_d + \gamma_a] + \mathbb{E}[\tilde{\beta}_{asd}]s_{id} + \log \epsilon_i \quad (7)$$

Unlike the aggregate average wage function in Equation (3), the individual-level earnings also depend on ϵ_i , 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 ($\text{corr}(\epsilon_i, r_{id}) \neq 0$); high-ability workers earn

¹⁶Willis (1986) shows how such formulations are derived from conventional utility functions. Intuitively, maximizing lifetime earnings maximizes lifetime consumption, and thereby utility when consumption is optimally allocated across periods.

¹⁷Since 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) suggests the quadratic costs from the observation that each subsequent year of education is even more expensive than before, as (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).

¹⁸For tractability, I suppress any time subscripts. The expectations operator captures expectations on future earnings at the time the schooling decisions are made. Notice that at the time of schooling decision, schooling levels and abilities are known, and hence have no expectations operator. Yet, there may be uncertainty on future earnings by skill, cohort and district.

high wages but have lower costs of schooling. Crucial to notice is that the returns to education $\tilde{\beta}_{asd}$ 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. Unlike Equation (3) on aggregate returns, here there is an expectations operator, as schooling decisions respond to expected, rather than actual returns (Hastings et al., 2018; Jensen, 2010).

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 γ_a given returns to experience or other cohort-specific productivities captured in Equation (4). From the first-order conditions one can obtain the optimal years of education for person i :

$$s_{id}^* = \frac{\mathbb{E}[\tilde{\beta}_{asd}] - r_{id}}{\Gamma} \quad (8)$$

The variation in s_{id}^* within a district d is driven entirely by the variation in the marginal cost parameter r_{id} .¹⁹ The marginal cost parameter for person i in district d is a function of the district-level costs of going to school, and an individual component η_i that captures individual heterogeneity in the costs of schooling. The district-level costs depend on the access to schooling A_d (distance to the nearest school) and the monetary price of school p_d (like school fees).

$$r_{id} \equiv -\Psi A_d + p_d + \eta_i \quad (9)$$

Ψ represents how access to education affects individual i . More schools in regions that did not initially have many, lower transportation costs of going to school, but may also lower competitive equilibrium fees, even for private schools. As such, the individual costs of schooling are a direct function of the funds received under the program, R_d .

The schooling response in Equation (8) also depends on expected wage returns $\mathbb{E}[\tilde{\beta}_{asd}]$. It is not necessary that households have perfect foresight over their expected wage, for instance, as they may be unable to predict the magnitudes of the general equilibrium effects. If expectations (whether accurately or not) suggest substantially low returns, then cost reductions may not induce many to get education. As such, expectations will affect how a fall in the costs of schooling increase education (Heckman et al., 1998a,b,c).²⁰ Yet, some of the later analysis is focused on examining the change in *ex post* education returns as skill-prices adjust in response to a *given* change in schooling. Here, expectations matter less. As I show below, when measuring the economic benefits, once again, expectations play an important role.

In Appendix B.VI, I expand the model to include migration decisions. This is a period in India with well documented evidence of low migration rates despite wide regional-wage disparities (Munshi and Rosenzweig, 2009, 2016; Topalova, 2010), and empirically I fail to

¹⁹However, the distribution of earnings 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 models of variation in earnings (Roy, 1950) discuss the importance of ‘abilities’ ϵ_i , like “health, strength, skill, and so on.”

²⁰If expectations adapt slowly, we may even generate interesting dynamics in future cohort schooling decisions, like cobweb-style models (Freeman, 1975). Indeed, students in India seem to react to expected returns (Adukia et al., 2020; Khanna, 2020; Khanna and Morales, 2019).

detect a migration response.²¹ Theoretically, migration may affect how schooling subsidies translate into an increase in the size of the skilled workforce. Yet for much of the analysis, I examine how a given increase in the size of the skilled workforce reduces the skill-premium. If skilled wages fall, then skilled workers may emigrate, muting the outward shift in the supply of skilled workers, and mitigating any GE effects. As such, the GE effects are for a given change in the skill distribution. In the model extension detailed in Appendix B.VI, migration is an individual decision, and varies across cohorts and education level. The extension clarifies that migration changes the slope of the labor supply curve, even as one of my primary aims is to estimate the slope of the (relative) labor demand curve for a given shift in labor supply.

In Model Appendix B, I have a detailed discussion of the supply of both public and private schooling (Appendix B.I), and the model’s equilibrium is described in Appendix B.IV. Public schools depend on funds received from the government, R_d , and expand the supply of schooling in response. Private schools make an entry or exit decision based on expected profits. If the expansion of public schooling lowers equilibrium school fees, p_d , private schools may be crowded out. Alternatively, if demand externalities raise the demand for private schooling, or if public expansions lower entry-costs for private schools, they may be crowded in (Appendix B.I).

3 Data

I combine a number of large datasets, merged at the district level, which is 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. Data details can be found in Data Appendix D.

Data for school construction and inputs come from the District Information System for Education (DISE), covering about 1.45 million schools by 2012. I compile data at the school level for all waves between 2005 and 2012, including the number of schools, when they were built, whether public or privately owned, number of teachers by education level, and infrastructure. Table A.1 summarizes variables in 2005, at the end of DPEP. 27% of schools in 2005 were built post 1993, and while 20% are government, the remaining 7% are private schools.

To study educational outcomes, I use household surveys and Census data. I create a panel of districts using the 1991 Census as a baseline. District splits-and-merges are well documented by official Census crosswalks, so I normalize changes to districts to be at the 1991 level. The 1991 Census female literacy rate, used by policymakers, is the running variable for the RD.

I use three rounds of the National Sample Survey (NSS), between 2004 and 2010, to examine impacts 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 weekly earnings for each individual in the household. It also includes a detailed expenditures module that in-

²¹Many studies on India are explicit about ignoring migration in the main analysis as the numbers are low (Anderson, 2005; Banerjee et al., 2008; Deshingkar and Anderson, 2004; Duflo and Pande, 2007).

cludes educational expenditures. Summary statistics for the 2009 NSS round are presented in Table A.2. In 2009, only about 66% of the population finished primary school, and on average, people had 6.8 years of education and earned about \$37 a month. The 2009 NSS round is the first large-sample labor-force round after the end of DPEP, and has the added advantage of allowing time for students affected by the policy to enter the labor market. I restrict individuals to be between 17 and 75 years of age, but results are robust to relaxing this constraint.

To study firm behavior, I use a census of manufacturing firms from the Annual Survey of Industries (ASI), and obtain the District (Gross) Domestic Product from each state’s statistical office. Finally, I use the Annual Status of Education Report (ASER) to study impacts on test scores. ASER is collected annually by an NGO (Pratham), and surveys children between the ages of 3 and 16. These surveys are done at home on weekends, so as to capture school dropouts, and those who never attended schools. I use all available surveys between 2007 and 2012.

4 Estimation and Identification Using an RD

The DPEP targeted low-literacy districts. Districts that had a female literacy below the national average (based on the previous 1991 Census) were eligible for the program. However, not all such districts were selected, generating a fuzzy Regression Discontinuity design using the 1991 female literacy rate as a running variable. To my knowledge, no other programs use the district-level 1991 female-literacy rate as a cutoff.

As the RD cutoff was around the national average, there is a large density at the cutoff (Figure 1a), allowing for robust identification. As such, my estimates are representative of a (policy-relevant) average-literacy district induced into taking up treatment. Since we should not expect any discontinuity in the baseline distribution of individual labor-market abilities or costs of schooling around the cutoff, the RD estimator is consistent. Indeed, at the cutoff, we expect no discontinuity in pre-policy labor market characteristics, skill-biased capital and regional output. I show that cohorts that were too old to change their schooling by the time the policy was implemented have no discontinuity in educational attainment. In order to estimate the GE effects, I further exploit variation in cohort exposure and skill levels.

Since more able workers may also be 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, while identical students in non-policy regions do not. Following students into the labor market, I 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 1b. There is a sharp discontinuity at the cutoff which provides a causal estimate of 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 students who were induced into more schooling and lived in districts near the cutoff that

took up the policy. Similarly, the GE effects depend on what type of students get induced into 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, 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 here as the cutoff was the national average of the female literacy rate from the previous Census. McCrary (2008) tests indicate no discontinuity in the density of districts around the cutoff (Figure 1a), and the p-value of the change in density is 0.71. Importantly, Figure 1a also shows that there is a high density of districts at the national average, providing sufficient data for RD estimation. Other falsification tests discussed below justify the RD assumptions that there were no other discontinuities at the same cutoff.²²

While I present RD results graphically, the coefficients of interest are calculated using two different optimal bandwidths procedures, by Calonico et al. (2014) and Imbens and Kalyanaraman (2012) respectively. Imbens and Kalyanaraman (2012) use a data-driven bandwidth selection algorithm to identify the optimal bandwidth for a local linear regression given a squared loss function, whereas Calonico et al. (2014) perform a bias-correction and develop robust standard errors. Results using both bandwidth procedures are presented, and are robust to using other parametric approaches (Hahn et al., 2001; Imbens and Lemieux, 2008).²³

4.1 Using Policy Changes to Estimate Parameters

Variation in schooling, s_{id}^* , is driven 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 ($Cov(\eta_i, \epsilon_i) \neq 0$), and there are underlying baseline differences in the skill distribution and skill-biased capital across these markets (Equation (4)), OLS regressions of earnings on education will be biased. In Section 2 and Appendix B.I, we derived that the equilibrium amount of aggregate schooling in a district, S_d^* , is affected by schooling expansions:²⁴

$$S_d^* = \phi_1 \mathbb{E}[\tilde{\beta}_{asd}] + \phi_2 R_d - \frac{\eta_d}{\Gamma} \quad (10)$$

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.I). The term $\phi_1 \mathbb{E}[\tilde{\beta}_{asd}]$, captures how changes in expected returns to education affect equilibrium schooling. If, for instance, the labor-market GE effects are expected to substantially lower the returns to education $\mathbb{E}[\tilde{\beta}_{asd}]$, then there may be no

²²Cattaneo 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.

²³The results are robust to using various alternative procedures (Appendix Table A.12). Bartalotti and Brummet (2017) allows for standard errors at an aggregated level. Calonico et al. (2017) allows for different-sized optimal bandwidths on either side of the cutoff and for nearest neighbor standard errors.

²⁴See Appendix B.I for a parameterization of ϕ_1 and ϕ_2 , where $\phi_1 \equiv \left(\frac{\theta_d^2}{\Gamma \theta_d^2 + z_{2d}} \right) > 0$ and $\phi_2 \equiv \left(\frac{z_{2d} + \Psi \theta_d^2}{\Gamma \theta_d^2 + z_{2d}} \right) (\Pi_m \frac{\alpha_m}{p_m} \alpha_m) > 0$, and $\eta_d = \mathbb{E}[\eta_i | i \in d]$.

increase in the equilibrium amount of schooling. The final term $\frac{\eta_d}{\Gamma}$ (baseline residual schooling costs) is unaffected by schooling expansions. We would, however, expect it to be correlated with other unobserved district characteristics. Yet, η_d are likely 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 just on the other side. In the neighborhood of the cutoff, we should therefore expect, for $\phi D_d \equiv \phi_1 \mathbb{E}[\tilde{\beta}_{asd}] + \phi_2 R_d$:

$$S_d = \phi D_d - \frac{\eta_d}{\Gamma} \quad \text{and} \quad \mathbb{E}[\eta_d | D_d = 1] = \mathbb{E}[\eta_d | D_d = 0] \quad (11)$$

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

4.1.1 Returns to Education and Disentangling Earnings

The model derives equations 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), reproduced below, ψ_a captures the cohort effect.²⁵ θ_{sd} captures the pure productivity effect and changes in skill-biased capital. 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:²⁶

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

I leverage variation along various dimensions (age cohorts, skill levels and treatment status) to disentangle the components of the change in earnings across the RD cutoff. 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 education at policy implementation, will be partially affected by labor-market GE effects. The GE effects that affect all cohorts can thus be isolated by looking at the impact on the skill-premium of older cohorts.

Earnings for the young will additionally be affected by cohort-specific GE effects as there are more skilled workers particularly in younger cohorts. As the young and old are not perfect substitutes, and may be complements in production, I estimate effects for each cohort separately. While the estimates are informative for the model, such natural experiments also provide well-identified evidence. To that end, for unbiased estimation, I follow a few rules similar to Jackson's (2019) analysis of Difference-in-Regression Discontinuity designs. First, I compare outcomes across the RD cutoff; second, I compare the same cohorts across the cutoff (to account for cohort effects); and last I compare the same skill group across the cutoff.

For ease of exposition, I restrict the analysis to two skill levels – skilled s and unskilled u

²⁵For clarity, I repeat the same equation number every time the same equation appears in the text.

²⁶While 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, and young and old. The functional form is to better understand the role played by underlying economic parameters.

workers. The fraction of each among the young y are represented by ℓ_{sy} and ℓ_{uy} respectively. For any two-skill groups: $\Delta\ell_{sy} \equiv (\ell_{sy,D=1} - \ell_{sy,D=0}) = -\Delta\ell_{uy} \equiv (\ell_{uy,D=1} - \ell_{uy,D=0})$.

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:²⁷

$$\log \frac{w_{as,D=0}}{w_{au,D=0}} = \underbrace{\log \frac{\theta_{s,D=0}}{\theta_{u,D=0}}}_{\text{Productivity}} + \underbrace{\left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E} \right) \log \frac{L_{s,D=0}}{L_{u,D=0}}}_{\text{Aggregate skill distribution}} - \underbrace{\frac{1}{\sigma_A} \log \frac{\ell_{as,D=0}}{\ell_{au,D=0}}}_{\text{Cohort specific skill distribution}} \equiv \beta_{as,D=0}, \quad (12)$$

where $\beta_{as,D=0}$ are the earnings returns to changing one's 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:

$$\log \frac{w_{as,D=1}}{w_{au,D=1}} = \log \frac{\theta_{s,D=1}}{\theta_{u,D=1}} + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E} \right) \log \frac{L_{s,D=1}}{L_{u,D=1}} - \frac{1}{\sigma_A} \log \frac{\ell_{as,D=1}}{\ell_{au,D=1}} \equiv \beta_{as,D=1}, \quad (13)$$

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. $\Delta\beta_{as}$ captures the GE effects on the returns to skill:

$$\Delta\beta_{as} = \underbrace{\left(\log \frac{\theta_{s,D=1}}{\theta_{u,D=1}} - \log \frac{\theta_{s,D=0}}{\theta_{u,D=0}} \right) + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E} \right) \left[\log \frac{L_{s,D=1}}{L_{u,D=1}} - \log \frac{L_{s,D=0}}{L_{u,D=0}} \right]}_{\text{GE effects on all cohorts}} - \underbrace{\frac{1}{\sigma_A} \left[\log \frac{\ell_{as,D=1}}{\ell_{au,D=1}} - \log \frac{\ell_{as,D=0}}{\ell_{au,D=0}} \right]}_{\text{Additional GE on young}} \quad (14)$$

To estimate returns and the GE effects I will not need to estimate every economic parameter (like σ_A and σ_E). Such an approach is similar to ones taken in Public Economics (Harberger, 1954, 1964; Hotelling, 1938), sometimes measuring general equilibrium effects (Goulder and Williams III, 2003). Similar to what Heckman and Vytlacil (2007) refer to as the Marschak (1953) Maxim, I only need to identify a combination of economic parameters rather than every primitive. This is less demanding of the data, while allowing for clean identification.

In order to disentangle the GE effects by cohort, I 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 ensures that the differences in earnings across the RD cutoff are only due to differences in the skill distribution and the amount of skill-biased capital.

The change in returns in Equation (14) can be split up into two components. The first

²⁷There is no expectations operator as these are the actual aggregate returns, and not individual expectations.

is the GE effect that affects all cohorts. To estimate this effect, I examine 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:²⁸

$$\underbrace{\log \frac{w_{so,D=1}}{w_{so,D=0}} - \log \frac{w_{uo,D=1}}{w_{uo,D=0}}}_{\text{GE effects on all cohorts}} = \underbrace{\left(\log \frac{\theta_{s,D=1}}{\theta_{u,D=1}} - \log \frac{\theta_{s,D=0}}{\theta_{u,D=0}} \right)}_{\text{Skill biased capital}} + \underbrace{\left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E} \right)}_{\text{Aggregate skill distribution}} \left[\log \frac{L_{s,D=1}}{L_{u,D=1}} - \log \frac{L_{s,D=0}}{L_{u,D=0}} \right] \quad (15)$$

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.

The second component of the GE effects from Equation (14) 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:²⁹

$$\underbrace{\left[\log \frac{w_{sy,D=1}}{w_{sy,D=0}} - \log \frac{w_{uy,D=1}}{w_{uy,D=0}} \right]}_{\text{Additional GE on young}} - \underbrace{\left[\log \frac{w_{so,D=1}}{w_{so,D=0}} - \log \frac{w_{uo,D=1}}{w_{uo,D=0}} \right]}_{\text{Age specific skill distribution}} = -\frac{1}{\sigma_A} \underbrace{\left[\log \frac{\ell_{ys,D=1}}{\ell_{yu,D=1}} - \log \frac{\ell_{ys,D=0}}{\ell_{yu,D=0}} \right]}_{\text{Age specific skill distribution}} \quad (16)$$

Migration directly affects the quantities in Equations (15) and (16). For instance, 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. In this way, the model incorporates migration in determining the GE effects. The individual decision to migrate, and the corresponding labor supply curve, is specifically modeled in Appendix B.VI.

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:³⁰

$$\log \frac{w_{y,D=1}}{w_{y,D=0}} = \ell_{sy,D=1} \log \frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=1} \log \frac{w_{uy,D=1}}{w_{uy,D=0}} + \underbrace{\Delta \ell_{sy} \log \frac{w_{sy,D=0}}{w_{uy,D=0}}}_{\beta_{ys,D=0}} \quad (17)$$

$$\log \frac{w_{y,D=1}}{w_{y,D=0}} = \ell_{sy,D=0} \log \frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=0} \log \frac{w_{uy,D=1}}{w_{uy,D=0}} + \underbrace{\Delta \ell_{sy} \log \frac{w_{sy,D=1}}{w_{uy,D=1}}}_{\beta_{ys,D=1}} \quad (18)$$

The change in the average earnings for the younger cohorts is a weighted average of the change in the young skilled wage (weighted by the fraction skilled), the young unskilled wage

²⁸Regardless of the specific form of the production function, the change in the skill premium for the old will be the GE effects on all cohorts, and estimates of returns and cohort-specific GE effects will empirically rely on the left hand sides of equations (15) and (16). For instance, this is true even if wage returns are determined in a purely signaling model. The right hand sides merely help us understand the underlying economic parameters.

²⁹Notice that if $\sigma_A < \sigma_E$ then the two components may be of opposite signs.

³⁰See Appendix B.V for detailed derivations of these equations.

(weighted by the fraction unskilled), and the returns to skill (weighted by the fraction of compliers). 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 changes across the cutoff in the following manner:

$$\begin{aligned}\Delta S &= (\ell_{sy,D=1}s_1 + \ell_{uy,D=1}s_0) - (\ell_{sy,D=0}s_1 + \ell_{sy,D=0}s_0) \\ &= \Delta\ell_{sy}s_1 + \Delta\ell_{uy}s_0 = \Delta\ell_{sy}(s_1 - s_0),\end{aligned}\tag{19}$$

where s_1 is the years of education for the skilled group, s_0 the years for the unskilled group, and $\Delta\ell_{sy}$ is the fraction of students induced into getting more skill.³¹

4.2 Outcomes and Economic Benefits

The economic benefits depend on the changes in the wage distribution across the cutoff. Yet, the benefits to different types of workers depend on a few crucial elasticities, as in [Harberger \(1964\)](#).³² For instance, the labor market benefits to workers induced into getting more skill are the sum of partial equilibrium returns and the GE effect on skilled wages. This highlights the importance of estimating *both* parameters in order to measure economic benefits.

The first determinant of the changes in overall benefits is the reduction in costs of schooling for younger cohorts. This is particularly meaningful for infra-marginal students that were always going to attend school, even if the policy had not induced marginal students to get more education. Studies that focus on only the enrollment response in education interventions may miss this large component of benefits that affects infra-marginal students.

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 in the skill distribution. At the same time, the GE effects will have distributional consequences. The welfare of older cohorts is unaffected by the reduction in the costs of schooling. The skilled old however are adversely affected by GE effects that affect all skilled workers, whereas the unskilled old benefit from increases in their earnings.

If agents do not have perfect foresight, we need to distinguish between welfare *ex ante* and *ex post* of GE effects.³³ *Ex post* welfare depends on $\beta_{as,D=0}$, the actual returns to skill in untreated districts, and $\beta_{as,D=1}$ the returns including the GE effects. The *ex post* welfare for a young high-skill person that would acquire skill even in the absence of the policy rises by the reduction in the total costs of education, but is dampened by GE effects that affects all cohorts and additional GE effects on the young. Labor market welfare for them is $\log \frac{w_{as,D=1}}{w_{as,D=0}}$. In the absence of GE effects, the young always-skilled are only affected by reductions in education costs. For those who would never acquire skill, even in the presence of the policy, the difference

³¹Recall that the per-year rate of return is a product of the returns to skill and the education gap between the skilled and unskilled: $\tilde{\beta}_{asd} = \frac{\beta_{asd}}{(s_1 - s_0)}$.

³²Similar to what [Heckman and Vytlacil \(2007\)](#) refer to as the [Marschak \(1953\)](#) Maxim, I only need to identify a combination of economic parameters rather than every primitive. This puts less constraints on the data by focusing identification on certain parameters.

³³I measure labor and education market ‘welfare’ in log-monetary units, as they depend on costs and earnings.

in the unskilled wage at the cutoff is their *ex post* labor market welfare: $\log \frac{w_{au,D=1}}{w_{au,D=0}}$.

For younger cohorts, who are induced into getting more skill, the *ex post* labor market welfare change depends on the skilled wage in treated districts and the unskilled wage in untreated districts, $\log \frac{w_{ys,D=1}}{w_{yu,D=0}}$, and thereby the returns $\beta_{ys,D=0}$:

$$\log \frac{w_{ys,D=1}}{w_{yu,D=0}} = \log \frac{w_{ys,D=0}}{w_{yu,D=0}} + \log \frac{w_{ys,D=1}}{w_{ys,D=0}} = \beta_{ys,D=0} + \log \frac{w_{ys,D=1}}{w_{ys,D=0}} \quad (20)$$

As such, changes in *ex post* labor market benefits for those induced into getting more skill consist of two components: the partial equilibrium returns and the GE change in economic benefits to the ‘always skilled.’ This is why it is important to causally estimate both partial equilibrium returns and GE effects. In the absence of any GE effects, the change in earnings for a person induced into getting more education would simply be $\beta_{ys,D=0}$.

Ex ante welfare depends on expectations. If expectations were myopic, then wage expectations in regions that receive the policy $D = 1$, would equal steady state values in the absence of the policy (i.e. wages in untreated regions $D = 0$). As such, $\mathbb{E}[\beta_{as,D=1}] = \beta_{as,D=0}$. This *ex ante* myopic wage-return is the partial equilibrium *ex post* returns that I estimate. *Ex ante* change in welfare would be similar to the partial equilibrium *ex post* welfare mentioned above.³⁴

If agents had perfect foresight, then *ex ante* welfare changes would be the same as the *ex post* GE changes above. If expectations were neither myopic nor of perfect foresight, then *ex ante* welfare depends on how wages are expected to change. For the young always-skilled, this is the expected change in skilled wages, and for the young never-skilled this is expected changes in unskilled wages. For those induced into skill, welfare depends on how both skilled and unskilled wages are expected to change. I measure *ex post* welfare changes for different groups (by cohort and skill), and *ex ante* changes under perfect foresight and myopia.

To compare labor market gains to reduced schooling costs, and estimate welfare changes, I discount labor market gains by the real interest rate over the period. For a student induced into more education, the costs include tuition and the opportunity cost of foregone unskilled wages. The benefits include the present discounted value of a skilled worker’s earnings stream.

5 Results

5.1 Public and Private School Building

The primary objective of the program was to build new schools. Figure 1c shows the effect of the program on schools built once the program was underway in 1994. I trace out the longer-terms effects by studying how the coefficient in Figure 1d changes over time. The first coefficient plotted for the year 2005 shows a large discontinuity in the fraction of new schools, whereas other coefficients in later years show a smaller difference among districts on either side of the cutoff, as funding declined. In the absence of funds, regions on the untreated side of the cutoff catch up over time by building schools at a relatively more rapid rate. As a falsification

³⁴For the young always-skilled, myopic *ex ante* welfare change would just be the reduction in schooling costs. The young ‘never-skilled’ see no change in myopic *ex ante* welfare. For the young induced into skill, the myopic *ex ante* labor market welfare change is simply the partial equilibrium returns $\beta_{ys,D=0}$.

test, in Figure 1e and 1f, I show no differential impacts on the fraction of schools that were built in the twenty year period prior to the program (1973 to 1993).³⁵

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 driven by such a large-scale program. In Figure A.3b there is no evidence of crowd-out, and if anything, there is mild evidence of crowd-in.³⁶

5.1.1 School Quality

While the primary focus of DPEP was to increase educational attainment by building schools, there may have been quality improvements given the large funding and hiring of teachers. In Table A.3, I use the ASER data: of six different test score variables, only one (the ability to identify numbers) shows a statistically significant 5 percentage point increase. This is, at best, weak evidence of better test scores that may attenuate the negative GE effects. Either way, there is no detectable evidence on declining quality. On the other hand, better ‘quality’ in terms of better infrastructure may have made it easier for students to finish a grade and so further lower the marginal costs of schooling.

5.2 Education and Earnings

Since DPEP was introduced in 1994, we would not expect changes in education for students who were past school-going age at the time. In Appendix F.I, I discuss how cohorts are defined.³⁷ I find no detectable discontinuities in education, literacy, or different levels of education for older individuals in the left panels of Figures 2 and A.4, and top two panels of Tables 1 and A.6.

Table 1 and Figure 2 show RD impacts on education and earnings for those who reported earnings, across different bandwidths and age groups. The ITT estimates on the young are 0.7 more years of education, and a 0.11 log points increase in earnings. The estimates for the full sample in Table A.4 show that the young attain 0.217 more years of schooling. Treatment on the treated (TOT) estimates, scaled up by treatment probabilities, are in Table A.6. Older populations have no discontinuity in education and earnings. The old could still have GE effects, as *average* earnings conflate falls in skilled and rises in unskilled wages.³⁸

The program was targeted towards primary and upper primary levels, and we expect the

³⁵In Appendix Figure A.3 I split up the sample by government and private schools, and show alternative versions where I plot the total schools per capita, and the dynamic trends for old private schools.

³⁶What drives the crowd in? On the one hand, the demand externality could raise the equilibrium tuition and draw in private schools; on the other, a fall in operating costs may induce private school entry and lower the equilibrium tuition. In Appendix B.I.2, I discuss how I determine which mechanism is stronger by seeing how tuitions changes. Later in Section 5.4 – specifically Table A.14 – I show that household expenditure on schooling falls, suggesting that the cost-reduction mechanism is stronger.

³⁷As I show, results are robust to alternative age cutoffs. I show appendix tables with multiple age groupings and wider age restrictions. In Difference-in-Differences specifications I show impacts on each age separately.

³⁸I find some negative effects on average wages for close substitutes: cohorts close to treated cohorts. In Table A.7, the sample is sliced thin into more age groups, and even though they are imprecisely estimated, there do seem to be negative earning effects on 36 to 45-year-olds, the closest age group to the treated cohorts.

largest impacts at those levels. Table A.5 and the right panel of Figure A.4, show discontinuities in different levels of education for the young. Literacy rates are higher by 3 percentage points, and the likelihood of finishing primary school by 5.8 percentage points. Later, I define ‘skilled’ to be those who finished upper-primary, as that is where the program was targeted, where the largest effects are, and roughly divides the sample into equal halves. Table A.4 and Figure A.5 show analogous results for the full sample, rather than for those reporting earnings.³⁹

5.2.1 Heterogeneity by Gender, and Robustness

Since men may be likely to be in occupations that benefit from education, while women may be more likely to be engaged in domestic work, we may expect men to be more responsive to these interventions (Dreze and Sen, 2002; Kingdon, 1998). In Appendix Tables A.9 and A.10, I find the effects are concentrated among males, which is similar to the related literature (Ashraf et al., 2020; Breierova and Duflo, 2003; 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.⁴⁰

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 (Appendix Table A.11). I try more in-progress RD bandwidth selection procedures and standard error estimation methods in Table A.12, 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.7, and to age restrictions by including a larger sample of ages in Table A.13.

5.3 Returns to Education

5.3.1 OLS and Conventional IV 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. In general, IV-LATE estimates are found to be larger, perhaps as a reduction in marginal costs that affects all students equally will induce those with higher returns into getting more education (Carneiro et al., 2011; Imbens and Angrist, 1994; Oreopoulos, 2006).

Canonical IV-Wald methods estimate returns to education by using the RD cutoff to first

³⁹Comparing Tables 1 and A.4, 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.8 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. In the NSS, the probability that earnings are reported is uncorrelated with working in agriculture or being self-employed. 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.

⁴⁰One can add an additional nest to the CES production function, that captures differences across genders.

estimate the change in education, and then the corresponding change in earnings for the same cohort. By taking the ratio of the change in log earnings to the change in education, one estimates the returns to schooling. Under the assumption that the policy only induces some young workers to get more education, this method identifies the change in earnings due to an additional year of schooling for this marginal group. Yet, as my model stresses, the policy simultaneously affects both the skill premium and overall output in the district. Since changes in average earnings are not just driven by the switch in the fraction of students from unskilled to skilled groups, but also by the changes in skilled and unskilled wage, the estimated individual returns are conflated with changes in output and the skill premium.

The estimates in Table 1 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 1 shows the 2SLS-LATE version of this exercise. This estimate is not statistically indistinguishable from numbers as low as 7%, and lies within the range of comparable estimates found in the literature (Banerjee and Duflo, 2005; Psacharopoulos and Patrinos, 2004).⁴¹

The 2SLS return of 15.5% using the IV method described above ignores GE effects, and so is neither the partial equilibrium return, nor the return with the GE effects. As I show at the start of Section 5.3.3 below, it is a weighted average of both and lies in between.

5.3.2 Difference-in-Differences Designs

Difference-in-differences (DID) designs compare students across two dimensions: (1) whether the region received the policy, and (2) whether cohorts were young enough to change schooling. As I show in Appendix E.I, it is challenging to recover the partial (or GE) returns to schooling using an estimator that relies on ‘what happens to the old.’ I derive what a DID estimates, and the assumptions needed to derive meaningful parameters from DID.

There are at least two challenges with a DID estimator. The first is the ‘composition effect.’ The young and the old differ in the fraction of skilled workers. Due to the GE effects, skilled wages fall and unskilled wages rise. If there are more unskilled older workers, then the average wage for the old may actually rise. As such, the old may not be a good counterfactual for the young, as skill-compositions differ by cohort. The second reason is a ‘complementarity effect.’ The young and old are not perfect substitutes in production. More skilled young workers affect the wages of skilled older workers in different ways given the cohort elasticity of substitution σ_A . If they are complements, then the wage for skilled older workers may actually rise. Again, the old may not be good counterfactuals for the young.

In Table 2 and Figure A.6, I compare DPEP districts to non-DPEP districts, and the older cohorts to the younger cohorts.⁴² I estimate the difference-in-differences coefficient for three

⁴¹Recent experimental estimates of returns in a developing country are 13% (Duflo et al., 2017). While the bottom panel of Table 1 shows Log Earnings for the old; as there is no meaningful first stage response (in education), we should not interpret these as identified earnings returns.

⁴²For person i in age cohort a and district d , the following DID regression was estimated:

$$y_{ida} = \beta_{DID}T_{da} + \mu_d + \varpi_a + \epsilon_{ida} , \tag{21}$$

different subsamples. For the full sample, there is an increase in 0.3 years of education, and a 5.5 percentage point increase in literacy rates. 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 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, taking the ratio of the change in log earnings and years of education. This DID-2SLS “return” is 15.9%. It is also possible to measure how the skill-premium changes differentially for younger rather than older cohorts (an additional impact on the young). This component depresses the returns by 7.9 percentage points (Table 2).

5.3.3 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. Equation (17) described these returns:

$$\underbrace{\log \frac{w_{y,D=1}}{w_{y,D=0}}}_{\text{Avg earnings}} = \underbrace{\ell_{sy,D=1}}_{\text{Skilled}} \underbrace{\log \frac{w_{sy,D=1}}{w_{sy,D=0}}}_{\neq 0} + \underbrace{\ell_{uy,D=1}}_{\text{Unskilled}} \underbrace{\log \frac{w_{uy,D=1}}{w_{uy,D=0}}}_{\neq 0} + \underbrace{\Delta \ell_{sy}}_{\text{Compliers}} \underbrace{\log \frac{w_{sy,D=0}}{w_{uy,D=0}}}_{\beta_{ys,D=0}} \quad (17)$$

If there are no GE effects, then neither the skilled nor unskilled wage should change, and so for changes in partial equilibrium, $\log \frac{w_{uy,D=1}}{w_{uy,D=0}} = \log \frac{w_{sy,D=1}}{w_{sy,D=0}} = 0$. The average wage would rise only because more people go from earning a low unskilled-wage to a high-skilled wage ($\Delta \ell_{sy} > 0$). The change in average earnings across the RD cutoff would then 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, as I do in Section 5.3.1. 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 (17) are estimatable, allowing me to recover $\beta_{ys,D=0}$. Ignoring the GE effects produces an estimate (as in the bottom panel of Table 1) that lies between the partial and general equilibrium returns to skill.

Average earnings of all persons in treated districts are affected by changes in overall output. At the same time, the change skill distributions and the adoption of skill-biased capital affect skill-groups and cohorts differently, as in Equation (14). 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 4.1.1, and specifically, Equations (17) and (18) to derive the returns to education with and without the labor market general equilibrium effects. Scaling up by the treatment probability, there was a 17 percentage point increase in skilled workers across the cutoff (third panel of Table A.6).⁴³

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 parallel trends assumption, β_{DiD} is the DID parameter.

⁴³To reiterate, I define skilled workers as those finishing upper primary school, as the policy targeted this

As returns depend on shares of skilled and unskilled (Equation (17)), in Table 3 I bootstrap fifteen hundred draws with replacement, creating a null, jointly permutating the running variable, treatment status, and treatment probability. Appendix F.IV discusses the details. Estimated returns in the absence of GE effects are 19.9% per year (Table 3). The returns with GE effects, however, are only 13.4%, a 33% decrease in returns attributable to GE effects.

This change in the skill-premium can be split up into the portion that affects all cohorts, and additional impacts only on the young. To do this, I use Equations (15) and (16). Table 3 implies 85.3% of the change in GE effects are from the ‘additional impact on young’ term. The GE effect on all cohorts may be small as the two components that determine this effect counteract each other – an increase in the relative supply of skilled workers $\left(\log \frac{L_{s,D=1}}{L_{u,D=1}}\right)$ lowers the skill premium, but adoption of skill-biased capital $\left(\log \frac{\theta_{s,D=1}}{\theta_{u,D=1}}\right)$ increases the premium.

Furthermore, the additional impacts on the young is high, implying that the young and old are not close substitutes in production. Looking only at the GE effects on older cohorts considerably understates the GE effects on the young, and my paper is among the first to estimate the GE effects on all cohorts. Even as estimating the GE effects do not depend on the specifics of the model, they allow us to recover reliable 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.

5.3.4 External Validity, Compliance, and Extrapolating Away from the Cutoff

There are two issues related to external validity. The first is related to the difference between complier and non-complier districts (the RD is fuzzy), and the second, is about extrapolating away from the RD cutoff. Appendix E.II has a comprehensive discussion of these issues. The tests, with caution, lend support to the external validity of my estimates.

First, Appendix E.II.1 tests for differences in outcomes between treated complier districts and always-takers, and between untreated compliers and never-takers. Following Bertanha and Imbens (2019), in Figure E.1 and Table E.1 I test for discontinuities in education: (a) conditioning on receiving the program, and (b) conditioning on not receiving it. Then, in Appendix E.II.2, I estimate the Marginal Treatment Effect (MTE) across the distribution of the unobserved net costs of being treated, and find that the MTE is relatively stable, suggesting the lack of meaningful treatment effect heterogeneity (Brinch et al., 2017). I also test for the size of the first stage by different district characteristics in Appendix E.II.3, and fail to detect meaningful differences in compliance probability across different occupations and industries, suggesting that baseline economic characteristics are unlikely to drive compliance probabilities.

Second, I discuss effects away from the RD cutoff. The advantage of my context is that the Indian government chose the cutoff to be the national average of female literacy, and so my estimates are already relevant for an attractively representative set of districts. In Appendix

level, and because the largest earnings increase in OLS regressions on untreated districts comes at finishing upper primary school. In going from literate-below primary to finishing primary school, average earnings increase by 10%, whereas in going from primary to upper primary average earnings increase by 20%.

Section E.II.4, I closely follow [Dong and Lewbel \(2015\)](#), who estimate the marginal threshold treatment effect (MTTE), or how the treatment effect changes from a marginal change in the RD cutoff. The results in Table E.3 show that, at least locally, the MTTE is statistically indistinguishable from zero and economically small. These results suggest that locally extrapolating around the cutoff would produce minor changes to the estimated treatment effects. Yet, we should be cautious in extrapolating this to extremes of the running variable.

5.3.5 Returns vary across levels of schooling

Following [Lochner and Moretti \(2015\)](#), in Appendix E.III I examine returns across different levels of education. The OLS estimates in Figure E.4 suggest that other than at the college-level, there are no substantial non-linearities across different levels of education. [Montenegro and Patrinos \(2014\)](#) find that in India, the annual OLS returns are higher for college graduates (20.8%), than for primary (5.8%) and secondary (6%) students (which have similar returns). Figure E.4 reflects this, but makes another important point – the 2SLS weights on college are 0, but are high on finishing middle (upper primary school); perhaps as the policy was targeted and mainly induces children to finish upper primary school. As such, the returns we estimate are not capturing returns to college.

5.3.6 Labor Market Distortions

How do labor market frictions affect the analysis at hand? Let us consider two types of frictions described in the literature: (i) misallocation of factor inputs across sectors (or firms), and (ii) monopsonistic labor markets ([Muralidharan and Niehaus, 2017](#)). In Appendix B.VII I discuss in detail the consequences of the former distortion. If there are J sectors, with factor misallocation, then the wages are no longer the same across sectors. I show that much of the analysis is unaffected, and derive the conditions under which the estimated returns are simply a weighted average of sector-specific returns under certain conditions. In Appendix B.VIII, I examine the consequences of market power in the labor market. While market power does not affect the private returns to workers, it will affect the impacts on productivity and output. To elaborate, the private returns to workers will continue to depend on the skill-premium, but the difference in the skilled and unskilled wage no longer captures the difference between skilled and unskilled productivity. I derive an adjustment term that depends on the *relative* (between skilled and unskilled) labor supply elasticities, allowing us to identify the productivity consequences of upskilling given known elasticities. As some recent development literature suggests little difference in labor supply elasticities across sub-groups ([Goldberg, 2016](#)), this adjustment term is likely to be small.

5.4 Total Output, Consumption, and Educational Expenditure

Changes in overall output depend on the productivity of different skill levels and the shift in the labor force across skill levels. As workers acquire skill, and/or if skill-biased capital is adopted, overall productivity and output in the region increases. The top panel of Figure A.7 shows the impact on the District Domestic Product. These regressions are underpowered and

standard errors large. The (imprecise) point estimates indicate that in 2000-04, the increase in GDP associated with the policy was between 0.068 and 0.11 log points (Table A.15).

Changes in total output lead to changes in total consumption. The top panel of Table A.14 shows 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.⁴⁴ This fall in educational expenditure is persistent even five years after the program ended, driven by lower spending on tuition and fees (bottom panel of Appendix Figure A.7). There is, in fact, an increase in complementary expenditures, like books and stationery (Table A.14).

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 younger cohorts who attend school. The fall in the costs of schooling even benefit inframarginal households, who would always get education, even in the absence of the policy, and even as they may not change their enrollment decisions.

5.5 Migration

Local economies receiving 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 utilize their increase in skill and adopt new technologies in production. The adoption of skill-biased capital, therefore, increases the returns to skill, and affects the GE effects.

The migration-model extension in Appendix B.VI explicitly allows migration to change quantities of labor in Equation (14). Worker mobility tends to equalize wages across regions and mitigate negative GE effects on the skilled or positive GE effects on the unskilled. This would attenuate the GE effects, as skilled workers would leave once their wages fell.⁴⁵ Yet, permanent worker migration was extremely low in India, making it unlikely that newly skilled workers emigrate from these districts (Munshi and Rosenzweig, 2009, 2016; Topalova, 2010).⁴⁶

By analyzing the NSS 2007-8 waves, which asks questions on migration, I find 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. The migration-model extension in Appendix B.VI also derives an intuitive result: the change in migration rates de-

⁴⁴Jalan and Glinskaya (2013) measure a 20-40% fall in household educational expenditure.

⁴⁵Regions around the RD cutoff are geographically dispersed, making the migration of capital or workers across cutoff regions less likely. Indeed, the little flows that occur were to big cities, and not other poor areas.

⁴⁶Many studies on India ignore migration as the numbers are low (Anderson, 2005; Banerjee et al., 2008; Das-Gupta, 1987; Duflo and Pande, 2007; Foster and Rosenzweig, 1996). Munshi and Rosenzweig (2009) show that cross-village migration rates were only 8.7%, most of which perhaps takes place within the district. Deshingkar and Anderson (2004) show that rural-urban migration is much lower in India than other countries. Munshi and Rosenzweig (2016) show that worker migration is extremely low despite large regional wage gaps.

pend on migration probabilities conditional on skill. Since empirically, $P(\text{migrate}|\text{skilled}) - P(\text{migrate}|\text{unskilled}) \approx 0$ and low for work-related migrants, we may not expect to see a migration response. Panel B of Appendix Table A.8 shows that the policy did not meaningfully impact either any migration or work-related migration, for either the young or old cohorts.

5.6 Productivity, Capital Adoption, and District Heterogeneity

Local enterprises may adopt newer technologies or capital in response to the policy (Ghani et al., 2015), as I discuss in Appendix F.III. The empirical results using the ASI data are in Figure F.1. Even at the establishment level, average wages paid to workers increase as more educated workers start joining the labor market around 2004. Furthermore, I classify firms based on their products as ‘high-skill’ firms. The figure shows a steady increase in the fraction of firms that produce more mechanized products. This suggests that either existing firms shifted production and employed high-skill workers, or newer firms entered and hired these skilled workers. Both are suggestive of the adoption of skill-biased capital in these regions.

In general, there are some clear changes to the labor market for the workers in these regions. The bottom half of Appendix Table A.8 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).

In addition, it is important to note that the returns are not an immutable parameter, but rather an endogenous outcome tied to local skill composition, productivity and capital. In Appendix F.II, I discuss heterogeneity in returns across district and labor market characteristics, such as the share employed in different sectors, and strength of the labor market.

5.7 Overall Economic Benefits

The program raised aggregate economic benefits, but had meaningful distributional consequences. 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 3 shows the returns by skill group, which helps back out the parameters and the changes in yearly labor market benefits shown in Table 4. These estimates depend on the TOT effects on earnings, scaled up by the probability of treatment. For these calculations, I use the average real interest rate during that period (5%). A gap of 10 years is assumed between when the costs of education are borne and labor market returns are realized. Finally, I distinguish between welfare *ex post* or *ex ante* of GE changes, as in Section 4.2.⁴⁷

In the top panel of Table 4, I present results for those in younger cohorts that were induced into getting more skill because of the policy. This is about 17% of the young population. Their *ex post* welfare (or *ex ante* perfect foresight welfare) increases by 0.12 log points. The GE effects depress this increase in *ex post* welfare by 23.8%. As such, *ex post* welfare in partial

⁴⁷The 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.

equilibrium or *ex ante* myopic welfare would have increased by 0.157 log points.

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. This is their *ex post* GE (or *ex ante* perfect foresight) labor market welfare change. Whereas for workers who were always going to be unskilled, their *ex post* labor market welfare (which is also their *ex ante* perfect foresight welfare) rises by 0.014 log earnings points. *Ex post* welfare in partial equilibrium (or *ex ante* welfare for those with myopic expectations) would not change for these groups.⁴⁸

Since unskilled workers are better off and skilled workers worse off *ex post*, I estimate the transfer in labor-market benefits from the skilled to the unskilled due to the GE effects. Among the old this transfer is 0.007 log points, and among the young 0.05 log points. An analysis that ignores GE effects would overestimate the welfare on the young induced-into-skill by 23.8%, and miss the transfer in welfare from the always-skilled to the never-skilled for each cohort.

To measure the lifetime welfare change for students induced into more schooling, I compare the cost of an additional year of schooling to the benefits in Table 4. These costs include not just tuition fees but also the opportunity cost of foregone unskilled wages. The benefits are the present discounted value of the skilled earnings stream. Everyone benefits from increases in overall output, and the educated young benefit from reductions in schooling costs.

6 Conclusion

Large-scale education investments can and do generate substantial GE 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 a schooling-expansion program, which measurably increased educational inputs, and years of education and earnings for students. Leveraging the policy, I estimate the parameters of a GE model using an RD design. My estimates imply that the per-year return to acquiring skill is 13.4%; or 6.6 percentage points lower than it would be in the absence of GE 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 costs of education and increases in local economic output.⁴⁹

These findings have two important implications. First, the results in this paper help explain why scaled up government policies may have different impacts than researcher-led micro interventions (Acemoglu, 2010; Deaton, 2010; Heckman et al., 1999). Identifying who

⁴⁸Note 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.

⁴⁹These results do not necessarily indicate that the policy was cost effective. I have shown that the direct impacts were concentrated on certain cohorts, and had low persistence. Given the large amounts of funds invested, the overall cost effectiveness of this policy is questionable, and is left for future research.

benefits and who does not, and what works and what does not, is key to making such large-scale infrastructure investments more targeted and effective. The methods in this paper can be used to quantify the expected impacts of scaled-up micro-interventions in other contexts. It is clear that understanding all the consequences of large GE effects is crucial for both researchers and policy-makers when considering nation-wide interventions in public policy.

Second, it speaks to the large body of literature using large-scale variation (tuition subsidies, compulsory schooling laws, distance based measures, and school building) to estimate the returns to education. 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.

While the methodology I develop is applicable to other similar settings across the world, my estimates are for local labor markets and not for the entire country. As such, they are not generalizable to regions further away from the RD cutoff in the absence of stronger assumptions. Yet, the great advantage of the RD cutoff is that it was for districts around the average female literacy; therefore, we should think of these results as pertaining to the average district.

The debates about the role of the government in education investments usually center on the economic benefits of the policy. I show that economic benefits to households 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, education can have other welfare consequences as well, such as better health or more informed political participation (Clark and Royer, 2013; Sen, 1999). Exploring these relationships is left for future research.

References

- Abbott, B., Gallipoli, G., Meghir, C., and Violante, G. L. (2019). Education Policy and Intergenerational Transfers in Equilibrium. *Journal of Political Economy*, 127(6):2569–2624.
- Acemoglu, D. (2010). Theory, General Equilibrium, and Political Economy in Development Economics. *Journal of Economic Perspectives*, 24(3):17–32.
- Adukia, A. (2016). Sanitation and Education. *American Economic Journal: Applied Economics*.
- Adukia, A., Asher, S., and Novosad, P. (2020). Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction. *American Economic Journal: Applied Economics*, 12(1):348–76.
- 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.
- 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.
- Ashraf, N., Bau, N., Nunn, N., and Voena, A. (2020). Bride Price and the Returns to Education for Women. *Journal of Political Economy*, 128(2).
- 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., 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.
- Bertanha, M. and Imbens, G. (2019). External Validity in Fuzzy Regression Discontinuity Designs. *Journal of Business and Economic Statistics*, pages 1–39.
- Bianchi, N. (2016). The Effects of Educational Expansions: Evidence from a Large Enrollment Increase in STEM Majors. *Mimeo*, Northwestern.
- 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*.
- Breza, E., Kaur, S., and Krishnaswamy, N. (2019). Propping Up the Wage Floor: Collective Labor Supply Without Unions. *Working Paper*.
- Brinch, C., Mogstad, M., and Wiswall, M. (2017). Beyond LATE with a Discrete Instrument. *Journal of Political Economy*, 125(4):985–1039.

- Bryan, G. and Morten, M. (2018). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy*.
- 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. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326.
- Card, D. (1999). The Causal Effect of Education on Earnings. *Handbook of Labor Economics*, 5:1801–63.
- 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., Das, J., and Reis, H. (2019). The Value of Private Schools: Evidence from Pakistan. *Working Paper*.
- 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.
- Clark, D. and Royer, H. (2013). The Effect of Education on Adult Mortality and Health: Evidence from Britain. *American Economic Review*, 103(6):2087–2120.
- 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.
- Dong, Y. and Lewbel, A. (2015). Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models. *Review of Economics and Statistics*, 97(5):1081–1092.
- Dreze, J. and Sen, A. (2002). India. Development and Participation. *Oxford University Press*.
- 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*.

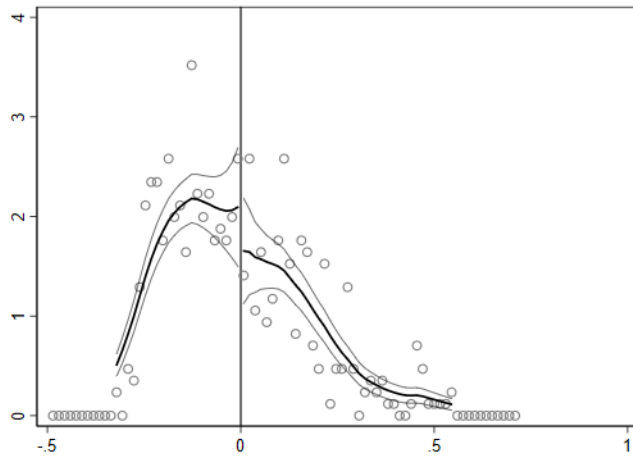
- Duflo, E., Hanna, R., and Ryan, S. (2012). Incentives Work: Getting Teachers to Come to School. *American Economic Review*, 102(4):1241–78.
- Duflo, E. and Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, 122(2):601–646.
- Egger, D., Haushofer, J., Miguel, T., and Walker, M. (2021). General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya. *Econometrica*. forthcoming.
- 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/eoj.12207.
- GOI (1994). *DPEP Principles*, volume New Delhi. Government of India Report.
- Goldberg, J. (2016). Kwacha Gonna Do? Experimental Evidence about Labor Supply in Rural Malawi. *American Economic Journal: Applied Economics*, 8(1):129–49.
- Goulder, L. H. and Williams III, R. C. (2003). The Substantial Bias from Ignoring General Equilibrium Effects in Estimating Excess Burden, and a Practical Solution. *Journal of Political Economy*, 111(4):898–927.
- 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.
- Harberger, A. (1954). Monopoly and Resource Allocation. *American Economic Review*, pages 77–87.
- Harberger, A. (1964). The Measurement of Waste. *American Economic Review*, 54(3):58–67.
- Hastings, J., Neilson, C., and Zimmerman, S. (2018). The Effects of Earnings Disclosure on College Enrollment Decisions. *NBER Working Paper No. 21300*.
- Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6):1251–71.
- Heckman, J., LaLonde, R., and Smith, J. (1999). The Economics and Econometrics of Active Labor Market Programs. *Handbook of Labor Economics*, 3A:1856 – 2097. Ashenfelter and 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.
- Heckman, J. J., Lochner, L. J., and Taber, C. (1998c). Tax Policy and Human-Capital Formation. *American Economic Review*, 88(2):293–297. Papers and Proceedings.
- Heckman, J. J. and Vytlacil, E. C. (2007). Econometric Evaluation of Social Programs, Part II: The Marginal Treatment Effect to Organize Alternative Econometric Estimators To Evaluate Social Programs, and to Forecast Their Effects in New Environments. *Handbook of Econometrics*, 6B.
- Hotelling, H. (1938). The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates. *Econometrica*, 6(3):242–269.
- 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.
- Jackson, C. K. (2019). Can Introducing Single-sex Education into Low-performing Schools Improve Academics, Arrests, and Teen Motherhood? *Journal of Human Resources*. forthcoming.
- Jagnani, M. and Khanna, G. (2020). The Effects of Elite Public Colleges on Primary and Secondary Schooling Markets in India. *Journal of Development Economics*. Forthcoming.
- 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.
- Kaur, S. (2019). Nominal Wage Rigidity in Village Labor Markets. *American Economic Review*, 109(10):3585–3616.
- Kazianga, H., Levy, D., Linden, L., and Sloan, M. (2013). The Effects of Girl-Friendly Schools: Evidence from the BRIGHT School Construction Program in Burkina Faso. *American Economic Journal: Applied Economics*, 5(3):41–62.
- Khanna, G. (2020). Does Affirmative Action Incentivize Schooling? Evidence from India. *Review of Economics and Statistics*, 102(2):219–233.
- Khanna, G., Liang, W., Mobarak, M., and Song, R. (2019a). The Productivity Consequences of Pollution-induced Migration in China. *Working Paper*.
- Khanna, G. and Morales, N. (2019). The IT Boom and Other Unintended Consequences of Chasing the American Dream. *Working Paper*.
- Khanna, G., Theoharides, C., and Yang, D. (2019b). Abundance from Abroad: Migrant Earnings and Economic Development in the Philippines. *Working Paper*.
- Kingdon, G. (1998). Does the Labour Market Explain Lower Female Schooling in India. *Journal of Development Studies*, 35(1).
- Kowalski, A. (2019). Behavior within a Clinical Trial and Implications for Mammography Guidelines. *NBER Working Paper, No. 25049*.
- 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.
- Lochner, L. and Moretti, E. (2015). Estimating and Testing Models with Many Treatment Levels and Limited Instruments. *Review of Economics and Statistics*, 97(2):387–397.
- 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.
- Marschak, J. (1953). Economic Measurements for Policy and Prediction. in *W. Hood, T. Koopmans (eds.), Studies in Econometric Method*, pages 1–26. Wiley, New York.

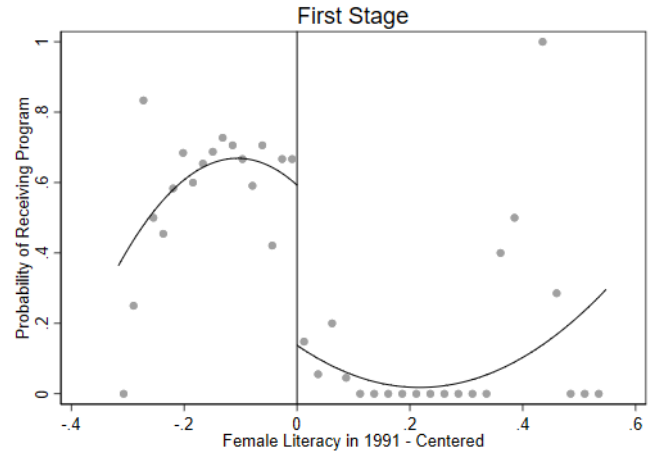
- McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*, 142(2):698–714.
- Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4):281–302.
- Montenegro, C. and Patrinos, H. (2014). Comparable Estimates of Returns to Schooling Around the World. *World Bank Policy Research Working Paper*, (7020).
- 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. (2016). Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap. *American Economic Review*, 106(1):46–98.
- Muralidharan, K. (2013). *Priorities for Primary Education Policy in India's 12th Five-year Plan*. NCAER-Brookings India Policy Forum.
- Muralidharan, K. and Niehaus, P. (2017). Experimentation at Scale. *Journal of Economic Perspectives*, 31(4):103–24.
- Muralidharan, K., Niehaus, P., and Sukhtankar, S. (2017). General Equilibrium Effects of (Improving) Public Employment Programs: Experimental Evidence from India. *NBER Working Paper w23838*.
- Muralidharan, K. and Prakash, N. (2017). Cycling to School: Increasing Secondary School Enrollment for Girls in India. *American Economic Journal: Applied Economics*, 9(3):321–50.
- 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.
- 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 \(2008\)](#) Test, the First Stage, and School Building



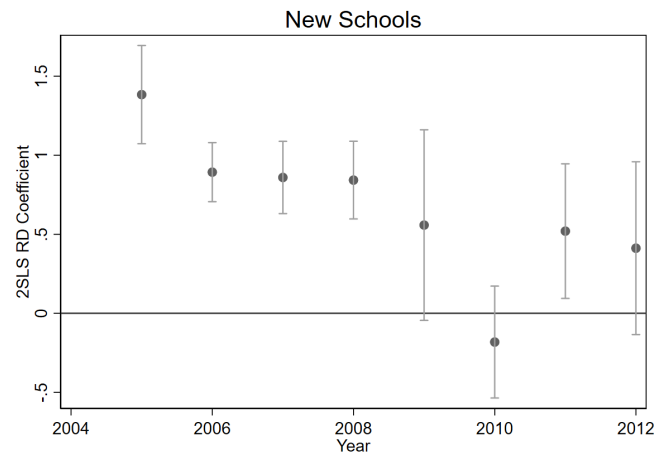
(a) District Level McCrary Test



(b) First Stage of DPEP



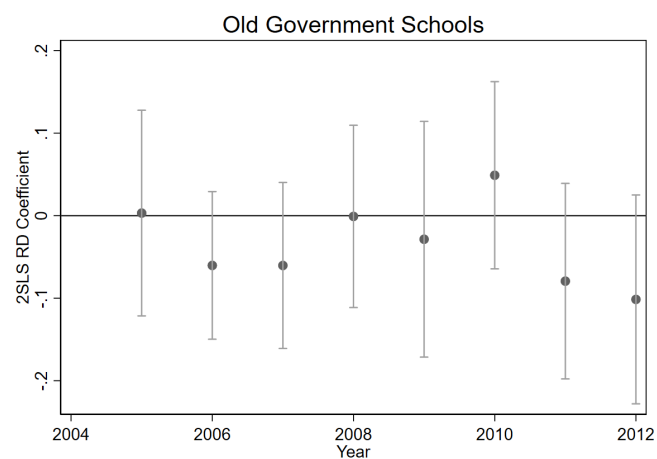
(c) Fraction of All Schools Built Post 1993



(d) Frac New Schools (2SLS) Over Time



(e) Total Number of Old Gov Schools (pre 1993)



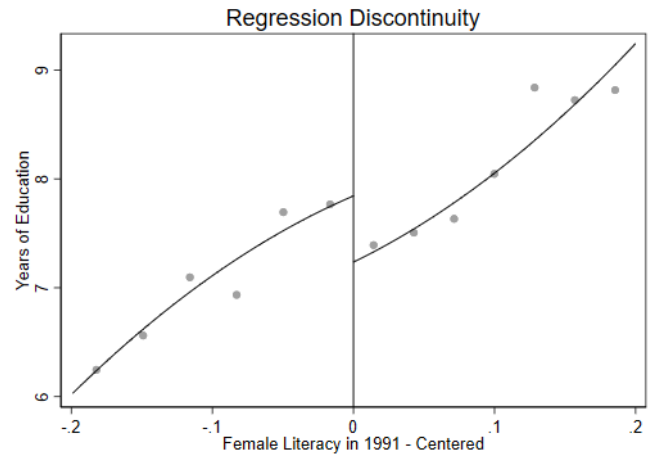
(f) Fraction Old Government Schools (1973-93)

Figure 1a is the [McCrary \(2008\)](#) test for discontinuity in density at the cutoff. Figure 1b is the first stage graph showing probability that a district received DPEP funds, using [Calonico et al. \(2017\)](#). Figures 1c - 1f use the DISE 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. \(2014\)](#) procedure.

Figure 2: RD Years of Education and Earnings



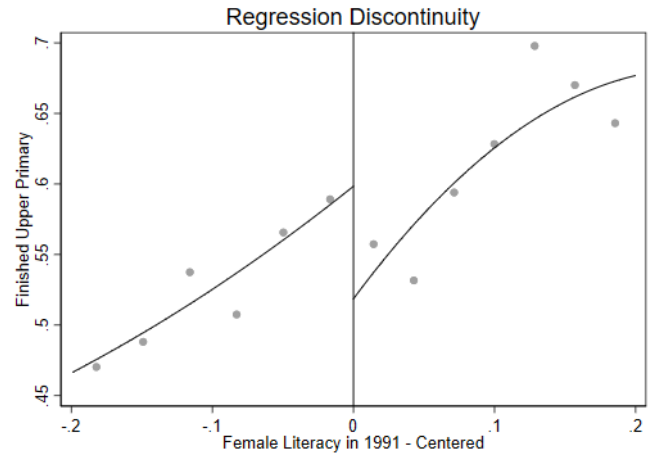
Years of Education - Older (Placebo)



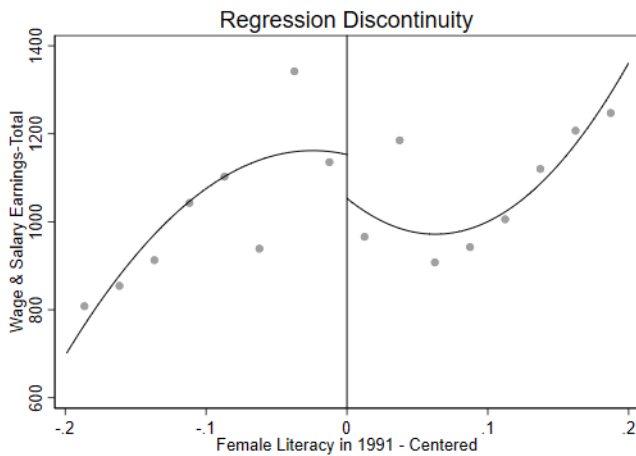
Years of Education - Younger



Upper Primary - Older (Placebo)



Upper Primary - Younger



Earnings (Rupees) - Younger



Log(Earnings) - Younger

National Sample Survey 2009 for those who reported earnings. Figures made using [Calonico et al. \(2014\)](#) 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: Education and Earnings for those with Reported Earnings

Panel A: First Stage				
Years of Education	Young	Old	Young	Old
RD Estimate	0.720 (0.199)***	-0.0856 (0.218)	0.698 (0.173)***	0.100 (0.188)
Observations	10,175	11,293	14,277	16,007
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper-primary				
RD Estimate	Young	Old	Young	Old
RD Estimate	0.0743 (0.0197)***	-0.0169 (0.0184)	0.0733 (0.0169)***	0.000212 (0.0158)
Observations	9,045	7,729	10,175	9,920
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Panel B: Reduced Form				
Log Earnings	Young	Old	Young	Old
RD Estimate	0.112 (0.0312)***	-0.0114 (0.0372)	0.145 (0.0269)***	0.0432 (0.0318)
Observations	10,175	11,293	14,277	16,007
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Panel C: 2SLS IV-LATE Conventional Method Returns				
Log Earnings	Young	Old	Young	Old
Years of Education	0.155 (0.0427)***	0.129 (0.303)	0.208 (0.0460)***	0.442 (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. \(2014\)](#) 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 2: Difference-in-Differences (Full Model)

Full Sample	Years of Education	Literate	Finished Primary	Finished Upper Primary
Estimate	0.332*** (0.0927)	0.0551*** (0.00783)	0.0386*** (0.00754)	0.0196** (0.00898)
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*** (0.106)	0.0426*** (0.00764)	0.0302*** (0.00834)	0.0209** (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** (0.155)	0.0558*** (0.0111)	0.0410*** (0.0119)	0.0299** (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** (0.0251)		0.159*** (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** (0.0281)	0.0183 (0.0211)	-0.0794** (0.0320)	
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 3: Returns, and Wage Parameters

	Fraction Switched	Change in Returns $\Delta\tilde{\beta}$	
Estimate	0.171	-0.066	
SE	(0.045)	(0.030)	
	Returns without GE $\tilde{\beta}_{y,D=0}$	Returns with GE $\tilde{\beta}_{y,D=1}$	% Change in returns
Estimate	0.199	0.134	-33%
Bootstrapped p-val	[0.00]	[0.00]	
$\Delta\tilde{\beta}$	Change for older cohorts	Additional on Young	% Change on young
	-0.0097	-0.057	85.3%

National Sample Survey 2009-10. The estimation follows the procedures described in the Model section 4.1.1, and detailed in Appendix B.V, specifically Equations (14), (17) and (18).

The bootstrapping procedure, and different components underlying the table estimates are described in detail in Appendix F.IV.

Younger cohorts are those between 17 and 35, whereas older cohorts are between 36 and 50.

P-values for returns with GE $\tilde{\beta}_{y,D=1}$ and returns without GE $\tilde{\beta}_{y,D=0}$ were bootstrapped using 1500 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 4: 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
0.120	0.157	-23.8%	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
0.007	0	0.052	0
Change in Lifetime Welfare for Induced Students			
Costs	Benefits	Net	% Change (due to GE)
5.15	6.55	1.4	-23.8%

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.

‘With GE’ is *ex post* welfare when GE effects are included. It also tells us *ex ante* welfare for those with perfect foresight and full information on returns. ‘Without GE’ is *ex post* welfare if there were no GE effects. It also tells us *ex ante* welfare for those with perfectly myopic expectations and know the wage distribution at the time they make schooling decisions.

‘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%.

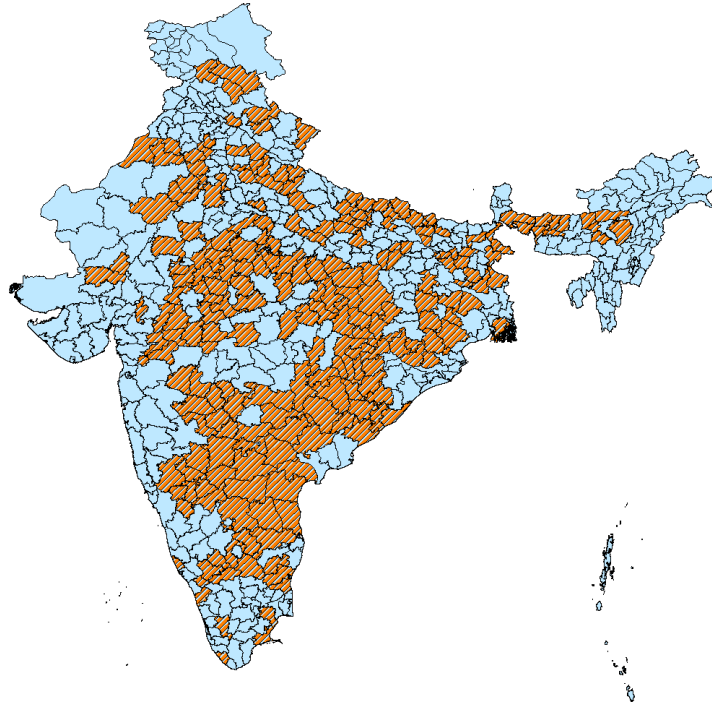
‘Change in Lifetime Welfare for Induced Students’: Costs include (a) opportunity cost of foregone earnings for unskilled work, and (b) tuition costs for students in DPEP districts near the cutoff. Costs are calculated in 2004 (NSS 61st round).

‘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

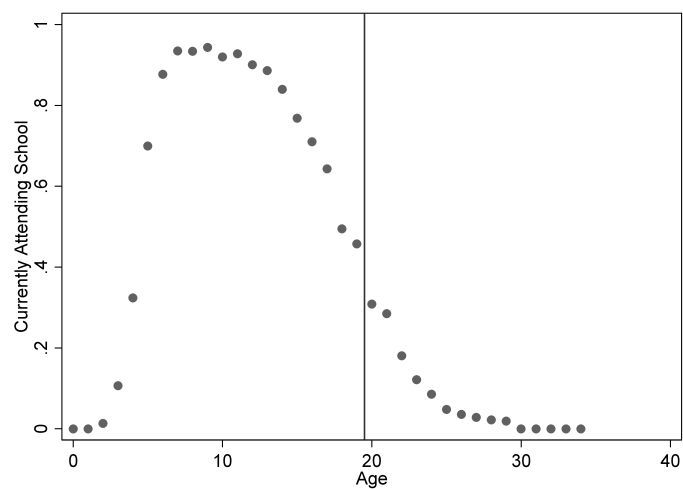
A Additional Tables and Figures

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

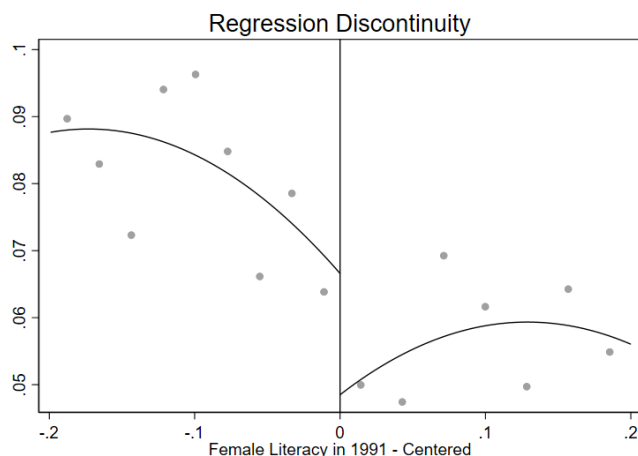


National Sample Survey 2009. The largest drop in school enrollment occurs between the ages of 19 and 20 - representing a 15 percentage point fall.

Figure A.3: Public and Private Schools, Old and New Schools



(a) Fraction of Government schools Built Post 1993



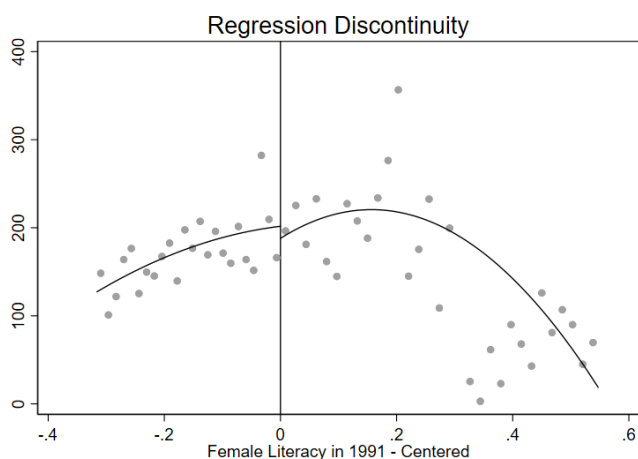
(b) Fraction of Private schools built post 1993



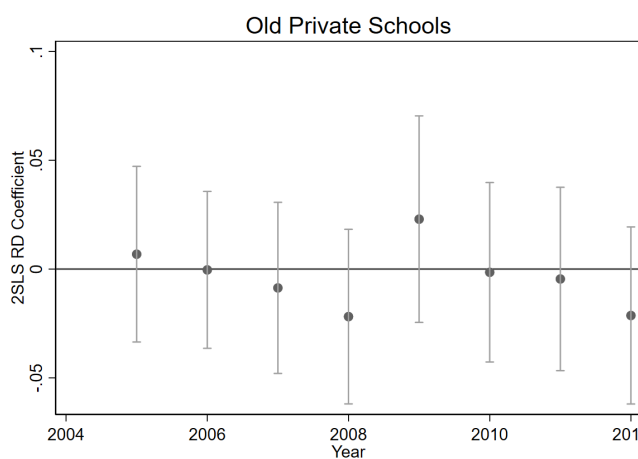
(c) Total Schools (per cap) Built Post 1993



(d) Total Government Schools (per cap) post 1993



(e) Total Number of Old Schools (built pre-1993)



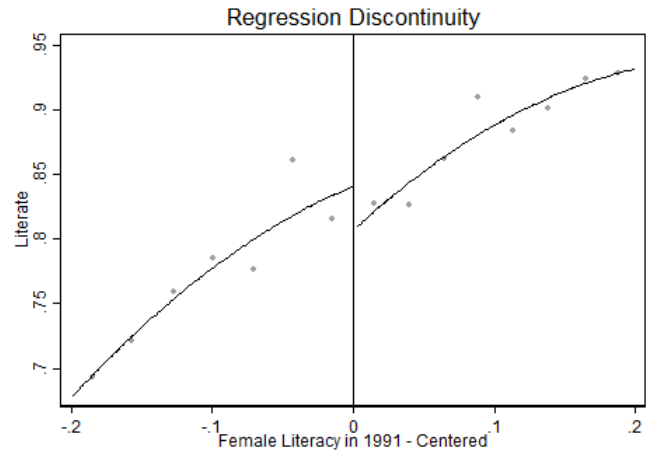
(f) Frac of Private Schools that are 'old' (1973-93)

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. \(2014\)](#) procedure. 'per cap' figures normalized by total population in district.

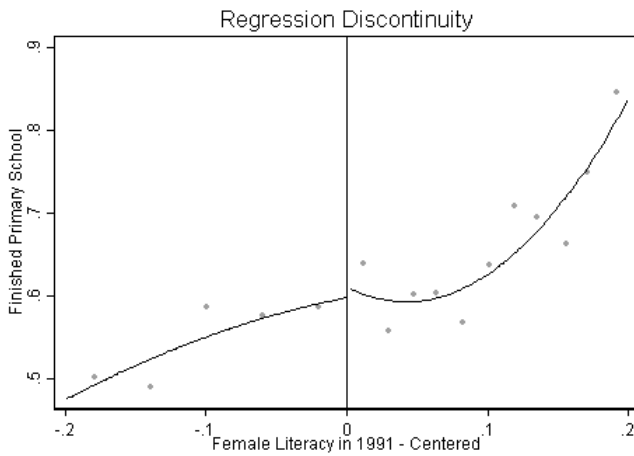
Figure A.4: RD figures - Levels of Education



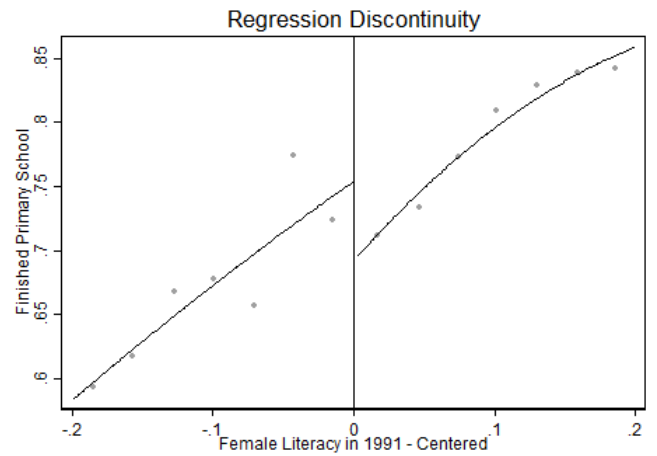
Literate - Older (Placebo)



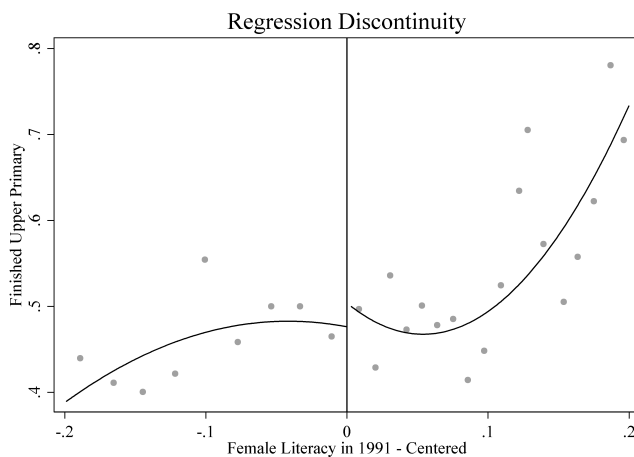
Literate - Younger



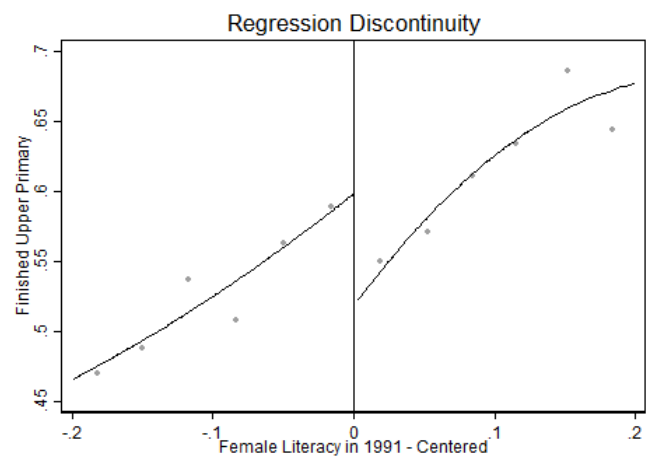
Finished Primary - Older (Placebo)



Finished Primary - Younger



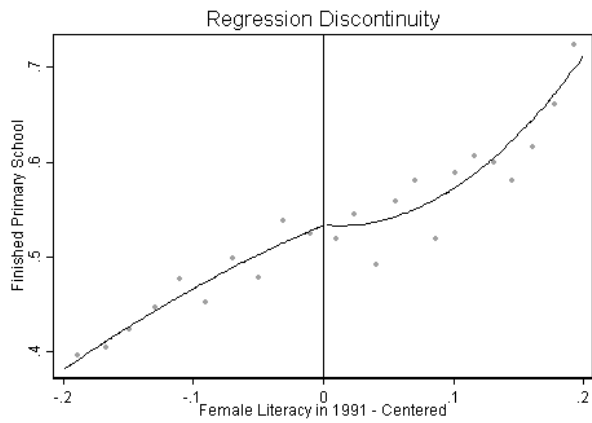
Upper Primary - Older (Placebo)



Upper Primary - Younger

National Sample Survey 2009 for persons who report earnings in primary occupation. Appendix Figure A.5 shows the analogous graphs for the full sample of persons. Figures made using [Calonico et al. \(2014\)](#) 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.

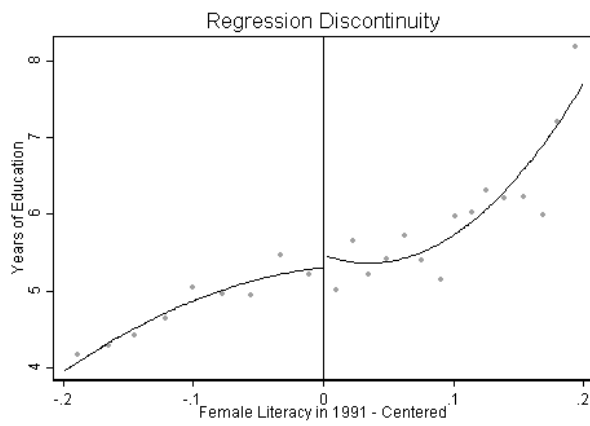
Figure A.5: RD Figures - Education - Full Sample



Finished Primary School - Old



Finished Primary - Young



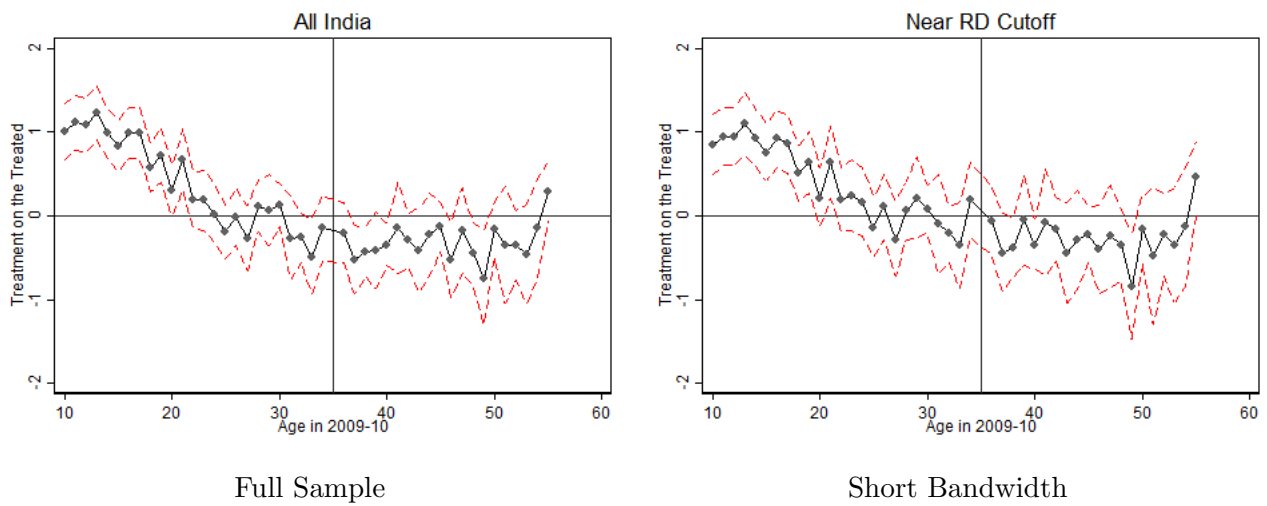
Years of Education - Old



Finished Upper Primary - Young

National Sample Survey 2009 for all persons. Figures made using [Calonico et al. \(2014\)](#) 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.6: Difference-in-Differences: Years of Education

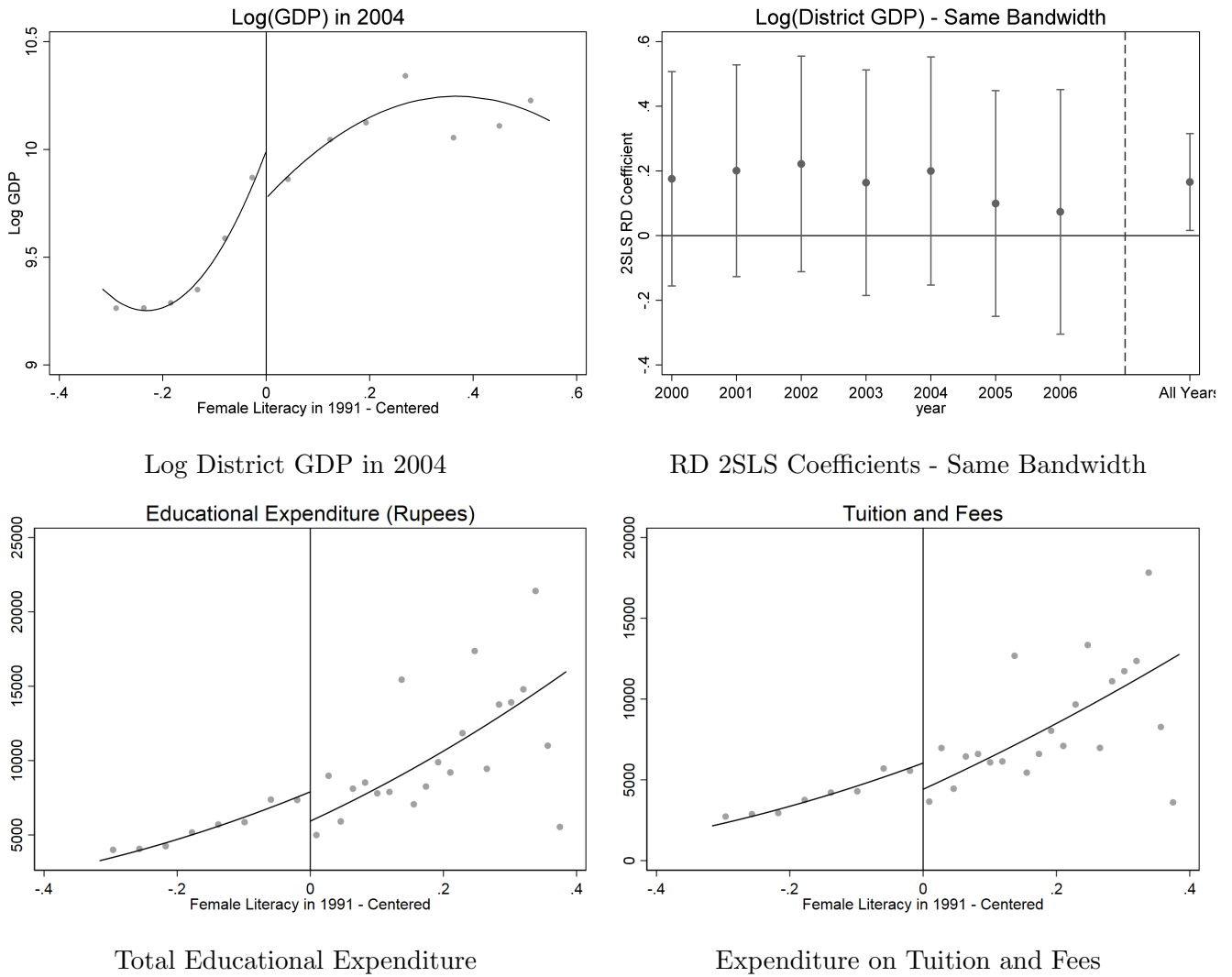


Full Sample

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



RD graph optimal binning and 2SLS RD coefficients calculated using [Calonico et al. \(2014\)](#) 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.

Table A.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 A.2: Summary Statistics: Household Level

	Non DPEP Mean	Non DPEP SD	DPEP Mean	DPEP SD	All Mean	All SD
Finished Primary School	0.70	0.46	0.59	0.49	0.66	0.47
Finished Upper Primary	0.57	0.49	0.47	0.5	0.53	0.5
Years of Education	7.23	5.21	5.97	5.33	6.79	5.28
Male	0.52	0.5	0.51	0.5	0.51	0.5
Age	37.77	14.43	37.37	14.38	37.63	14.42
Monthly Wage Earnings	40.59	50.63	30.5	38.07	37.25	47.09

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

Table A.3: Test Scores

Panel A: Reading 2008	Read Letter	Read Word	Reading Level 1	Reading Aggregate
RD Estimate	0.00411 (0.0107)	-0.0158 (0.0118)	-0.0147 (0.0120)	-0.0256 (0.0180)
Bandwidth	CCT	CCT	CCT	CCT
Panel B: Math 2008	Numbers 1-9	Numbers 10-99	Subtraction	Math Aggregate
RD Estimate	0.0531 (0.0116)***	0.0197 (0.0136)	0.0216 (0.0145)	0.0383 (0.0246)
Bandwidth	CCT	CCT	CCT	CCT
Panel C: Reading 2012	Read Letter	Read Word	Reading Level 1	Reading Aggregate
RD Estimate	-0.0143 (0.0148)	0.0164 (0.0141)	0.0196 (0.0137)	0.00221 (0.0313)
Bandwidth	CCT	CCT	CCT	CCT
Panel D: Math 2012	Numbers 1-9	Numbers 10-99	Subtraction	Math Aggregate
RD Estimate	0.0514 (0.0156)***	-0.0277 (0.0184)	0.0351 (0.0183)*	0.0633 (0.0219)***
Bandwidth	CCT	CCT	CCT	CCT

Source: Annual Status of Education Report (ASER) Data – years 2008 and 2012 – for children (aged 3 through 15).

Sample of 540 districts. Number of unique households range from about 331,490 in 2008 to about 337,315 households in 2012.

Bandwidths: ‘CCT’ is the [Calonico et al. \(2014\)](#) 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.

Reading Aggregate and Math Aggregate are the sum of the standardized values of each component in the reading and math categories.

Table A.4: Education Changes - Full Sample

Years of Education	Young	Old	Young	Old
RD Estimate	0.217 (0.0891)**	0.122 (0.120)	0.244 (0.0767)***	0.167 (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 (0.00893)***	0.0107 (0.0112)	0.0313 (0.00762)***	0.0166 (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. \(2014\)](#) method. ‘I and K’ is the [Imbens and Kalyanaraman \(2012\)](#) method.

Table A.5: Fraction of People that Have Finished At Least a Given Level of Education

Literate	Young	Old	Young	Old
RD Estimate	0.0328 (0.0143)**	-0.0121 (0.0169)	0.0291 (0.0124)**	0.00711 (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 (0.0174)***	-0.0117 (0.0180)	0.0574 (0.0150)***	0.00401 (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 (0.0197)***	-0.0169 (0.0184)	0.0733 (0.0169)***	0.000212 (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. \(2014\)](#) 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: Treatment on the Treated using Two-Stage Least Squares Fuzzy RD

Panel A: Full Sample				
Years of Education	Young	Old	Young	Old
RD Estimate	0.573 (0.190)***	0.278 (0.237)	0.571 (0.185)***	0.304 (0.219)
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 (0.458)***	-0.177 (0.451)	1.569 (0.390)***	0.211 (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				
Years of Education	Young	Old	Young	Old
RD Estimate	0.171 (0.0454)***	-0.0350 (0.0381)	0.165 (0.0380)***	0.000448 (0.0333)
Observations	9,045	7,729	10,175	9,920
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log Earnings				
Years of Education	Young	Old	Young	Old
RD Estimate	0.258 (0.0720)***	-0.0235 (0.0769)	0.326 (0.0605)***	0.0910 (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. \(2014\)](#) method. ‘I and K’ is the [Imbens and Kalyanaraman \(2012\)](#) method.

Table A.7: 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 (0.262)***	0.519 (0.282)*	0.963 (0.206)***	0.548 (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 (0.363)	0.396 (0.429)	-0.221 (0.316)	0.301 (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 (0.0420)***	0.0607 (0.0441)	0.195 (0.0325)***	0.123 (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 (0.0589)	0.0163 (0.0750)	-0.0287 (0.0509)	0.0880 (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. \(2014\)](#) method. ‘I and K’ is the [Imbens and Kalyanaraman \(2012\)](#) method.

Table A.8: Migration, Earnings Reported, Paid Monthly, and Unemployment

Panel A: Work Structure				
P(Wages Reported)	Young	Old	Young	Old
RD Estimate	-0.00366 (0.00488)	-0.0105 (0.00865)	-0.00366 (0.00513)	-0.0105 (0.00844)
Observations	37,201	42,316	32,742	39,823
Bandwidth selection procedure	CCT	CCT	I and K	I and K
P(Unemployed)				
P(Unemployed)	Young	Old	Young	Old
RD Estimate	-0.0125 (0.00225)***	-0.00421 (0.00186)**	-0.0157 (0.00272)***	-0.00394 (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)				
P(Paid monthly)	Young	Old	Young	Old
RD Estimate	0.0828 (0.0211)***	0.00998 (0.0264)	0.0874 (0.0198)***	0.0170 (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 (0.0106)	-0.00293 (0.0101)	-0.000416 (0.00124)	0.000628 (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. \(2014\)](#) method. ‘I and K’ is the [Imbens and Kalyanaraman \(2012\)](#) method.

Table A.9: Education and Earnings - Men

Panel A: Full Sample				
Years of Education	Young	Old	Young	Old
RD Estimate	0.391 (0.142)***	0.188 (0.130)	0.285 (0.0978)***	0.188 (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 (0.211)***	0.0582 (0.283)	0.681 (0.196)***	0.207 (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				
Years of Education	Young	Old	Young	Old
RD Estimate	0.0697 (0.0217)***	-0.00119 (0.0246)	0.0720 (0.0200)***	0.0151 (0.0179)
Observations	6,947	6,589	9,841	13,236
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log Earnings				
Years of Education	Young	Old	Young	Old
RD Estimate	0.146 (0.0324)***	-0.0332 (0.0460)	0.154 (0.0299)***	0.0632 (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. \(2014\)](#) 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: Education and Earnings - Women

Panel A: Full Sample				
Years of Education	Young	Old	Young	Old
RD Estimate	0.0933 (0.152)	-0.0271 (0.143)	0.0675 (0.157)	-0.0227 (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 (0.479)*	-0.0653 (0.443)	0.782 (0.418)*	-0.0780 (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				
Years of Education	Young	Old	Young	Old
RD Estimate	0.0762 (0.0422)*	-0.0295 (0.0349)	0.0882 (0.0364)**	-0.0297 (0.0360)
Observations	2,620	2,157	2,250	1,998
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log Earnings				
Years of Education	Young	Old	Young	Old
RD Estimate	-0.0455 (0.0745)	-0.0595 (0.0769)	0.0351 (0.0643)	-0.0686 (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. \(2014\)](#) 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.11: District-Age Cells and the Parametric RD

Panel A: District - Age Cells				
Years of Education	Young	Old	Young	Old
RD Estimate	0.603 (0.262)**	0.0458 (0.321)	0.644 (0.232)***	0.524 (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				
Years of Education	Young	Old	Young	Old
RD Estimate	0.0765 (0.0224)***	0.000168 (0.0265)	0.0755 (0.0229)***	0.0480 (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*** (0.224)	-0.146 (0.279)	0.884*** (0.225)	-0.124 (0.283)
Observations	10,038	11,088	10,038	11,088
Polynomial	Linear	Linear	Quad	Quad
Finished Upper-primary				
Years of Education	Young	Old	Young	Old
RD Estimate	0.0882*** (0.0219)	-0.0249 (0.0215)	0.0875*** (0.0220)	-0.0219 (0.0217)
Observations	10,038	11,088	10,038	11,088
Polynomial	Linear	Linear	Quad	Quad

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. \(2014\)](#) 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.12: 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 (0.336)**	-0.0881 (0.406)	0.698 (0.301)**	0.0902 (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 (0.0328)**	-0.0175 (0.0301)	0.0733 (0.0284)***	-0.00107 (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	0.671 (0.186)***	-0.188 (0.307)	0.539 (0.200)***	-0.343 (0.373)
Observations	9757	8143	7329	6287
Bandwidth selection procedure	MSE-2	MSE-2	CER-2	CER-2
Finished Upper-primary	Young	Old	Young	Old
RD Estimate	0.0801 (0.0163)***	-0.0259 (0.0237)	0.0818 (0.0184)***	-0.0391 (0.0283)
Observations	10367	7857	7300	6095
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. \(2014\)](#) 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. \(2014\)](#) 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.13: Robustness: Widening Age Restrictions

Panel A: Full Sample				
Years of Education	Young	Old	Young	Old
RD Estimate	0.257 (0.0680)***	0.104 (0.116)	0.259 (0.0734)***	0.140 (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 (0.194)***	-0.180 (0.269)	0.732 (0.180)***	-0.0557 (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				
Years of Education	Young	Old	Young	Old
RD Estimate	0.0728 (0.0204)***	-0.0262 (0.0231)	0.0727 (0.0201)***	-0.0162 (0.0174)
Observations	9,662	7,734	10,117	13,441
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Log Earnings				
Years of Education	Young	Old	Young	Old
RD Estimate	0.116 (0.0308)***	-0.0116 (0.0372)	0.130 (0.0283)***	-0.00263 (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. \(2014\)](#) 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.14: Household Expenditures

Log(Consumption Expenditure)				
	2004-5		2009-10	
RD Estimate	0.0664 (0.0125) ^{***}	0.0659 (0.0120) ^{***}	0.0589 (0.0177) ^{***}	0.0575 (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 (0.0581)	-0.216 (0.0492) ^{***}	-0.0296 (0.0538)	-0.0333 (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 (0.0806) ^{**}	-0.389 (0.0679) ^{***}	-0.215 (0.0927) ^{**}	-0.229 (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 (0.0550)	-0.0250 (0.0416)	0.175 (0.0379) ^{***}	0.197 (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. \(2014\)](#) 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.15: District GDP 2000 and 2004

Log(District GDP)	Y 2000	Y 2000	Y 2004	Y 2004
RD Estimate	0.0684 (0.141)	0.0890 (0.139)	0.110 (0.114)	0.144 (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 (2,951)	2,415 (3,222)	2,834 (2,369)	3,449 (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. \(2014\)](#) method. ‘I and K’ is the [Imbens and Kalyanaraman \(2012\)](#) method.

B Model Derivations and Extensions

B.I Education Sector

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.⁵⁰ 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 , such as distance to the nearest school (Carneiro et al., 2019).⁵¹

$$r_{id} \equiv -\Psi A_d + p_d + \eta_i \quad (9)$$

B.I.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 .⁵²

$$\max_{\mathbf{x}_m} A_d(\mathbf{x}_m) \quad \text{s.t.} \quad \sum_{m=1}^M \mathbf{p}_m \mathbf{x}_m \leq \mathbf{R}_d, \quad (\text{A.1})$$

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.

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

$$A(\mathbf{x}_m) = \prod_m \mathbf{x}_m^{\alpha_m}, \quad (\text{A.2})$$

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:

$$A_d(R_d, \mathbf{p}_m) = \mathbf{R}_d \prod_m \left(\frac{\alpha_m}{\mathbf{p}_m} \right)^{\alpha_m} \quad (\text{A.3})$$

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

B.I.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. In order to flexibly allow for the possibility of a crowd-out or crowd-in on of private schools, I outline a framework based on current known evidence.

⁵⁰Students choose the lowest cost school regardless of whether they are privately or publicly owned.

⁵¹Restricting 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.

⁵²The set-up is agnostic about heterogeneity in public schools – some may be more efficient than others.

The assumptions made here can be relaxed in many ways, but allow us to test whether private schools are crowded out.

Private schools are assumed to be 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. Private schools may, however, be more cost-effective. Private schools, in my model, therefore, have the same output as public schools, but may do so at a different cost; and there is heterogeneity in these costs ([Kremer and Muralidharan, 2007](#)).⁵³

Total educational output (in student-years) Q_{jd} by school j is a function of X_{jd} its aggregate inputs: $Q_{jd} = \bar{\theta}_d X_j$, and the average skill level of the district $\bar{\theta}_d$. This captures demand externalities ([Birdsall, 1985](#)), 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:

$$\max_{X_j} p_d \bar{\theta}_d X_j - Z(X_j) \quad (\text{A.4})$$

The costs $Z(X_j) = z_{1j} X_j + \frac{1}{2} z_{2d} X_j^2$ have a simple quadratic formulation.⁵⁴ 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:

$$Q_{jd} = \bar{\theta}_d X_j^* = \bar{\theta}_d \frac{p_d \bar{\theta}_d - z_{1j}}{z_{2d}} \quad \text{and} \quad \pi_{jd} = \frac{(p_d \bar{\theta}_d - z_{1j})^2}{2z_{2d}} \quad (\text{A.5})$$

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: $F(\bar{\theta}_d p_d)$. Notice what guides the entry and exit decision of schools is the average productivity level in the district $\bar{\theta}_d$, the price p_d , and consequently the cost z_{2d} which depends on the infrastructure levels.

The overall supply of private schooling is therefore:

$$S_{pvt,d}^{sy} = \int_0^{p_d \bar{\theta}_d} \bar{\theta}_d \frac{p_d \bar{\theta}_d - z_{1j}}{z_{2d}} f(\tilde{z}_{1j}) dz_{1j} = \frac{\bar{\theta}_d}{z_{2d}} [p_d \bar{\theta}_d - \mathbb{E}_d(z_{1j} | z_{1j} < p_d \bar{\theta}_d)], \quad (\text{A.6})$$

where $f(\tilde{z}_{1j})$ 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:

$$\Pi = \int_0^{\bar{\theta}_d p_d} \frac{(p_d \bar{\theta}_d - z_{1j})^2}{z_{2d}} dF(z_{1j}) \quad (\text{A.7})$$

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

⁵³Alternatively, they could have been modeled as having heterogeneous productivities, with the same result.

⁵⁴While 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 \(2020\)](#) and [Pal \(2010\)](#) find that physical infrastructure upgrades induce private-school entry in India.

B.I.3 Education Market Equilibrium and Changes in Policy

As shown in the text, the demand for schooling is determined by the household decisions, where $s_{id}^* = \frac{\mathbb{E}[\tilde{\beta}_d] - r_d - \eta_i}{\Gamma}$. Given a distribution for $\eta_i \sim H(\eta)$, the overall demand for schooling in district d comes from households:⁵⁵

$$S_d^{Dd} = \int \frac{\mathbb{E}[\tilde{\beta}_d] + \Psi A_d - p_d - \eta_i}{\Gamma} dH(\eta) = \frac{\mathbb{E}[\tilde{\beta}_d] + \Psi A_d - p_d - \bar{\eta}_d}{\Gamma}, \quad (\text{A.8})$$

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

$$S_d^{Sy} = \frac{\bar{\theta}_d}{z_{2d}} [p_d \bar{\theta}_d - \mathbb{E}_d(z_{1j} | z_{1j} < p_d \bar{\theta}_d)] + A_d \quad (\text{A.9})$$

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 (A.8) and (A.9) 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

$$p_d^* = \frac{\mathbb{E}[\tilde{\beta}_d] + (\Psi - \Gamma)A_d - \bar{\eta}_d}{\Gamma \left(\frac{\bar{\theta}_d^2}{z_{2d}} \right) + 1} \quad \text{and} \quad S_d^* = \frac{\bar{\theta}_d^2 \left(\mathbb{E}[\tilde{\beta}_d] + \Psi A_d \right) + z_{2d} A_d}{\Gamma \bar{\theta}_d^2 + z_{2d}} - \frac{\bar{\eta}_d}{\Gamma} \quad (\text{A.10})$$

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.

$$r_{id} = -R_d \Psi \prod_m \frac{\alpha_m^{\alpha_m}}{p_m} + p_d^*(R_d) + \eta_i \quad (\text{A.11})$$

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

$$S_d^* = \phi_1 \mathbb{E}[\tilde{\beta}_d] + \phi_2 R_d - \frac{\eta_d}{\Gamma}, \quad (\text{A.12})$$

where $\phi_1 \equiv \left(\frac{\bar{\theta}_d^2}{\Gamma \bar{\theta}_d^2 + z_{2d}} \right)$ and $\phi_2 \equiv \left(\frac{(z_{2d} + \Psi \bar{\theta}_d^2) (\prod_m \frac{\alpha_m^{\alpha_m}}{p_m})}{\Gamma \bar{\theta}_d^2 + z_{2d}} \right)$. In Equation (A.12) the equilibrium amount of schooling is affected by the expansion of public schooling.

⁵⁵For exposition, in this appendix, I suppress the cohort and skill subscripts from the returns β_d .

⁵⁶Alternatively, the public-school "supply" can be separated from access A_d . For example, the supply of public schools could be $x_{school}^* = R_d \frac{\alpha_{school}}{p_{school}}$. Doing this, would not change the model's predictions.

⁵⁷If the supply of public schools was instead modeled as x_{school}^* , then the equilibrium quantity would be $S_d^* = \frac{\bar{\theta}_d^2 (\mathbb{E}[\tilde{\beta}_d] + \Psi A_d) + z_{2d} \left(R_d \frac{\alpha_{school}}{p_{school}} \right)}{\Gamma \bar{\theta}_d^2 + z_{2d}} - \frac{\bar{\eta}_d}{\Gamma}$. This would produce the same qualitative results going forward.

B.II 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:

$$\log w_{asd} = \log \left(\frac{\partial Y_d}{\partial \ell_{asd}} \right) = \log \theta_{sd} + \log \psi_a + \left(\left(\frac{1}{\sigma_E} - 1 \right) \left(\frac{1}{\varrho} \right) \log Y_d - \left(\frac{1}{\sigma_E} - 1 \right) \left(\frac{1-\varrho}{\varrho} \right) \log \bar{K}_d \right) + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E} \right) \log L_{sd} - \frac{1}{\sigma_A} \log \ell_{asd}, \quad (\text{A.13})$$

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.III 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 (A.16), which includes an elasticity of substitution between labor ℓ_{sd} and skill biased capital k_{sd} represented by σ_s :

$$Y_d = L_d^\varrho K_d^{(1-\varrho)} \quad (\text{A.14})$$

$$L_d = \left(\sum_s \theta_{sd} L_{sd}^{\frac{\sigma_E-1}{\sigma_E}} \right)^{\frac{\sigma_E}{\sigma_E-1}} \quad (\text{A.15})$$

$$L_{sd} = \left(\Lambda_s k_{sd}^{\frac{\sigma_s-1}{\sigma_s}} + (1 - \Lambda_s) \ell_{sd}^{\frac{\sigma_s-1}{\sigma_s}} \right)^{\frac{\sigma_s}{\sigma_s-1}} \quad (\text{A.16})$$

$$\ell_{sd} = \left(\sum_a \psi_a \ell_{asd}^{\frac{\sigma_A-1}{\sigma_A}} \right)^{\frac{\sigma_A}{\sigma_A-1}} \quad (\text{A.17})$$

Given this set up, earnings can be represented by Equation (A.18), instead of Equation (3):

$$\log w_{asd} = \log \tilde{\varrho} + \log \psi_a + \log \theta_{sd} (1 - \Lambda_s) + \frac{1}{\sigma_E} \log Y_d + \left(\frac{1}{\sigma_s} - \frac{1}{\sigma_E} \right) \log L_{sd} + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_s} \right) \log \ell_{sd} - \frac{1}{\sigma_A} \log \ell_{asd} \quad (\text{A.18})$$

This new set up does not change the estimation or the interpretation of the estimates. In the following equation, that replaces Equation (15) to estimate the GE effects on all workers, the skill-biased capital term is captured by $L_{sd}(k_{sd}, \ell_{sd})$, and depends on the elasticity of substitution between skill-biased capital and labor σ_s :⁵⁸

$$\log \frac{w_{so,D=1}}{w_{so,D=0}} - \log \frac{w_{uo,D=1}}{w_{uo,D=0}} = \left(\log \frac{\theta_{sd}(1 - \Lambda_s)}{\theta_{ud}(1 - \Lambda_u)} \right) + \left(\frac{1}{\sigma_s} - \frac{1}{\sigma_E} \right) \left[\log \frac{L_{s,D=1}}{L_{u,D=1}} - \log \frac{L_{s,D=0}}{L_{u,D=0}} \right] + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_s} \right) \left[\log \frac{\ell_{s,D=1}}{\ell_{u,D=1}} - \log \frac{\ell_{s,D=0}}{\ell_{u,D=0}} \right] \quad (\text{A.19})$$

B.IV Definition of an Equilibrium

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 B.I.1).

⁵⁸The additional impact on the young, Equation (16), stays the same as before.

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 B.I.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 4.1). 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 exogenous elements are the utility, cost and production functions, the amount of government spending on schooling, and the expectations on future wage profiles. What is endogenous is the amount of schooling and returns to schooling, the optimal inputs in schools and schooling supply, firm output, and the equilibrium price and quantity of schooling. Appendix B.I 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: cost functions $\kappa(s, r, \Gamma)$, a student indirect-utility function increasing monotonically with the earnings function $\log w(s)$; 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 skill β_{asd} that varies by district, age and skill level; expectations over these returns $\mathbb{E}[\beta_{asd}]$; the distribution of the optimal years of schooling S_d^* , and price of going to school p_d^* ; optimal inputs into the access function $x_m^*(R_d, \alpha_m, \mathbf{P}_m)$; optimal private school inputs $X_j^*(p_d, z_{1j})$; equilibrium earnings w_{asd} and quantities of each type of worker ℓ_{asd} .*

B.V Deriving Equations (17) and (18)

In Equations (17) and (18) 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:

$$\begin{aligned}
\log \frac{w_{y,D=1}}{w_{y,D=0}} &= (\ell_{sy,D=1} \log w_{sy,D=1} + \ell_{uy,D=1} \log w_{uy,D=1}) - (\ell_{sy,D=0} \log w_{sy,D=0} + \ell_{uy,D=0} \log w_{uy,D=0}) \\
&= \ell_{sy,D=1} (\log w_{sy,D=1} - \log w_{sy,D=0}) + (\ell_{sy,D=1} - \ell_{sy,D=0}) \log w_{sy,D=0} + \\
&\ell_{uy,D=1} (\log w_{uy,D=1} - \log w_{uy,D=0}) + (\ell_{uy,D=1} - \ell_{uy,D=0}) \log w_{uy,D=0} \\
&= \ell_{sy,D=1} \log \frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=1} \log \frac{w_{uy,D=1}}{w_{uy,D=0}} + \\
&(\ell_{uy,D=1} - \ell_{uy,D=0}) \log w_{uy,D=0} + (\ell_{sy,D=1} - \ell_{sy,D=0}) \log w_{sy,D=0} \\
&= \ell_{sy,D=1} \log \frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=1} \log \frac{w_{uy,D=1}}{w_{uy,D=0}} + \underbrace{\Delta \ell_{sy} \log \frac{w_{sy,D=0}}{w_{uy,D=0}}}_{\beta_{as,D=0}} \tag{A.20}
\end{aligned}$$

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

$$\begin{aligned}
\log \frac{w_{y,D=1}}{w_{y,D=0}} &= (\ell_{sy,D=1} \log w_{sy,D=1} + \ell_{uy,D=1} \log w_{uy,D=1}) - (\ell_{sy,D=0} \log w_{sy,D=0} + \ell_{uy,D=0} \log w_{uy,D=0}) \\
&= \ell_{sy,D=0} (\log w_{sy,D=1} - \log w_{sy,D=0}) + (\ell_{sy,D=1} - \ell_{sy,D=0}) \log w_{sy,D=1} + \\
&\quad \ell_{uy,D=0} (\log w_{uy,D=1} - \log w_{uy,D=0}) + (\ell_{uy,D=1} - \ell_{uy,D=0}) \log w_{uy,D=1} \\
&= \ell_{sy,D=0} \log \frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=0} \log \frac{w_{uy,D=1}}{w_{uy,D=0}} + \\
&\quad (\ell_{uy,D=1} - \ell_{uy,D=0}) \log w_{uy,D=1} + (\ell_{sy,D=1} - \ell_{sy,D=0}) \log w_{sy,D=1} \\
&= \ell_{sy,D=0} \log \frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=0} \log \frac{w_{uy,D=1}}{w_{uy,D=0}} + \underbrace{\Delta \ell_{sy} \log \frac{w_{sy,D=1}}{w_{uy,D=1}}}_{\beta_{as,D=1}} \tag{A.21}
\end{aligned}$$

B.VI Deriving Migration-modified Labor Supply

In this appendix, I derive a modified version of the labor supply curve that explicitly incorporates the migration decision, building on other work (Bryan and Morten, 2018; Khanna et al., 2019a,b). Empirically, as I show, the probability of migration across skill-groups appear to be similar and low, suggesting empirically why there is no migration response to the policy.

I modify wages earned to depend not just on the origin district d , but also the destination d' . Workers pay an iceberg migration cost $\tau_{dd'}$ in migrating from d to d' , where $\tau_{dd} = 0$ for non-migrants. Finally, workers draw preferences for different destinations d' (that generates heterogeneity in the migration response). This draw $\varepsilon_{id'}$, affects wage-utility:

$$\widetilde{w_{aisdd'}} = w_{aisd'}(1 - \tau_{dd'})\varepsilon_{id'} \quad ,$$

where w_{aisd} is the same wage function as in the main text (Equation (3)). Assuming that $\varepsilon_{id'}$ is drawn from a Frechet distribution with shape parameter ν , we can derive the share of workers with skill s that migrate from d to d' . We denote this migration probability to be $\pi_{asdd'}$, where:

$$\pi_{asdd'} = \frac{(w_{asd'}(1 - \tau_{dd'}))^\nu}{\sum_k (w_{ask'}(1 - \tau_{dk}))^\nu} \tag{A.22}$$

The proof is as follows. Workers will pick the destination with the highest value of $w_{aisd'}(1 - \tau_{dd'})$. The probability that they pick destination 1 is given by:

$$\begin{aligned}
\pi_{asd1} &= Pr [\widetilde{w_{ais1}\varepsilon_1} > \widetilde{w_{aisd'}\varepsilon_{d'}}] = Pr \left[\varepsilon_{d'} < \frac{\widetilde{w_{ais1}\varepsilon_1}}{\widetilde{w_{aisd'}}} \right] \quad \forall d' \neq 1 \\
&= \int \frac{dF}{d\varepsilon_1} (\varepsilon_1, \alpha_1\varepsilon_1, \dots, \alpha_D\varepsilon_D) d\varepsilon_1 \tag{A.23}
\end{aligned}$$

where we define $\alpha_{d'} \equiv \frac{w_{ais1}(1 - \tau_{d1})}{w_{aisd'}(1 - \tau_{dd'})}$. We assume that the abilities are distributed with the following Frechet distribution:

$$F(\varepsilon_1, \dots, \varepsilon_D) = exp \left\{ - \left[\sum_{d=1}^D \varepsilon_d^{-\nu} \right] \right\} \tag{A.24}$$

So the derivative of the CDF is given by:

$$\frac{dF}{d\varepsilon} = \nu\varepsilon^{-\nu-1} \exp \left\{ - \left[\sum_{d=1}^D \varepsilon_d^{-\nu} \right] \right\} \quad (\text{A.25})$$

This derivative evaluated at $(\varepsilon_1, \alpha_1\varepsilon_1, \dots, \alpha_D\varepsilon_D)$, allows us to determine the probability of choosing destination 1:

$$\begin{aligned} \pi_{asd1} &= \int \nu\varepsilon^{-\nu-1} \exp \left\{ - \left[\sum_{d=1}^D (\alpha_d\varepsilon)^{-\nu} \right] \right\} d\varepsilon \\ &= \frac{1}{\sum_{d=1}^D \alpha_d^{-\nu}} \int \left(\sum_{d=1}^D \alpha_d^{-\nu} \right) \varepsilon^{-\nu-1} \exp \left\{ - \left[\varepsilon^{-\nu-1} \left(\sum_{d=1}^D \alpha_d^{-\nu} \right) \right] \right\} d\varepsilon \\ &= \frac{1}{\sum_{d=1}^D \alpha_d^{-\nu}} \int dF(\varepsilon) = \frac{1}{\sum_{d=1}^D \alpha_d^{-\nu}} \cdot 1 \\ &= \frac{(w_{asd'}(1 - \tau_{dd'}))^\nu}{\sum_k (w_{ask'}(1 - \tau_{dk}))^\nu} \end{aligned} \quad (\text{A.26})$$

The third line comes from the properties of the Frechet distribution, where we know that the term in the integral of the second line is simply the PDF with a shape parameter ν , and a scale parameter $\sum_{d=1}^D \alpha_d^{-\nu}$.

From the Frechet assumptions, we can also derive indirect wage-utility as workers now derive utility from being in different destinations:

$$\begin{aligned} E[V_{asd'}|d'] &= w_{asd'}(1 - \tau_{dd'})E[\varepsilon_{id'}|d'] \\ &= \left(\sum_{d'} (w_{asd'}(1 - \tau_{dd'}))^\nu \right)^{\frac{1}{\nu}} \Gamma \left(1 - \frac{1}{\nu} \right), \end{aligned} \quad (\text{A.27})$$

where Γ is the Gamma-function, and constant across destinations. Intuitively, the first part of the equation (let us define it as $\overline{V_{asd'}}$) is simply the geometric mean of the different destination options – or the option value of migration.

We know that the supply of workers ℓ_{asd} of skill level s from age-cohort a working in district d , depends on the probability that workers stay in their home districts π_{asdd} by skill and age cohort. From the derivation above, and seeing that the migration costs $\tau_{dd} = 0$ for staying home is zero, the modified labor-supply curve is simply:

$$\log \ell_{asd} = \nu \log w_{asd} - \nu \log \overline{V_{asd}},$$

Here, we see that including migration changes the slope of the labor *supply* curve in the model. The labor demand portion of the model remains the same as before. Furthermore, acquiring more skill changes the flows of migration in the following manner:

$$\Delta \text{Emigration}_{ad} = \Delta \ell_{asd} \sum_k (\pi_{asdk} - \pi_{audk}) \quad (\text{A.28})$$

where $\Delta \ell_{asd}$ is simply the fraction of people induced into getting more skill due to the policy – this is the same variable and notation defined in Section 4.1.1 of the paper and used in the Equations (17) - (18). Equation (A.28) here makes an intuitive point: the change in the emigration rate is the product of two terms (1) the fraction of people induced into going from unskilled to skilled $\Delta \ell_{asd}$ and (2) the difference in the migration probabilities for the skilled and unskilled $(\pi_{asdd'} - \pi_{audd'})$.

In the 2008 migration round of the NSS data, I find that the probability of migration for economic reasons, for the skilled $\pi_{asdd'} = 0.04$ and unskilled $\pi_{audd'} = 0.041$ are roughly equal, and low. In the context of my framework, it would imply that migration costs $\tau_{dd'}$ are

high. Given the lack of meaningful differences between migration probabilities by skill, it is unsurprising we empirically fail to find changing migration rates as a result of the policy.

B.VII Labor Market Distortions: Misallocation Across Sectors

One type of labor market distortion arises from the fact that there may be multiple sectors (or firms) with misallocation of factors between them, such that each sector (or firm) may have a different return to education. For instance, differential access to skill-biased capital may lead to differences in skilled wages across sectors. To address this point, I begin with modifying the production function (Equation (1)), to include a sector j subscript to all quantities and parameters:

$$Y_d = \left(\sum_j y_{dj}^{\frac{\xi-1}{\xi}} \right)^{\frac{\xi}{\xi-1}} \quad \text{where} \quad y_{dj} = L_{dj}^{\rho_j} K_{dj}^{(1-\rho_j)} \quad \text{and} \quad L_{dj} = \left(\sum_s \theta_{sdj} L_{sdj}^{\frac{\sigma_{Ej}-1}{\sigma_{Ej}}} \right)^{\frac{\sigma_{Ej}}{\sigma_{Ej}-1}} \quad (1)b$$

Total output in the district is a CES aggregate of each sector's output, with the possibility that that as $\xi \rightarrow \infty$, it is simply the sum of what each sector produces. The aggregate skill supply is again a nested CES of each cohort a 's supply similar to Equation (2):

$$L_{sdj} = \left(\sum_a \psi_{aj} \ell_{asdj}^{\frac{\sigma_{Aj}-1}{\sigma_{Aj}}} \right)^{\frac{\sigma_{Aj}}{\sigma_{Aj}-1}} \quad (2)b$$

As a result, the new relative labor demand curve is a modified version of Equation (3):

$$\begin{aligned} \log w_{asdj} &= \log \left(\frac{\partial Y_d}{\partial \ell_{asdj}} \right) = \log \left(Y_d^{\frac{1}{\xi}} \frac{\partial y_{dj}}{\partial \ell_{asdj}} \right) \quad (3)b \\ &= \frac{1}{\xi} \log Y_d + \log \tilde{\varrho}_j + \log \theta_{sdj} + \log \psi_{aj} + \frac{1}{\sigma_{Ej}} \log y_{dj} + \left(\frac{1}{\sigma_{Aj}} - \frac{1}{\sigma_{Ej}} \right) \log L_{sdj} - \frac{1}{\sigma_{Aj}} \log \ell_{asdj} \end{aligned}$$

Wage differences across sectors $\log w_{asdj} \neq \log w_{asdj'}$ arise from the differences in access to factor inputs and productivities. As a result, we can define a sector j specific skill premium in lieu of Equation (5):

$$\log \frac{w_{asdj}}{w_{audj}} = \log \frac{\theta_{sdj}}{\theta_{udj}} + \left(\frac{1}{\sigma_{Aj}} - \frac{1}{\sigma_{Ej}} \right) \log \frac{L_{sdj}}{L_{udj}} - \frac{1}{\sigma_{Aj}} \log \frac{\ell_{asdj}}{\ell_{audj}} \equiv \beta_{asdj} \quad (5)b$$

The sector-specific returns highlight the differences in access to, say, skill-biased capital across sectors $\log \frac{\theta_{sdj}}{\theta_{udj}} \neq \log \frac{\theta_{sdj'}}{\theta_{udj}'}$. I describe the relationship between the sector-specific returns β_{asdj} and the aggregate district-level return β_{asd} below. While the subsequent discussion (starting from Section 2.2 on the education decision side) are unaffected by the sector differences, they are relevant for when we estimate the returns to education in Section 4.1.1.

Like Equations (14) and (15), we derive sector-specific changes to returns $\Delta \beta_{asdj}$, but to interpret these we need to decompose wages and re-derive β_{asd} as in Appendix B.V. That is, we re-derive Equations (17) and (18). Equation (17) for the young cohorts is altered:

$$\begin{aligned}
\log \frac{w_{y,D=1}}{w_{y,D=0}} &= \sum_j (\ell_{syj,D=1} \log w_{syj,D=1} + \ell_{uyj,D=1} \log w_{uyj,D=1}) \\
&\quad - \sum_j (\ell_{syj,D=0} \log w_{syj,D=0} + \ell_{uyj,D=0} \log w_{uyj,D=0}) \\
&= \sum_j \left(\ell_{syj,D=1} \log \frac{w_{syj,D=1}}{w_{syj,D=0}} + \ell_{uyj,D=1} \log \frac{w_{uyj,D=1}}{w_{uyj,D=0}} \right) \\
&\quad + \sum_j (\Delta \ell_{syj} \log w_{syj,D=0}) + \sum_j (\Delta \ell_{uyj} \log w_{uyj,D=0}) \\
&= \sum_j \left(\ell_{syj,D=1} \log \frac{w_{syj,D=1}}{w_{syj,D=0}} + \ell_{uyj,D=1} \log \frac{w_{uyj,D=1}}{w_{uyj,D=0}} + \underbrace{\Delta \ell_{syj} \log \frac{w_{syj,D=1}}{w_{uyj,D=0}}}_{\beta_{ysj,D=0}} \right) \tag{17b}
\end{aligned}$$

In going from Step 1 to 2 above, we simply add and subtract the terms $\ell_{syj,D=1} \log w_{syj,D=0}$ and $\ell_{uyj,D=1} \log w_{uyj,D=0}$. In going from Step 2 to 3, we use the fact that $\Delta \ell_{syj} = (\ell_{sy,D=1} - \ell_{sy,D=0}) = -(\ell_{uy,D=1} - \ell_{uy,D=0})$. Let us for notational purposes define the share of the skill s workforce in a district d and age cohort a that works in sector j as $\widetilde{\ell}_{asdj} = \frac{\ell_{asdj}}{\ell_{asd}}$.⁵⁹

$$\begin{aligned}
\log \frac{w_{y,D=1}}{w_{y,D=0}} &= \ell_{sy,D=1} \sum_j \left(\widetilde{\ell}_{syj,D=1} \log \frac{w_{syj,D=1}}{w_{syj,D=0}} \right) + \ell_{uy,D=1} \sum_j \left(\widetilde{\ell}_{uyj,D=1} \log \frac{w_{uyj,D=1}}{w_{uyj,D=0}} \right) + \\
&\quad \Delta \ell_{sy} \sum_j \left(\widetilde{\ell}_{syj,D=1} \log w_{syj,D=0} - \widetilde{\ell}_{uyj,D=1} \log w_{uyj,D=0} \right) \tag{17c}
\end{aligned}$$

As I explain below, this type of market distortion implies that the returns to skill are a weighted average of the skilled and unskilled wages by sector. Furthermore, in the absence of sector-switching (i.e., the misallocation is strict, such that cross-sectoral factor mobility does not equalize factor prices), we also get that $\log \frac{w_{sy,D=1}}{w_{sy,D=0}} = \sum_j \left(\widetilde{\ell}_{syj,D=1} \log \frac{w_{syj,D=1}}{w_{syj,D=0}} \right)$ and that $\log \frac{w_{uy,D=1}}{w_{uy,D=0}} = \sum_j \left(\widetilde{\ell}_{uyj,D=1} \log \frac{w_{uyj,D=1}}{w_{uyj,D=0}} \right)$. Together, these allow us to derive a familiar version of the original Equation (17):

⁵⁹To derive the last line of the above equation, we use the following fact on the share of marginal individuals induced into becoming skilled by sector j , and skilled wages:

$$\begin{aligned}
\sum_j (\ell_{syj,D=1} - \ell_{syj,D=0}) \log w_{syj,D=0} &= \sum_j (\ell_{sy,D=1} - \ell_{sy,D=0}) \left(\widetilde{\ell}_{syj,D=1} \log w_{syj,D=0} \right) \\
&= (\ell_{sy,D=1} - \ell_{sy,D=0}) \sum_j \left(\widetilde{\ell}_{syj,D=1} \log w_{syj,D=0} \right) \\
&= \Delta \ell_{sy} \sum_j \left(\widetilde{\ell}_{syj,D=1} \log w_{syj,D=0} \right)
\end{aligned}$$

Furthermore, since $(\ell_{sy,D=1} - \ell_{sy,D=0}) = -(\ell_{uy,D=1} - \ell_{uy,D=0})$, we also get the term $-\Delta \ell_{sy} \sum_j \left(\widetilde{\ell}_{uyj,D=1} \log w_{uyj,D=0} \right)$ for unskilled wages. So together:

$$\begin{aligned}
\sum_j (\ell_{syj,D=1} - \ell_{syj,D=0}) \log w_{syj,D=0} &+ (\ell_{uy,D=1} - \ell_{uy,D=0}) \log w_{uyj,D=0} \\
&= \Delta \ell_{sy} \sum_j \left(\widetilde{\ell}_{syj,D=1} \log w_{syj,D=0} - \widetilde{\ell}_{uyj,D=1} \log w_{uyj,D=0} \right)
\end{aligned}$$

$$\log \frac{w_{y,D=1}}{w_{y,D=0}} = \ell_{sy,D=1} \log \frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=1} \log \frac{w_{uy,D=1}}{w_{uy,D=0}} + \Delta \ell_{sy} \sum_j \left(\widetilde{\ell_{syj,D=1}} \log w_{syj,D=0} - \widetilde{\ell_{uyj,D=1}} \log w_{uyj,D=0} \right) \quad (17)d$$

That is, we recover the first few terms of original Equation (17). We can also understand under what conditions that last terms of Equation (17) and (17)d can be equated. As such, this framework also allows us to understand the relationship between sector-specific returns in partial equilibrium $\beta_{asj,D=0}$ and aggregate returns $\beta_{as,D=0}$ in the face of factor misallocation. We can compare the last terms of Equations (17)b and (17)c:

$$\sum_j \Delta \ell_{saj} \underbrace{\log \frac{w_{saj,D=0}}{w_{uaj,D=0}}}_{\beta_{asj,D=0}} = \Delta \ell_{sa} \underbrace{\sum_j \left(\widetilde{\ell_{saj,D=1}} \log w_{saj,D=0} - \widetilde{\ell_{uaj,D=1}} \log w_{uaj,D=0} \right)}_{\beta_{as,D=0}} \quad (A.29)$$

The right side shows that in the face of factor input misallocation across sectors, the estimated return is a weighted average of the skilled and unskilled wages by sector, weighted by the factor shares in that sector. If, also, the share of each skill group that works in sector j is the same across sectors (i.e., if $\widetilde{\ell_{saj,D=1}} = \widetilde{\ell_{uaj,D=1}} \equiv \widetilde{\ell_{aj,D=1}}$), then the estimated return is simply a weighted average each sectors skill-premium. That is (if we make this rather strong assumption), we get:

$$\Delta \ell_{sa} \underbrace{\sum_j \left(\widetilde{\ell_{saj,D=1}} \log w_{saj,D=0} - \widetilde{\ell_{uaj,D=1}} \log w_{uaj,D=0} \right)}_{\beta_{as,D=0}} = \Delta \ell_{sa} \underbrace{\sum_j \left(\widetilde{\ell_{aj,D=1}} \underbrace{\log \frac{w_{saj,D=0}}{w_{uaj,D=0}}}_{\beta_{asj,D=0}} \right)}_{\beta_{as,D=0}}$$

In other words, if we assume equal shares, the aggregate skill premium is a weighted average of the sector-specific premia: $\beta_{as,D=0} = \sum_j \ell_{aj,D=1} \beta_{asj,D=0}$. Yet, in many settings, the assumption that the share of skilled workers that work in sector j is equal to the share of unskilled workers that work in sector j' , is unlikely to hold. In which case, in the face of meaningful factor-input misallocation, we should interpret $\beta_{as,D=0}$ as defined in Equation (A.29). That is, as the difference in the weighted average of the skilled and the weighted average of the unskilled wages (rather than the weighted average of the sector-specific skill premia).

Empirically, a few patterns describe how sector composition affects the exercise. Appendix F.II (and specifically Table F.1) shows how returns vary by the sectoral composition of the district. While expectantly, places that have a higher manufacturing share, tend to have a higher return, than places with a lower share, these returns are not statistically indistinguishable from each other.⁶⁰

B.VIII Market Power and Wage Rigidities

Recent work on Indian labor markets describes wage rigidities (Breza et al., 2019; Kaur, 2019) and monopsonistic labor markets (Muralidharan et al., 2017). While frictionless labor markets are not necessary to derive the private returns (nor the GE effects on private returns), it does affect how we think about the overall productivity consequences. To elaborate, an individual who goes from being unskilled to skilled will still see their earnings increase be

⁶⁰In Table E.2 we also establish that treatment probability seems unrelated to sectoral (occupational) composition.

determined by $\log \frac{w_s}{w_u}$, but that is no longer the increase in productivity (and hence output). So the private return to the individual stays similar to before, and if changes in the supply of skilled workers affect the skill premium in general equilibrium, the GE effects on the private returns are again similar to before. Yet, to understand the effects on changes to overall productivity, we require information on the labor supply elasticity (for a detailed discussion of market power in India, see [Muralidharan et al. \(2017\)](#)).

The profit maximizing wage under a situation of labor market power also depends on the labor supply elasticity ν_s :

$$w_{asd} = MRP_L \Big/ \left(1 + \frac{1}{\nu_s}\right) = \left(\frac{\partial Y_d}{\partial \ell_{asd}}\right) \Big/ \left(1 + \frac{1}{\nu_s}\right)$$

As a result, the skill premium also depends on the relative labor supply elasticities:

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

From the workers point of view, the private returns in partial equilibrium, and the GE effects are driven solely by what the equilibrium wages are (regardless of market power). That is, Equations (17) and (18) continue to tell us the returns to education, and how they change in general equilibrium.

Yet, we may be interested in the consequences on economic productivity. That is, what happens to output when an individual moves from being unskilled to skilled depends on the first three terms of Equation (5)c (excluding the labor supply elasticities).

Estimates of the labor supply elasticity in the face of rigidities and market power range from 3.07 ([Muralidharan et al., 2017](#)) to 3.89 ([Breza et al., 2019](#)). Yet, what is important for us to know is not the level of the labor-supply elasticity, but rather the *relative* labor supply elasticities between the skilled and unskilled. If both the skilled and unskilled have a similar elasticity (i.e., if $\nu_s \approx \nu_u$), the skill premium already captures the corresponding gain to productivity. The authors of [Breza et al. \(2019\)](#) and [Muralidharan et al. \(2017\)](#) suggest their setup is not ideal for estimating the labor supply elasticity, and so do not show heterogeneity by skill group. Instead, we can find well-identified estimates of this elasticity in [Goldberg \(2016\)](#) who randomizes wage offers in Malawi and measures the labor supply response. [Goldberg \(2016\)](#) finds a low labor supply elasticity, and (more importantly) fails to reject the null that these elasticities are the same across sub-groups. All the while suggesting that $\log \frac{\left(1 + \frac{1}{\nu_s}\right)}{\left(1 + \frac{1}{\nu_u}\right)} \approx 0$, and precluding the need to make any adjustments to the analysis.⁶¹

Nevertheless, we can still try and do some sensitivity analyses. If we started with the ([Muralidharan et al., 2017](#)) estimate of 3.07 and assume that one of the skill groups has a smaller elasticity of say 2.5, then to derive the productivity component of the skill premium, we would adjust $\log \frac{w_{asd}}{w_{aud}}$ by 0.02 log points. That is, instead of a 13.4% productivity gain, we would have either a 11.4% (if the skilled had less elastic labor supply) or a 15.4% (if the unskilled were less elastic) gains. If instead, one of the skill groups had a larger elasticity of say 3.5, we would need to adjust the estimates by 0.0157 log points (productivity gains range from 11.82% to 14.9%). These adjustments are small relative to the mean return of 13.4%.

One final note on this discussion: for convenience, we presume (as is conventional in the literature) that the labor supply elasticities by skill are constant, and not a function of the labor supply (i.e., η_s , does not vary with L_{sd}).

⁶¹In the context of inter-city migration in China, ([Khanna et al., 2019a](#)) estimate a migration-supply elasticity of 1.3 for skilled workers, and 1.01 for unskilled workers. Yet, the migration response to differences in wages across districts is a different parameter than the within-district supply responsiveness to wages. Indeed, in China, migration restrictions for the unskilled are one possible reason for the lower elasticity.

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.⁶²

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.⁶³

The “*project would be a reconstruction of primary education as a whole in selected districts instead of piecemeal implementation of schemes*” (GOI, 1994). While most of the funds were directed towards the government schools, some were used towards a training drive for teachers of private and government-aided schools.

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. By 2002 the World Bank alone had committed about \$1.62 bn on DPEP, whereas the other donors concentrated on certain states. For example, in the first few years of the program, the EC spent ECU 150mn in Madhya Pradesh, the Netherlands spent \$25.8 mn in Gujarat, DFID spend 80 mn pounds in Andhra Pradesh and West Bengal, whereas UNICEF spent \$ 153 mn in Bihar. 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.

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).

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 district and appraised by the Bureau that also provided implementation support. The programs were evaluated and poorly performing subprojects are dropped.

⁶²Primary is usually grades 1 through 4 or 5, and upper primary is grades 5 or 6 through 8.

⁶³SSA was similar to the DPEP, but covered the entire country. There were, however, certain programs under SSA that targeted certain sub-districts.

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 Mother-Teacher Associations were tasked with creating local awareness, 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 education levels, and various infrastructural features. The DISE data was initially meant to cover only 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 A.1 for summary statistics in 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 A.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 is a yearly education survey for school-age (3-16 years) children in India. The sample is a representative repeated cross section at the district level.⁶⁴ Children are also tested in math and reading ability. It surveys children at home – whether they went to government school, private school, religious schools and even dropouts. It is administered each year on 2 to 3 weekends from the end of September to the end of November limiting considerations of spatially systematic seasonality in data collection, and endogenous sampling as in school children are likely not available on weekdays.

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: 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.

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

⁶⁴In each Indian district, 30 villages are sampled from the latest Census list of villages, using the PPS (Probability Proportional to Size) sampling technique. A team of two surveyors go to the village, meet the village head, and make a list of households in the village. They divide the village into 4 sections (‘hamlets’), and select 5 households from each hamlet, to get a total of 20 households per village. In each household, they record information about all children in the age 5-16 years. Children are interviewed at the household on weekends so as to include both school-going and unrolled students in the testing. This produces about 600 samples households per district, or about 300,000 households across India each year.

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 Empirical Strategy, Compliance, and Varying Returns

E.I What does the Difference-in-Differences Estimator Estimate?

A difference-in-differences (DID) estimator would usually compare (1) districts that receive the policy to district that do not, and (2) cohorts young enough to change their schooling decisions to older cohorts. Keeping with the notation used in the main paper, let y represent the young cohort and o the older cohort, and let $D = 1$ be for treated regions, while $D = 0$ be regions that did not get the policy. Differencing out the wages by cohort and region helps us estimate:

$$W_{DiD} = \log \frac{w_{y,D=1}}{w_{y,D=0}} - \log \frac{w_{o,D=1}}{w_{o,D=0}} \quad (\text{A.30})$$

These average wages can be split up by skilled s and unskilled u groups. Like in the paper, let ℓ_{sy} represent the fraction of young that are skilled. For ease of notation, let us also define how the skilled and unskilled wages by cohort will change due to the GE effects. Let $\Delta\omega_{sy} \equiv \log \frac{w_{sy,D=1}}{w_{sy,D=0}}$ be how the equilibrium skilled wage for youth changes due to the GE effects. Similarly, $\Delta\omega_{uy} \equiv \log \frac{w_{uy,D=1}}{w_{uy,D=0}}$ is how the unskilled wage for the youth changes due to the GE effects. And these quantities for older cohorts would be $\Delta\omega_{so}$ and $\Delta\omega_{uo}$.

We can rewrite Equation (A.30) in terms of these quantities, the fraction of people who switch (compliers) from being unskilled to skilled $\Delta\ell_{sy}$ and the partial equilibrium returns to skill for that group $\beta_{y,D=1} \equiv \log \frac{w_{sy,D=1}}{w_{uy,D=0}}$:

$$W_{DiD} = \underbrace{[\ell_{sy,D=0}\Delta\omega_{sy} + \ell_{uy,D=0}\Delta\omega_{uy} + \Delta\ell_{sy}\beta_{y,D=1}]}_{\log \frac{w_{y,D=1}}{w_{y,D=0}}} - \underbrace{[\ell_{so,D=1}\Delta\omega_{so} + \ell_{uo,D=1}\Delta\omega_{uo}]}_{\log \frac{w_{o,D=1}}{w_{o,D=0}}} \quad (\text{A.31})$$

Notice that in order to estimate the returns, the term we wish to isolate is simply $\Delta\ell_{sy}\beta_{y,D=0}$. We may rewrite the above expression to group the sources of confounders:

$$W_{DiD} = \Delta\ell_{sy}\beta_{y,D=0} + [\ell_{sy,D=1}\Delta\omega_{sy} + \ell_{uy,D=1}\Delta\omega_{uy}] - [\ell_{so,D=1}\Delta\omega_{so} + \ell_{uo,D=1}\Delta\omega_{uo}] \quad (\text{A.32})$$

There are a few things to notice from Equation (A.32). First, if there no GE effects (i.e. all the $\Delta\omega = 0$), then the DiD estimator recovers what we want: $\Delta\ell_{sy}\beta_{y,D=0}$. Yet, when there are GE effects, there are two main types of effects that make the DiD estimate different from $\Delta\ell_{sy}\beta_{y,D=0}$.

The Composition Effect: First, suppose we assume that the young and the old are perfect substitutes (i.e. $\sigma_A = 0$), and also assume that the rise in skilled wages at the RD cutoff are equal in magnitude to the fall in unskilled wages. That is, suppose we assume, $\Delta\omega_{sy} = \Delta\omega_{so} = -\Delta\omega_{uy} = -\Delta\omega_{uo} \equiv \Delta\omega$. Even with these simplifying assumptions, the DiD estimator does not recover the partial or general equilibrium returns. Under these assumptions, Equation (A.32) can be rewritten as:

$$W_{DiD} = \Delta\ell_{sy}\beta_{y,D=0} + \underbrace{[[\ell_{sy,D=1} - \ell_{uy,D=1}] - [\ell_{so,D=1} - \ell_{uo,D=1}]]}_{\text{Composition Effect}} \Delta\omega \quad (\text{A.33})$$

Similarly, we can derive an equation in terms of the returns to skill after GE effects affect $\beta_{y,D=1}$.

$$W_{DiD} = \Delta\ell_{sy}\beta_{y,D=1} + \underbrace{[[\ell_{sy,D=0} - \ell_{uy,D=0}] - [\ell_{so,D=0} - \ell_{uo,D=0}]]}_{\text{Composition Effect**}} \Delta\omega \quad (\text{A.34})$$

These ‘‘Composition Effects’’ capture the fact that the young and old differ in the baseline composition of the skilled and unskilled workers. Suppose the GE effects lower skilled wages and raise unskilled wages for both the young and the old in the exact same manner (i.e. $\Delta\omega < 0$). Even then, the DiD will not directly recover either $\beta_{y,D=1}$ nor $\beta_{y,D=0}$. For instance, if there are a lot of unskilled older workers, then the average wage for older workers will rise. If there are a lot of skilled older workers, the average wage for the old will fall. Since the skill-composition of the old and young are usually different, the old may not be a meaningful counterfactual for the young.

The Complementarity Effect: The second reason why the DiD may not easily recover β_y has to do with the fact that the young and the old are not perfect substitutes. Even if we (unreasonably) assume that the composition of the skilled and unskilled are identical across all cohorts, we cannot identify the returns. If $\ell_{sy,D=1} = \ell_{uy,D=1} = \ell_{so,D=1} = \ell_{uo,D=1} = \ell$, we can rewrite Equation (A.32) to be:

$$W_{DiD} = \Delta\ell_{sy}\beta_{y,D=0} + \underbrace{[[\Delta\omega_{sy} + \Delta\omega_{uy}] - [\Delta\omega_{so} + \Delta\omega_{uo}]]}_{\text{Complementarity Effect}} \ell \quad (\text{A.35})$$

This ‘‘Complementarity Effect’’ arises out of the fact that changes to the skilled and unskilled wages will differ by cohort. For instance, if the young and the old are complements in production, then an increase in the skilled youth will tend to lower the skilled wage for the young but raise the skilled wage for the old.

Recovering Returns from DiDs: So under what assumptions can we recover the returns using a DiD estimator? Equation (A.32) suggests that these assumptions are rather strong. We would need to adjust the estimator for all the components on the right hand side of Equation (A.32). While these components are observable, they are not well-identified. That is, there may be systematic differences between treated and untreated districts in the levels of wages that would bias estimates of these components. Since the DiD assumptions are only based on parallel trends, and not levels, we would need stronger assumptions. For instance, to estimate $\Delta\omega_{sy} \equiv \log \frac{w_{sy,D=1}}{w_{sy,D=0}}$ we would need to assume that the only reason that the wage *levels* for skilled youth differ across treated and untreated districts is because of the GE effects of the policy, when in fact it may be for a whole host of other reasons that simultaneously determine the size of the local economy and schooling market. Under the RD assumptions, however, we require weaker assumptions and can estimate $\Delta\omega_{sy}$ if required.

E.II Compliance at Cutoffs, and Effects Away from the Cutoff

I explore issues related to the external validity of our estimates across other districts. To elaborate, the Indian government said districts with a female literacy below the national average were eligible to receive the program. Despite a strong discontinuity at the cutoff, some districts with low literacy did not receive it. I use a fuzzy RD to estimate outcomes at the cutoff. If the effects on compliers and non-complier districts at the cutoff are different, then that affects the external validity of our results. Furthermore, the RD is informative about the effects only for districts around the cutoff. Yet, with some assumptions, we can speak to what we may expect the effects to be away from the cutoff.

E.II.1 Are Complier Districts Different from Others at the Cutoff?

I follow [Bertanha and Imbens \(2019\)](#) and [Brinch et al. \(2017\)](#), to first test for any difference in outcomes between treated compliers and always-takers, and between untreated compliers

and never-takers. These tests are informative about the external validity of the LATE to other compliance groups (districts) at the threshold.

Bertanha and Imbens (2019) suggest the following tests. First, for the set of regions that received the policy ($D_i = 1$), evaluate the outcome Y_i as the running variable X_i nears the cutoff c :

$$\lim_{\epsilon \rightarrow 0} \mathbb{E}[Y_i | D_i = 1, c < X_i < c + \epsilon] = \lim_{\epsilon \rightarrow 0} \mathbb{E}[Y_i | D_i = 1, c - \epsilon < X_i < c] \quad (\text{A.36})$$

This tells us whether the treated compliers have similar outcomes as the always-takers. We can do a similar test for those who did not receive the policy ($D_i = 0$), which tells us about the equality in outcomes between untreated compliers and never-takers:

$$\lim_{\epsilon \rightarrow 0} \mathbb{E}[Y_i | D_i = 0, c < X_i < c + \epsilon] = \lim_{\epsilon \rightarrow 0} \mathbb{E}[Y_i | D_i = 0, c - \epsilon < X_i < c] \quad (\text{A.37})$$

We may also jointly test for the equalities in Equations (A.36) and (A.37), in a manner detailed in Bertanha and Imbens (2019). To implement these tests we test for a discontinuity in outcomes (at the cutoff) after conditioning on treatment status. Implementing the joint test requires us to have sufficient observations across all groups: compliers, always-takers and never-takers. If, say, there are not enough always-takers at the cutoff, then we may only be able to test the relationship in Equation (A.37).

The test suggested by Brinch et al. (2017) is similar (p 998 and p 1016), but is not adapted to an RD framework. As they stress, tests for the LATE’s external validity do not require the estimation of the Marginal Treatment Effects (MTE) model. The suggested test is therefore similar to Bertanha and Imbens (2019).

I perform these tests in Figures E.1 and Table E.1, using the code developed by Bertanha and Imbens (2019). The left panel of Figure E.1 shows the effects conditional on not receiving treatment. The figure shows no detectable discontinuity in the outcome (years of education for the young cohort), suggesting that there is continuity of expected outcomes. This result is confirmed by the top row of Table E.1, where again there is no statistically (or economically) detectable discontinuity in outcomes for the set of regions that did not receive the policy.⁶⁵

Table E.1: Tests for Discontinuity in Outcomes Conditional on Treatment Status

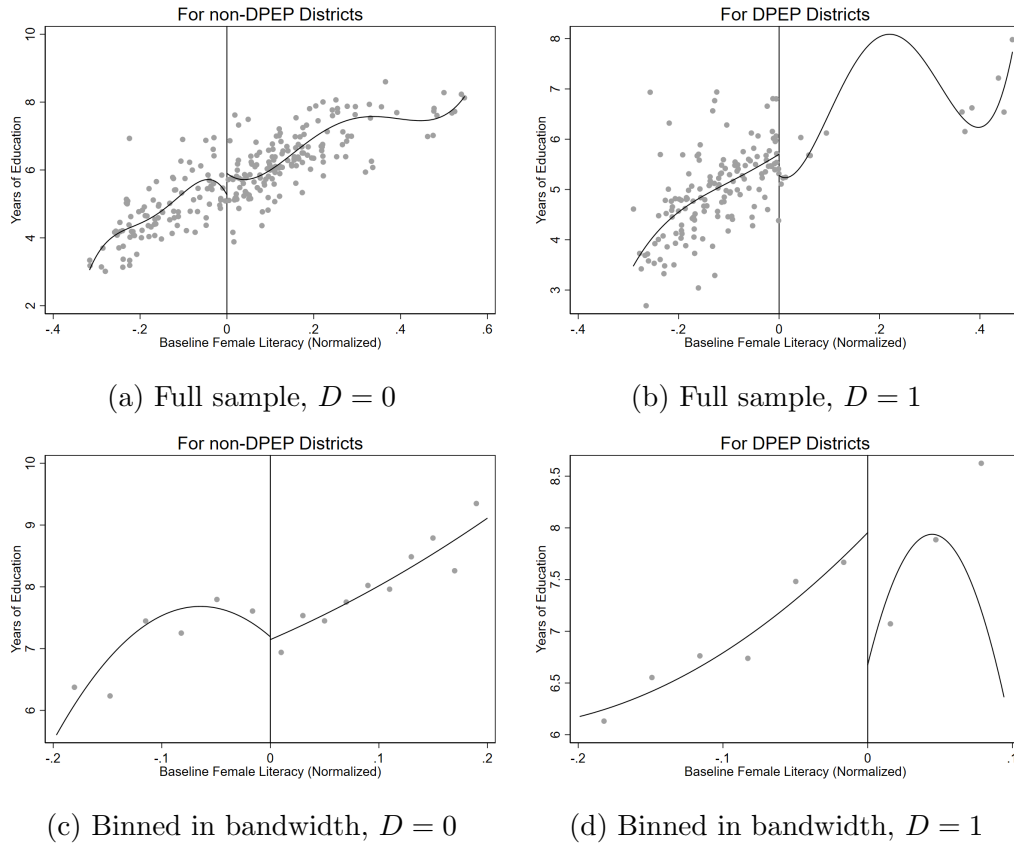
	Est	S.E.	pval	Est	S.E.	pval
	Young cohorts			Old cohorts		
$\mathbb{E}[Y D = 0, X = c+] - \mathbb{E}[Y D = 0, X = c-]$	0.274	0.272	0.313	0.357	0.406	0.379
$\mathbb{E}[Y D = 1, X = c+] - \mathbb{E}[Y D = 1, X = c-]$	-0.463	0.220	0.035	-0.933	0.490	0.057
Joint F-test	5.558		0.062	4.293		0.117

Notes: Tests for discontinuity in outcomes conditional on treatment status, using the code in Bertanha and Imbens (2019). The sample is for young cohorts only. The outcome Y is years of education, $D = 1$ if the district received DPEP, and X is the running variable (baseline female literacy). The bandwidth is determined by the code in Bertanha and Imbens (2019).

For the sample of treated regions $D = 1$, the right panel of Figure E.1 shows that there may be insufficient sample for regions with high-female literacy that were treated. Since there are almost no ‘always-takers’, there is no need to test if compliers are similar to always takers. Practically, at the cutoff, the estimates seem unstable and dependent on the functional form

⁶⁵One may also consider a similar test across compliance groups of individuals within each district (i.e. by conditioning on being skilled or unskilled), but as I show, by construction, the wages of the skilled in treated districts must be different from the wages of the skilled in untreated districts because of the GE effects.

Figure E.1: Testing for differences in outcomes across compliance groups



Notes: Figures test for equality of outcomes (here, years of education for the young cohort) between treated compliers and always-takers, and untreated compliers and never-takers, as suggested by [Bertanha and Imbens \(2019\)](#). The top panel shows results for the full sample, whereas the bottom panel shows binned averages after restricting it to a bandwidth around the cutoff. The left panel is for regions that did not receive DPEP ($D = 0$), whereas the right panel is for regions that did receive DPEP ($D = 1$). Plots made using the code in [Calonico et al. \(2014\)](#).

given the lack of sample for treated high-literacy regions. The code by [Calonico et al. \(2014\)](#) is unable to compute a bandwidth, yet the code by [Bertanha and Imbens \(2019\)](#) in Table E.1 suggests a discontinuity for $D = 1$. The joint F-test for both discontinuities suggests marginal statistical significance, driven by the $D = 1$ sample, where there is little observational support on the right of the cutoff.

As the sample for untreated regions $D = 0$ seems to have enough support on either side of the cutoff, we only rely on that test, where there is no detectable discontinuity. The lack of sample on the right of the cutoff for $D = 1$ undermines the need for a test for the treated sample. The results for the $D = 0$ sample lends support to the external validity of these estimates to other compliance groups at the threshold.

E.II.2 Marginal Treatment Effects (MTE) at the Cutoff

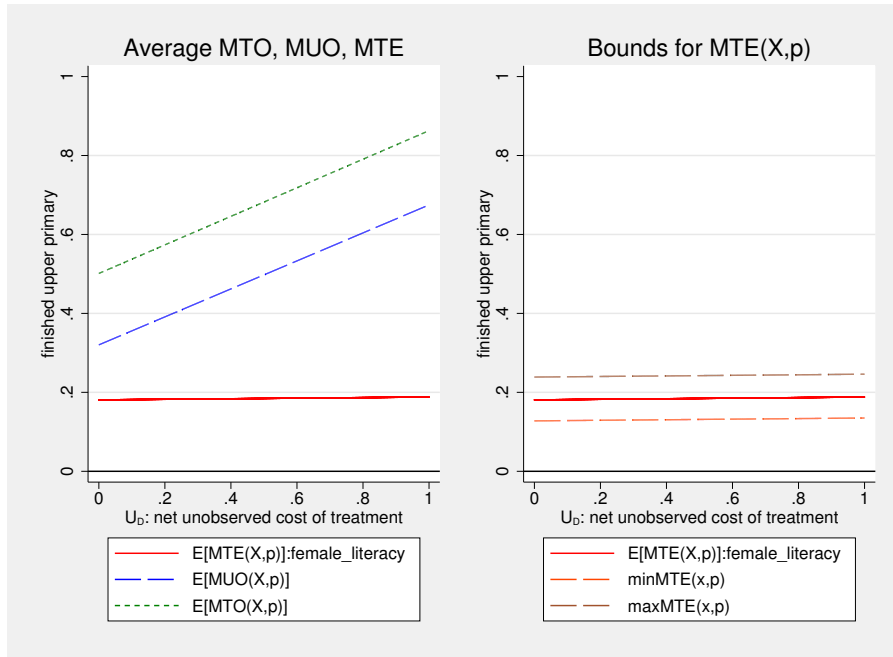
While the previous set of results suggest that different compliance groups at the threshold have similar responses, this can be examined in further detail using methods developed by [Brinch et al. \(2017\)](#); [Heckman and Vytlacil \(2007\)](#); [Kowalski \(2019\)](#), which estimate the Marginal Treatment Effect (MTE) across the distribution of the costs of being treated.

Intuitively, the net ‘benefit’ from being treated depends on observed variables (like the baseline female literacy rate) and unobserved net costs U_D . Being above the female literacy cutoff raises the likelihood of receiving the program. Yet, if the unobserved costs of being treated

are high, then a district chooses not to receive funds (‘never takers’). Brinch et al. (2017); Kowalski (2019) use a generalized Roy model to determine how to estimate the treatment effect across the distribution of U_D even with a single binary instrument (being above the cutoff, conditional on baseline female literacy).

If the MTE varies across U_D , then that may suggest that the LATE for complier districts is different from the LATE for never-takers or always-takers, where $LATE(x) = \frac{1}{p_1(x) - p_0(x)} \int_{p_0(x)}^{p_1(x)} MTE(x, p) dp$. Here, $p_1(x)$ is the propensity to be treated when above the cutoff, and $p_0(x)$ the treatment probability below the cutoff.

Figure E.2: Marginal Treatment Effects



Notes: Figures estimate the Marginal Treatment Effect (MTE), the Marginal Untreated Outcome (MUO) and Marginal Treated Outcome (MTO). Two-staged least squares regressions, where the first stage outcome is receiving the policy on being above the cutoff. Sample restricted to the Calonico et al. (2014) optimal bandwidth, and controlling linearly for the running variable. Outcome is ‘Finishing upper primary school’, which determines the skill level. Corresponding LATE for compliers only in third panel of Table A.6.

Figure E.2 shows the result of this exercise, where the first stage is receiving the policy, and the second stage is becoming skilled (finishing upper primary). The instrument is being above the cutoff, conditional on being in the optimal bandwidth, and controlling linearly for the running variable. The MTE is roughly flat across the entire distribution of U_D . Intuitively, this suggests that the MTE for the always takers, is similar to the MTE for the compliers and they are both similar to the MTE for the never takers. These results reinforce the lack of meaningful treatment effect heterogeneity across compliance groups, as in Appendix E.II.

E.II.3 Do District Characteristics Affect Treatment Probability?

An additional test of external validity is related to how the treatment probability varies by different district-level characteristics. For instance, if the discontinuity in treatment probability is differentially higher in regions with more retail trade then it may suggest that some economic differences across compliance groups. To implement this test, I estimate a parametric RD, in the following manner:

$$Received Program_{id} = \alpha + \beta_1 \mathbb{1}_{Literacy_d < 0} + \beta_2 (\mathbb{1}_{Literacy_d < 0} \times X_d) + f(Literacy_d, X_d) + \epsilon_{id} \quad d \in -D, D, \quad (A.38)$$

Table E.2: Changes in Treatment Probability By District Characteristics

	Received Program		
$\mathbb{1}_{Literacy_d < 0}$	0.499*** (0.105)	0.468*** (0.103)	0.531*** (0.0962)
$\mathbb{1}_{Literacy_d < 0} \times$ Skilled occu	0.0305 (0.170)		
$\mathbb{1}_{Literacy_d < 0} \times$ Non agriculture		0.102 (0.169)	
$\mathbb{1}_{Literacy_d < 0} \times$ Retail trade			-0.0358 (0.168)
Observations	367,043	367,043	367,043
R-squared	0.372	0.381	0.360

Notes: Parametric regressions in a bandwidth of 25 ppt around the cutoff for the full sample (all age groups, regardless of being wage earners). The running variable is female literacy from the 1991 Census, normalized at the national average (the cutoff). Outcome is the probability of receiving the DPEP program. ‘Skilled occupations’, defined according to the National Classification of Occupations 2004. The largest excluded (i.e. unskilled) categories are workers in agriculture, mining, forestry, fishing, construction, street vendors, and “other elementary occupations.” ‘Non agriculture’ = 1 if the district has above the median (within bandwidth) share of employment in non-agricultural occupations. ‘Retail trade’ = 1 if the district has above the median (within bandwidth) share of employment working in retail trade. Linear controls for a linear function of the running variable on each side of the cutoff (interacted with the heterogeneity X_d variable), whereas quadratic controls for the quadratic function of the running variable on each side of the cutoff. Standard errors clustered at the district level.

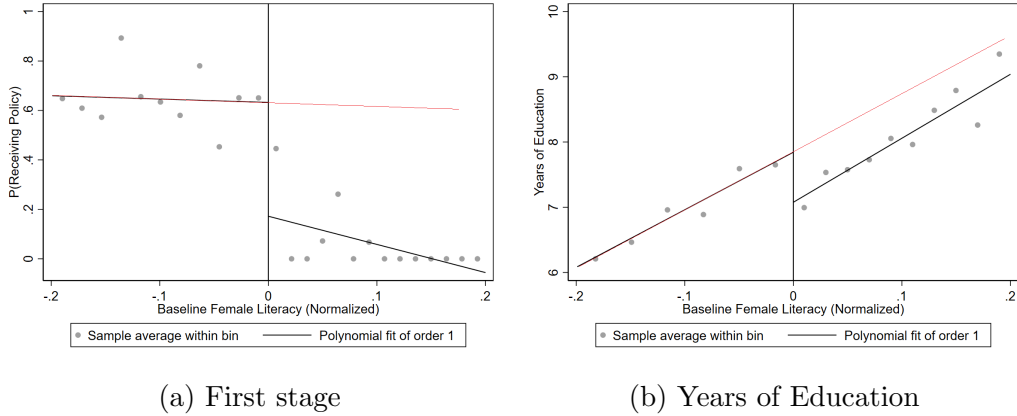
where $Received\ Program_{id}$ is an indicator for treatment status, $Literacy_d$ is the normalized baseline female literacy rate (the running variable), X_d is a district level characteristic, and $f(\cdot)$ is a control function. In Table E.2, the control function is either linear or quadratic in literacy, with slopes varying around the cutoff. I restrict the sample to 20 percentage points around the cutoff, and cluster errors at the district level.

I look at heterogeneity in district-level compliance across three dimensions. First, the share of workers working in high-skilled occupations in the district being above the median across districts. Second, the share of employment in non-agricultural occupations, and third the share in retail trade being above the median across districts. Across each of these three dimensions, there does not seem to be strong evidence of heterogeneity in compliance probabilities, as the interaction terms are statistically and economically insignificant. This lack of evidence may lend some additional support to the external validity of the estimates at the threshold.

E.II.4 Estimates Away from the Cutoff

Next, I examine the change in treatment effects as we locally vary the cutoff. Dong and Lewbel (2015) show that at the cutoff, we can estimate not just the size of the discontinuity, but also the change in the first derivative (or higher order derivatives) of the regression function. These higher order derivatives allows us to extrapolate away from the cutoff. Dong and Lewbel (2015) define the marginal threshold treatment effect (MTTE) to be the change in the treatment effect that would result from a marginal change in the threshold. Under the assumptions discussed by them, the estimation of the MTTE requires us to estimate the following model:

Figure E.3: Higher-order Derivative Based Extrapolation



(a) First stage

(b) Years of Education

Notes: Figures show the first-stage and years of education for young wage-earners after restricting the bandwidth to 20 ppt around the cutoff. Plots made using the code in [Calonico et al. \(2014\)](#).

$$Y_{id} = \alpha + \beta_1 \mathbb{1}_{Literacy_d < 0} + \beta_2 Literacy_d + \beta_3 (\mathbb{1}_{Literacy_d < 0} \times Literacy_d) + \epsilon_{id} \quad d \in -D, D, \quad (\text{A.39})$$

where Y_{id} is the outcome of interest, $Literacy_d$ is the normalized baseline female literacy rate (the running variable), and $\mathbb{1}_{Literacy_d < 0}$ is an indicator for literacy being below the cutoff. In a sharp design, and under their definition of local policy invariance, β_3 is a consistent estimator of the MTTE. While this is simply the standard parametric RD estimation technique, the observation is that these widely used local linear estimators also provide estimates of *derivatives* which are then used to recover the MTTE. For a fuzzy design, we may also look at the same relationship where the outcome is the first-stage.

Table E.3: The Change in the Treatment Effect for a Marginal Change in the Threshold

	P(Received Policy)		Years of Education	
$\mathbb{1}_{Literacy_d < 0}$	0.482*** (0.0974)	0.478*** (0.0974)	0.806*** (0.309)	0.809*** (0.307)
MTTE	-0.0116 (0.00755)	-0.0396 (0.0318)	0.0131 (0.0295)	0.0401 (0.104)
Specification	Linear	Quadratic	Linear	Quadratic
Observations	20,811	20,811	20,811	20,811
R-squared	0.434	0.436	0.020	0.020

Notes: Parametric regressions in a bandwidth of 20 ppt around the cutoff for young workers (below 35 in 2009) that are wage earners. Marginal Threshold Treatment Effect (MTTE) estimated as in [Dong and Lewbel \(2015\)](#), based on the interaction between $\mathbb{1}_{Literacy_d < 0}$ and the running variable. The running variable is female literacy from the 1991 Census, normalized at the national average (the cutoff). Outcome in the first two columns is the probability of receiving the DPEP program. Outcome in the last two columns is years of education for young wage earners. Linear controls for a linear function of the running variable on each side of the cutoff, whereas quadratic controls for the quadratic function of the running variable on each side of the cutoff. Standard errors clustered at the district level.

The intuition is that the change in the slope of the relationship between the outcome and running variable, across the cutoff, is informative of how the treatment effect may change as

we vary the cutoff locally along the running variable. If the slope of the relationship is similar on either side of the cutoff, then we may expect that as we marginally change the value of the cutoff, the magnitude of the treatment effect would not change.

Figure E.3 estimate the first-stage relationship, and the relationship for the main outcome (years of education) in a wide bandwidth around the cutoff. Table E.3 estimates the MTTE using the Equation (A.39) as outlined by Dong and Lewbel (2015). While the left-panel of Figure E.3 shows a slight change in the slope at the cutoff, this is not a statistically (nor economically) meaningful change as estimated by Table E.3. The right-panel of Figure E.3 shows almost no change in the slope for the main outcome of interest; a relationship that is confirmed by Table E.3. These results suggest that extrapolating at least locally around the cutoff, would produce minor changes to treatment effects. Yet, we may caution against extrapolating this to further from the cutoff.

E.III Returns Varying Across Schooling Levels

We have a single instrument (the RD) affecting a discrete endogenous regressor (levels of schooling), which in turn affects earnings. Lochner and Moretti (2015) make show that since OLS and IV estimators identify different weighted averages of the per-unit effects, the standard Hausman (1978) test is no longer valid. They develop a new exogeneity test, by reweighting the OLS per-unit effects with the IV weights. In a standard model, we estimate:

$$\text{Log Wages}_i = \beta \text{Schooling}_i + \mu_i, \quad (\text{A.40})$$

where Schooling_i is potentially endogeneous, but has more than two discrete values. We may expect the effects between wages and schooling to be somewhat non-linear:

$$\text{Log Wages}_i = \sum_{j=1}^S D_{ij} \beta_j + \epsilon_i, \quad (\text{A.41})$$

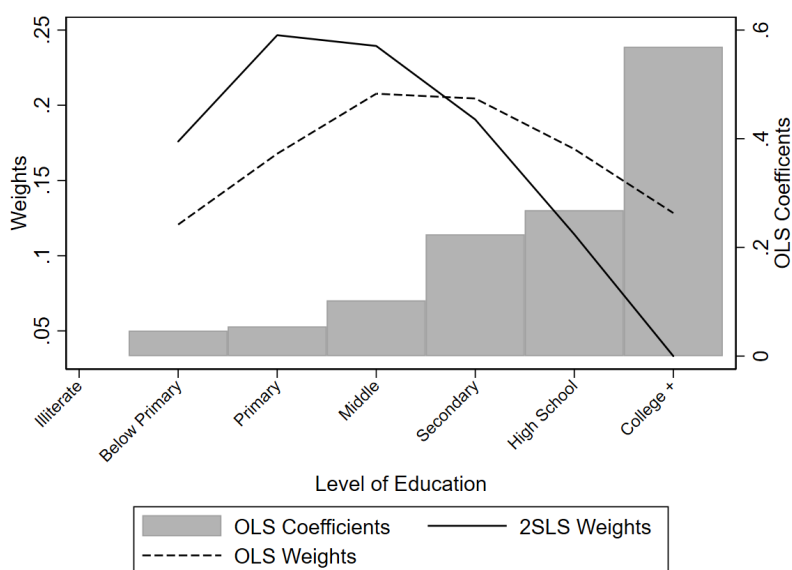
where D_{ij} are indicators for different levels of schooling j . In a case with a single discrete instrument (as in the current context), the weights on the IV and OLS estimates of β will differ. To compare the IV and OLS estimates to perform an exogeneity test, Lochner and Moretti (2015) suggest reweighting the OLS estimates of β_j to recover a comparable estimate.

Of course, this only tests if the weighted average of all OLS $\hat{\beta}_j$ asymptotic biases equals zero. It is not meant to test, for instance, if some OLS $\hat{\beta}_j$ are biased upwards and others downwards in a way that cancel each other out. Furthermore, this procedure does not allow us to recover any IV estimates of β_j as we hypothetically have only one instrument.

Figure E.4 show the OLS coefficients cumulatively relative to being illiterate. Higher levels of education are associated with more earnings. As the increments are roughly similar, these coefficients do not seem to display large amounts of non-linearities till we reach the college-level. College graduates earn a lot more, but the difference in years of education between high-school and college graduates are about five years, whereas the difference across other groups (like middle to secondary, or secondary to high school) are only two years. As such, we may think that adjusted for years of education, the increments in returns are somewhat similar.

Yet, the OLS and 2SLS weights differ. The IV induces students to have more primary and middle (upper primary) levels of education, and zero weight on college graduates. This is consistent with the results in the paper, and the policy's targeting of primary and upper-primary levels. Both the standard Hausman (1978) and Lochner and Moretti (2015) reject the exogeneity of the OLS estimate.

Figure E.4: OLS estimates of wage returns and IV weights



Notes: Dependent variable log earnings. Omitted level of education is “illiterate” – as such the returns are cumulatively relative to being illiterate. Sample of young wage earners in bandwidth of 20 ppt around the cutoff. Test as described by [Lochner and Moretti \(2015\)](#). IV control function includes a quadratic in the running variable.

F Heterogeneity, Robustness and Additional Results

F.I Defining Cohorts

The DPEP program was started at the end of 1993, and anyone who was past school-going age at that time should be relatively unaffected. Yet, we need to determine a cutoff for school-going age. 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 (such as The Factories Act of 1948 and the Mines Act of 1952) prevent many workplaces from hiring children below eighteen. 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. As I show, 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. 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.7](#)). 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.

F.II Heterogeneity by District Characteristics

The returns to education are not an unchanging economic parameter, but rather an endogenous variable that depends on the features of the local economy. To that end, one may explore how the Wald estimate for the returns to education (the ratio of log wages and education), differ by underlying economic characteristics. In Table [F.1](#), I divide the data at the district level, at the median value of some underlying feature of the labor or education market. Slicing the data thinly in an RD setup, however, is not ideal, and there may not be sufficient power to estimate differences across returns. Yet, the results suggest some meaningful heterogeneity.

For instance, as an example of possible heterogeneity in returns I explore how (the Wald estimate) returns differ by the fraction of workers with vocational training. This picks up the complementarities in production between the newly young skilled and old vocationally trained. Table F.1 shows that returns are higher (about 17.3%) in places where the old have less vocational education, rather than places with high vocational education (about 13.4%).

Similarly, places that have a higher manufacturing share, tend to have a higher return (about 17.3%), than places with a lower share (about 14.7%). These returns, however, are not statistically indistinguishable from each other, due to a lack of power.

F.III Evidence on Adoption of Capital and Skill-biased Capital

Firms may adopt more capital and technology in response to a locally skilled workforce. In the top few panels of Figure F.1, one can see that the average compensation to workers increases as more and more educated workers start joining the labor market around the year 2004. The figure also shows increases in the fraction of firms producing mechanized products. These are suggestive of the fact that either existing firms shifted production and employed more skilled workers, or new firms entered and hired these skilled workers. Both findings are suggestive evidence in support of the adoption of skill-biased capital in 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 F.1, 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.

F.IV Results and Distributions of Components of the Returns

In the text, I discuss how I calculate the returns to skill using Equation (17), and the change in returns using Equations (15) and (16). Here, I walk through the details of the implementation, and show the different components of the returns to skill. First, from Equation (17), we see that the estimates of the partial equilibrium returns to skill, $\beta_{ys,D=0}$, depend on quantities like the fraction of young skilled, $\ell_{sy,D=1}$, and unskilled, $\ell_{uy,D=1}$, workers. For inference, I bootstrap my estimates, sampling with repetition, 1500 times, so as to estimate standard errors for the returns.

$$\underbrace{\log \frac{w_{y,D=1}}{w_{y,D=0}}}_{\text{Avg earnings}} = \underbrace{\ell_{sy,D=1}}_{\text{Skilled}} \underbrace{\log \frac{w_{sy,D=1}}{w_{sy,D=0}}}_{\neq 0} + \underbrace{\ell_{uy,D=1}}_{\text{Unskilled}} \underbrace{\log \frac{w_{uy,D=1}}{w_{uy,D=0}}}_{\neq 0} + \underbrace{\Delta \ell_{sy}}_{\text{Compliers}} \underbrace{\log \frac{w_{sy,D=0}}{w_{uy,D=0}}}_{\beta_{ys,D=0}} \quad (17)$$

Equation (17) shows that the change in average wages for the young is a weighted average of the change in the young skilled wage (weighted by the fraction skilled), the young unskilled wage (weighted by the fraction unskilled), and the returns to skill (weighted by the fraction of compliers). Figure F.2 describes these components. The left hand side of the equation is already described in the bottom panel of Table A.6, a coefficient of 0.258.

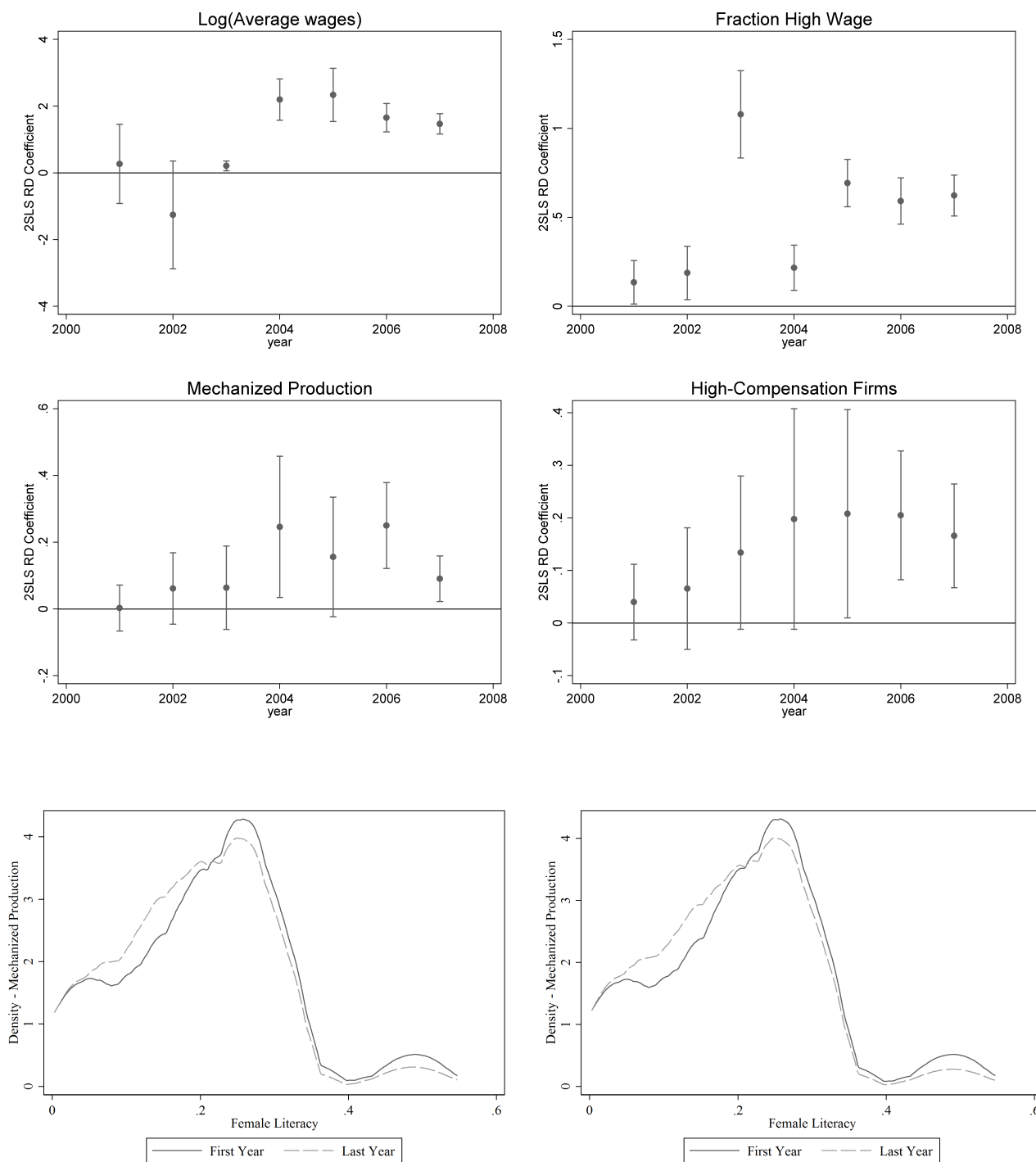
The top left panel of Figure F.2 shows the fraction skilled, $\ell_{sy,D=1}$, and thereby also the fraction unskilled, $\ell_{uy,D=1} = 1 - \ell_{sy,D=1}$. The top right panel and the middle left panel show the RD coefficients of the changes in skilled $\log \frac{w_{sy,D=1}}{w_{sy,D=0}}$ and unskilled wages $\log \frac{w_{uy,D=1}}{w_{uy,D=0}}$. These

Table F.1: Heterogeneity by district characteristics (young workers)

	High manufacturing share		Low manufacturing share	
	Years of Education	Log(wages)	Years of Education	Log(wages)
RD Estimate	0.781*** (0.257)	0.115*** (0.0412)	0.488* (0.291)	0.0844* (0.0465)
Observations	18,584	18,584	11,983	11,985
	High vocationally trained population		Low vocationally trained population	
	Years of Education	Log(wages)	Years of Education	Log(wages)
RD Estimate	0.694** (0.303)	0.0928* (0.0481)	0.585** (0.252)	0.101** (0.0407)
Observations	16,647	16,647	13,920	13,922
	High regular salaried employment		Low regular salaried employment	
	Years of Education	Log(wages)	Years of Education	Log(wages)
RD Estimate	0.949*** (0.283)	0.303*** (0.0459)	0.534** (0.264)	-0.0413 (0.0419)
Observations	17,090	17,091	13,477	13,478
	High retail share		Low retail share	
	Years of Education	Log(wages)	Years of Education	Log(wages)
RD Estimate	0.681** (0.344)	0.266*** (0.0578)	0.538** (0.238)	0.0634* (0.0381)
Observations	13,871	13,871	16,696	16,698
	High construction share		Low construction share	
	Years of Education	Log(wages)	Years of Education	Log(wages)
RD Estimate	0.0325 (0.294)	-0.0391 (0.0464)	0.938*** (0.261)	0.196*** (0.0419)
Observations	14,525	14,525	16,042	16,044

National Sample Survey 2009-10 for persons between 16-35 years of age (young cohort) who reported earnings. Optimal bandwidths from [Calonic et al. \(2014\)](#). Coefficients measure the change in outcomes on crossing the RD cutoff. ‘High vocational trained population’ is a sub-sample for if the fraction of population that received vocational training is above the median across districts. ‘High manufacturing/retail/construction share’ is a sub-sample for if the fraction of older-cohorts that work in the sector, lies above the median across districts. ‘High regular salaried employment’ is a sub-sample for if the share of workers that have regular salaried jobs, lies above the median across districts.

Figure F.1: Adoption of Skill Biased Capital: Firm-Level Data



Density Above Cutoff: Mechanized Production

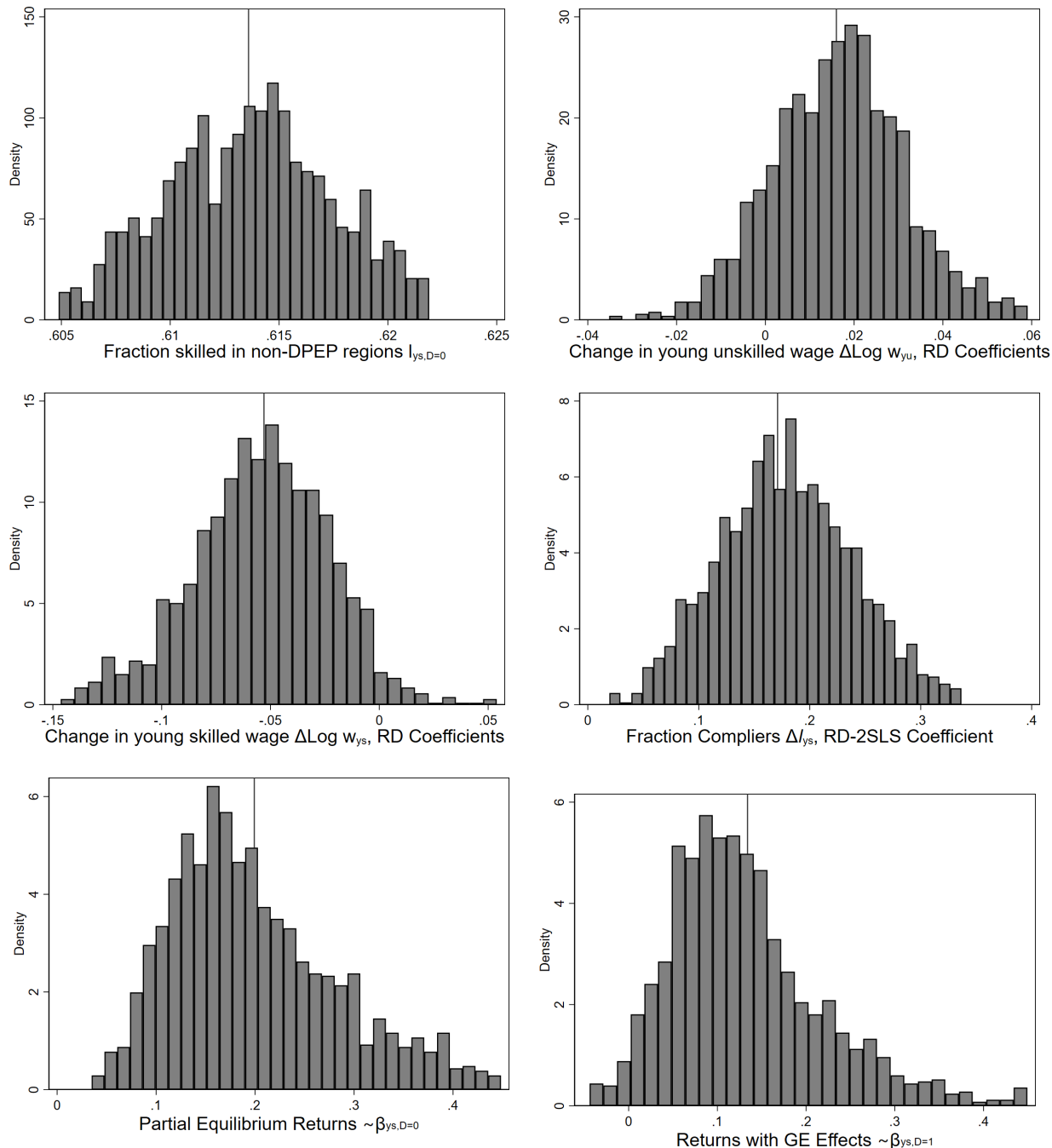
Density Above Cutoff: High Compensation

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. \(2014\)](#) 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.

RD coefficients are estimated using the optimal bandwidths. The middle right panel shows the fraction of young that switched from going from unskilled to skilled along the RD cutoff, $\Delta \ell_{sy}$. This is the same coefficient that is described in the ‘Finished upper-primary’ panel of [Table A.6](#),

a coefficient of 0.171. This is also what constitutes the top panel of Table 3. Using Equation (17), therefore, the corresponding partial equilibrium returns to education, $\tilde{\beta}_{ys,D=0} = \frac{\beta_{ys,D=0}}{\Delta s}$, are shown in the bottom left panel Figure F.2—a value of 0.199. This is shown in the middle panel of Table 3.

Figure F.2: Bootstrapped distributions of components of returns to skill



Figures show the bootstrapped distributions of the different components of the returns to skill, as described by Equation (17). We perform 1500 iterations, sampling with repetition. The 2SLS RD coefficients calculated using [Calonico et al. \(2014\)](#) procedure. Vertical lines indicate the mean of the distribution.

We can also decompose the categories that determine the change in returns (discussed in the top panel of Table 3). To reiterate, Equations (15) and (16) show the determinants of the changes in the returns.

Figure F.3 describes these changes. The GE effect on all cohorts can be seen by looking at the change in the skill premium of the old, as show by the left hand side of Equation (15). The top panel of Figure F.3 describes this change. The difference between the two is the part of the GE effect on all cohorts, discussed in the bottom panel of Table 3.

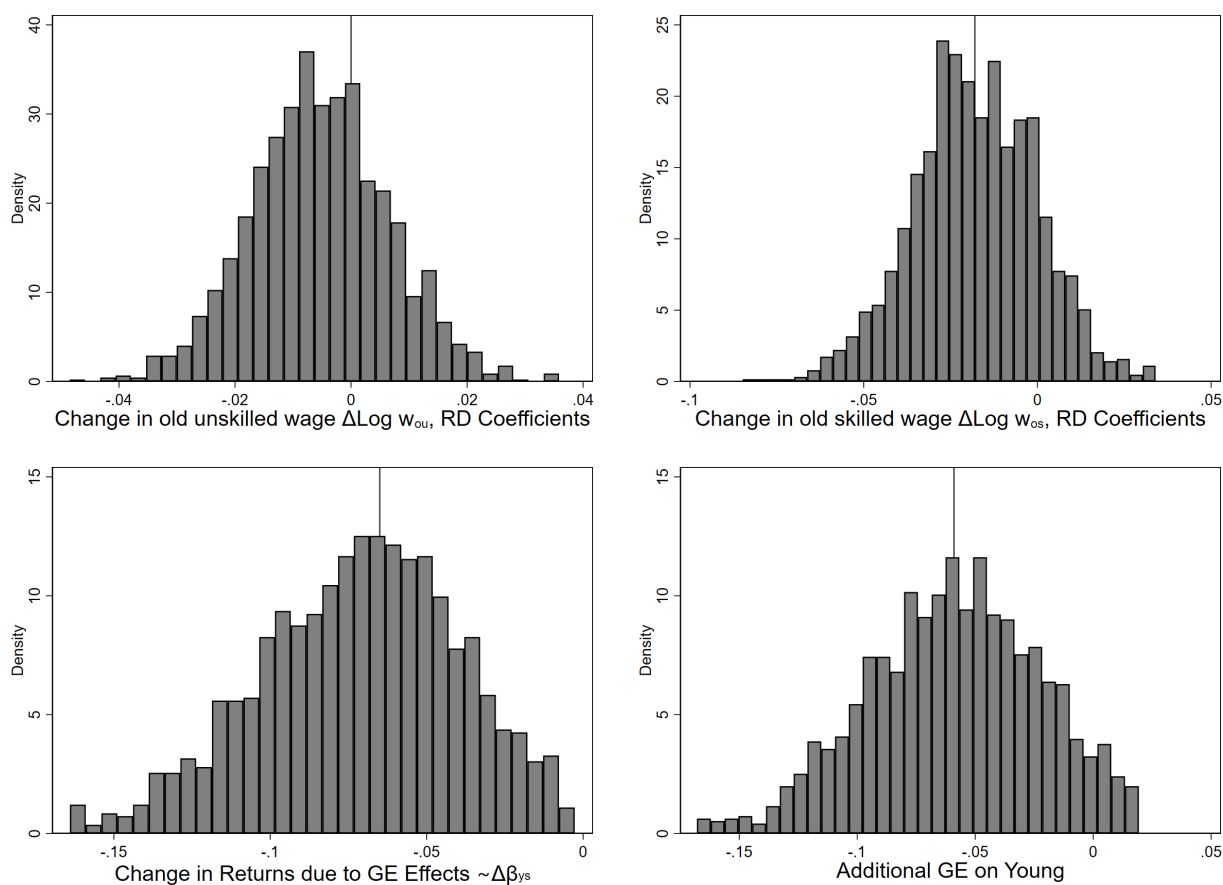
The change in the returns to education, $\Delta\tilde{\beta}_{ys}$, is the change in the skill premium for the young cohorts (the difference between the top-right panel and bottom left panel of Figure F.2). This 6.6p.p. is what is shown in the top panel of Table 3, and is the difference between partial equilibrium returns (19.9%), and the returns with GE (13.4%).

The additional GE effect on the young is described by Equation (15). Here, we need the change in returns, $\Delta\tilde{\beta}_{ys}$, and the ‘GE effect on all cohorts’ discussed above. This estimate is shown in the bottom panel of Table 3, and the bootstrapped distribution shown in the bottom panel of Figure F.3.

$$\underbrace{\log \frac{w_{so,D=1}}{w_{so,D=0}} - \log \frac{w_{uo,D=1}}{w_{uo,D=0}}}_{\text{GE effects on all cohorts}} = \underbrace{\left(\log \frac{\theta_{s,D=1}}{\theta_{u,D=1}} - \log \frac{\theta_{s,D=0}}{\theta_{u,D=0}} \right)}_{\text{Skill biased capital}} + \underbrace{\left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E} \right)}_{\text{Aggregate skill distribution}} \left[\log \frac{L_{s,D=1}}{L_{u,D=1}} - \log \frac{L_{s,D=0}}{L_{u,D=0}} \right] \quad (15)$$

$$\underbrace{\left[\log \frac{w_{sy,D=1}}{w_{sy,D=0}} - \log \frac{w_{uy,D=1}}{w_{uy,D=0}} \right]}_{\text{Additional GE on young}} - \underbrace{\left[\log \frac{w_{so,D=1}}{w_{so,D=0}} - \log \frac{w_{uo,D=1}}{w_{uo,D=0}} \right]}_{\text{Age specific skill distribution}} = -\frac{1}{\sigma_A} \left[\log \frac{\ell_{ys,D=1}}{\ell_{yu,D=1}} - \log \frac{\ell_{ys,D=0}}{\ell_{yu,D=0}} \right] \quad (16)$$

Figure F.3: Bootstrapped distributions of components of changes in returns



Figures show the bootstrapped distributions of the different components of the changes in returns to skill, as described by Equations (15) and (16). We perform 1500 iterations, sampling with repetition. The 2SLS RD coefficients calculated using [Calonico et al. \(2014\)](#) procedure. Vertical lines indicate the mean of the distribution.