

The Targeting Benefit of Conditional Cash Transfers

Katy Bergstrom

William Dodds



WORLD BANK GROUP

Development Economics

Development Research Group

January 2020

Abstract

Conditional cash transfers (CCTs) are a popular type of social welfare program that make payments to households conditional on human capital investments in children. Compared to unconditional cash transfers (UCTs), CCTs may exclude some low-income households as access is tied to normal investments in children. This paper argues that conditionalities on children's school enrollment offer an unexplored targeting benefit over UCTs: CCTs target money to households that forgo a discrete amount of child income. This paper shows that the size of this targeting benefit is directly related to the distribution of parental

incomes, the size of forgone child incomes, and two elasticities already popular in the literature: the income effect of a UCT and the price effect of a CCT. These elasticities are estimated for a large CCT program in rural Mexico, Progresa, using variation in transfers to younger siblings to identify income effects. In this setting, the analysis finds that the targeting benefit is almost as large as the cost of excluding some low-income households; this implies that 41 percent of the Progresa budget should go to a CCT over a UCT based on targeting grounds alone.

This paper is a product of the Development Research Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at kbergstrom@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

The Targeting Benefit of Conditional Cash Transfers*

Katy Bergstrom[†]

William Dodds[‡]

*We would like to thank the members of our committee Raj Chetty, Pascaline Dupas, Melanie Morten, and Petra Persson for their extremely helpful comments, guidance, and support on this project. We would also like to thank Arun Chandrasekhar, Marcel Fafchamps, Caroline Hoxby, Juan Rios, and Meredith Startz for their very helpful comments and suggestions, as well as participants at various seminars at Stanford University and the World Bank for their useful comments. Finally, we'd like to thank The Ric Weiland Graduate Fellowship in the School of Humanities and Sciences for financial support.

[†]Development Research Group, World Bank. Email: kbergstrom@worldbank.org.

[‡]Charles River Associates. Email: wdodds@crai.com.

I INTRODUCTION

Conditional cash transfers (CCTs), cash transfers targeted to poor households made conditional on investments in children’s human capital, have dramatically risen in prominence over the last two decades (Fiszbein and Schady, 2009). In 2016, 63 low- and middle-income countries had at least one CCT program, up from 2 countries in 1997 (Bastagli et al, 2016). CCTs aim to both alleviate current poverty by targeting transfers to poor households and reduce future poverty by tying access of transfers to investments in children’s human capital. However, it is argued that these two aims can be at odds with one another as low-income households may find the conditions too costly to comply with and thus be excluded from receiving aid; i.e., human capital investments in children are typically normal investments (Baird et al, 2011; Freeland, 2007). Unconditional cash transfers (UCTs), cash transfers targeted to poor households but with “no strings attached”, are therefore thought to be superior at alleviating current poverty.

This paper argues that there exists an unexplored targeting benefit of imposing conditions on cash transfers to send children to school. The central idea behind this benefit is that sending a child to school is a lumpy investment where the opportunity cost of this investment is forgone child income. CCTs are beneficial from a targeting perspective because they direct money to the set of households who forgo a large amount of child income. We argue that this benefit mitigates the adverse targeting effects that arise from imposing conditions on a normal investment. Thus, we argue that there exists a targeting trade-off when choosing how to allocate a budget between a CCT and a UCT, and that it may therefore be optimal to allocate some of a budget to a CCT based on targeting grounds alone. The objective of this paper is to theoretically highlight and empirically quantify this targeting trade-off so as to ascertain the extent to which the targeting benefit is an important advantage CCTs can offer over UCTs.

To illustrate this targeting trade-off, consider a set of parents earning heterogeneous incomes facing the decision of whether to send their one child to school or not. Parents value their child’s education but sending a child to school means forgoing the income they could earn by working instead. Thus, parents trade-off higher household consumption today with improved future outcomes for their child. All else equal, there will exist a parent with income \tilde{y} who is just indifferent between sending their child to school or to work, and parents with income above this cutoff will choose to send their child to school while those below will not. Just above this cutoff, household consumption will discretely drop by the amount of the child’s potential earnings; this discrete drop in household consumption is illustrated in Figure I below, where y_{child} denotes potential child income. Now consider distributing a budget to these households via a CCT vs a UCT. Relative to a UCT, a CCT will exclude those households earning parental incomes below \tilde{y} from

receiving aid.¹ However, a CCT will direct more money to those households forgoing child income, y_{child} , who are potentially lower consumption households.

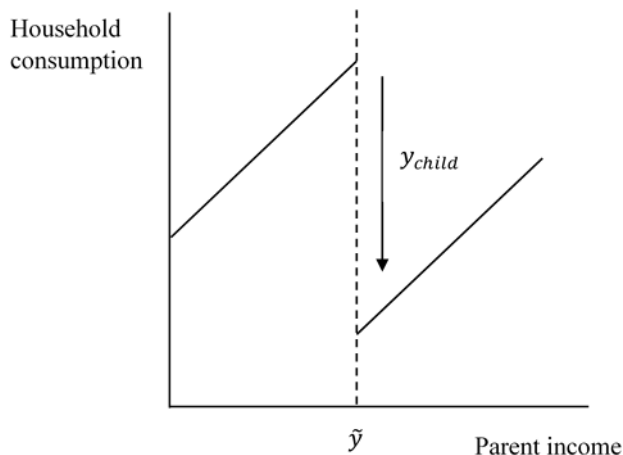


FIGURE I: HOUSEHOLD CONSUMPTION VS PARENT INCOME

The first goal of this paper is to develop a theoretical model that captures the targeting trade-off of CCTs versus UCTs. We consider a two-generation world in which households consist of a parent and a child and are heterogeneous with respect to parent income and child ability. Parents maximize household utility over the two generations by choosing whether to send their child to school or to work in the first generation. Schooling results in higher utility in the second generation but comes at the cost of a discrete loss in consumption in the first generation. We make the very realistic assumption that parents cannot borrow across generations, i.e., they cannot borrow against their child's future earnings. We consider a utilitarian social planner whose objective is to maximize the sum of lifetime utility across a set of eligible households, where the set of eligible households has been predetermined (say via proxy-means testing or geographical targeting).² The planner has a fixed budget to distribute to these households in the first generation and has to decide how to allocate this budget between a CCT versus a UCT. Because we wish to focus solely on the targeting effects of offering a CCT over a UCT, we assume parents make education decisions at the socially optimal level, thereby abstracting from conventional motives to condition transfers; consequently, the planner's sole objective is to transfer resources to the households that have the highest marginal utility of consumption in the first generation.³

1. For this simple example, we ignore behavioral responses to transfers.

2. As shown in Hanna and Olken (2018), moving from universal transfers to transfers that are targeted to a subset of the population (via say proxy means testing) typically results in large welfare gains. Moreover, the vast majority of CCT and UCT programs are only offered to a subset of the population (Baird et al, 2013). Of course, choosing the subset of the population with whom to distribute to is also a choice variable for the planner. For simplicity, we abstract from this decision and take the set of eligible households as given. Thus, our set-up can be viewed as solving the planner's interior problem.

3. Because the planner only has a budget to redistribute in the first generation, and because households

The planner faces the following trade-off: by increasing the CCT, she increases the share of money received by the schooling households (i.e., the households who forgo child income). We refer to the welfare gain experienced by the schooling households as the targeting benefit. However, this comes at the cost of decreasing the UCT, which in turn decreases the share of money received by the non-schooling households (i.e., the households who have lower average parental income). We refer to the welfare loss experienced by the non-schooling households as the exclusion cost. Our first result is that under some conditions, the targeting benefit can outweigh the exclusion cost, meaning that it is optimal for the planner to allocate some or potentially all of her budget to a CCT. This result goes against the current consensus that CCTs are unambiguously worse at targeting transfers within a set of beneficiary households.

Under what conditions will the targeting benefit exceed the exclusion cost? The targeting benefit can only dominate if the schooling households have, on average, higher marginal utility of consumption (i.e., lower consumption) in the first generation. We show that there are three factors that influence the size of the targeting benefit relative to the exclusion cost: (1) the concavity of utility of consumption, (2) the distribution of parental incomes among eligible households, and (3) the size of potential child incomes. First, in order for the targeting benefit to ever exceed the exclusion cost, there needs to be some degree of curvature in the utility of consumption. If utility is linear in consumption, marginal utilities are constant across all households, thus eliminating any motive for the planner to target transfers towards specific households. Second, we show that as parental income inequality increases among the eligible set of households, the targeting benefit falls while the exclusion cost rises. As income inequality rises, the households sending their child to school become richer relative to the households not sending their child to school (as schooling is a normal investment). Third, the larger potential child incomes, the greater the “lumpiness” of the investment, thus the greater the targeting benefit relative to the exclusion cost.

One potential concern with our framework is that while the schooling households may have lower consumption in the first generation, by revealed preference, they have higher lifetime utility. Thus, CCTs will direct more money to those households with higher lifetime utility. The reason it can still be optimal to give to higher lifetime utility households emerges from two modeling assumptions: (1) parents cannot borrow or save across generations, and (2) the planner only has a budget to give out in the first generation. Thus, the planner’s objective to maximize lifetime utility across all eligible households simply

cannot move resources across generations, the planner’s objective to maximize total lifetime utility of all eligible households translates into an objective to direct money to those households with the highest marginal utility of consumption in the first generation. This type of multi-generational welfare function is identical to that often used in the dynamic optimal taxation literature, e.g., Kocherlakota (2005) or Golosov (2003).

amounts to an objective to direct funds to those who value them the most today (as the planner has no way to directly increase household utility in the second generation). While assumption (1) is made for realism, assumption (2) is made because it allows us to capture the targeting trade-off in a parsimonious fashion: because parents cannot move resources across generations, the targeting trade-off the planner faces when distributing a budget today will be unaffected when considering a more complete set-up where the planner also has a budget to redistribute in future generations. When moving to such a set-up, the concern that we on net give more money to higher lifetime utility households is mitigated. This is because children that go to school today will have higher incomes tomorrow, so current beneficiaries of CCTs will be less likely to meet eligibility criteria for future transfers. Hence, households who do not send their child to school will receive higher transfers in future periods.⁴

The second goal of this paper is to investigate whether the empirical magnitude of the targeting benefit is large enough to warrant policy relevance. To do so, we first express the targeting benefit relative to the exclusion cost in terms of empirically observable objects. The objects we require are the distribution of parental incomes, the size of potential child incomes, and two elasticities already popular in the literature: the income effect of a UCT and the price effect of a CCT. The income effect measures the change in the share of children enrolled in school as the UCT increases, while the price effect measures the change in the share enrolled as the CCT increases. In a similar spirit to Chetty (2006), we use these elasticities to pin down the curvature of utility of consumption. A large income effect implies that the share of children enrolled in school increases sharply when we increase the value of the UCT. This implies that either (1) households' marginal utility of consumption is diminishing quickly so that the opportunity cost of schooling is decreasing quickly as households get wealthier, or (2) the density of households who are near indifferent to sending to school is high, or (3) the return to schooling is not changing quickly with child ability (i.e., schooling and child ability are not strong complements). However, if (2) or (3) are true, the price effect will also be large. Intuitively, if the *ratio* of the income effect to the price effect is high, this implies marginal utility is decreasing quickly, so that there is substantial curvature in utility of consumption.

We then move to our empirical application where we estimate the targeting trade-off in the setting of Progresa, a large CCT program in rural Mexico. Starting in 1997, Progresa was one of the first CCT programs and had the dual objectives of alleviating current poverty and increasing children's human capital so as to reduce the transmission of poverty. The largest component of Progresa was the cash transfer paid to mothers conditional on their school-age children attending school on a regular basis. These grants

4. Outside of CCTs, this insight is not new. For example, in developed countries, college students may be beneficiaries of the tax system while in college (when their consumption is low), but will likely be net-payers into the tax system over their lifetime.

were substantial: for example, a mother received 255 pesos per month, or 40% of the typical male laborer’s monthly earnings in these rural communities, if her ninth grade daughter was enrolled in school. Using this setting, we first estimate income and price effects. To identify price effects, we exploit the fact that the introduction of the CCTs was randomized at the locality level: 63% of the localities received transfers immediately, while the other 37% started two years later. To identify income effects, we use variation in transfers to younger siblings below the age of 12 years, as enrollment below the age of 12 is almost 100%, implying the conditions of these transfers were non-binding.⁵ In other words, transfers to younger siblings can be viewed as unconditional transfers to the household. Using detailed panel data from 1997-1999, we estimate income and price effects for secondary school aged children (children aged 12-15 years). We find substantial income effects, with average income effects around one-third as large as average price effects. These estimates imply a relatively high degree of curvature in the utility of consumption, with an implied coefficient of relative risk aversion of 1.05.⁶

Using these estimates, we proceed to evaluate the size of the targeting benefit relative to the exclusion cost for households in Progresa villages with secondary school age children. Under the observed Progresa transfers, we find that this ratio is substantial: the targeting benefit is 88% as large as the exclusion cost. We then calculate the share of the Progresa budget that should be allocated to a CCT over a UCT based solely on targeting grounds (i.e., we calculate the share that equates the targeting benefit to the exclusion cost). We find that 41% of the Progresa budget should go to a CCT, implying that the targeting benefit is a quantitatively important benefit of CCTs in this setting (in comparison, under the observed transfers, 65% of the budget goes towards a CCT).^{7,8} There are three factors driving this result. First, the opportunity cost of sending a teenage child to school in these villages is high. We estimate that 12-15 year-old children earn around 72% as much as their fathers. Second, parents sending their teenage children to school do not earn substantially more than parents not sending their teenage children to school, i.e., parental income inequality among eligible households is low. Third, we estimate substantial curvature in the utility of consumption. Taken together, these three factors imply that schooling households have higher marginal utility of consumption, on average, than non-schooling households. Hence, allocating some of the budget towards a CCT better targets transfers to households who place a higher value on receiving an extra dollar today.

5. We control for the direct effects that sibling composition has on enrollment.

6. I.e., we estimate the coefficient γ in the utility function $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$ to be 1.05.

7. I.e., of Progresa spending to households with a secondary school age child, 65% is spent on CCTs to these children, whereas 35% is spent on (effectively) unconditional transfers to these children’s younger siblings.

8. Moreover, if we restrict the planner’s choice set to either a pure UCT or a pure CCT, we find that social welfare is higher under a pure CCT.

The rest of the paper proceeds as follows: Section II discusses our contribution to the literature, Section III sets-up our theoretical framework and derives our main theoretical results, Section IV develops our sufficient statistics approach for estimating the size of the targeting benefit relative to the exclusion cost, Section V estimates the income and price effects for Progresa, Section VI calculates the size of the targeting benefit relative to the exclusion cost in the context of Progresa and then estimates the optimal share of the Progresa budget to be allocated to a CCT based on the targeting arguments alone. Finally, Section VII concludes.

II CONTRIBUTION TO THE LITERATURE

Our paper contributes to the literature in several ways. First we contribute to the literature on targeting of cash transfers in developing countries. While there exists a large literature on the various targeting strategies and how successful these strategies have been in practice (see, for example, Coady et al, 2004; Ravallion, 2009; Alatas et al, 2012; Alatas et al, 2016; Banerjee et al, 2018; Hanna and Olken, 2018), to the best of our knowledge, no one has investigated the targeting benefit associated with imposing conditions on school attendance. In doing so, we highlight a new welfare benefit of CCTs relative to UCTs, thus also contributing to the literature investigating the costs vs. benefits of imposing conditions on cash transfers (see, for example, Fiszbein and Schady, 2009; Baird et al, 2011; Baird et al, 2013; Benhassine et al, 2015).

Second, we contribute to the large literature on optimal redistribution. To the best of our knowledge, we are the first to analyze the CCT vs. UCT problem within an optimal redistribution framework. By doing so, we show that CCTs can outperform UCTs at transferring funds to the highest marginal utility households in society. Our result is in some ways an extension of the “tagging” literature that started with Akerlof (1978). Akerlof showed that it can be welfare improving to condition tax schedules on observable, immutable characteristics if such characteristics are negatively correlated with endowed earnings. We show that conditioning cash transfers on school attendance can be welfare improving because schooling households may have higher marginal utility of consumption due to forgone child income. However, we refer to this as the “targeting” benefit of CCTs rather than a “tagging” benefit because schooling is a choice variable as opposed to an immutable characteristic.

Third, we contribute to the literature on estimating the curvature of utility (see, for example, Mehra and Prescott, 1985; Barsky et al, 1997; Cohen and Einav, 2005; Kaplow, 2005; Chetty, 2006; Layard et al, 2008). We do so by extending the novel insights of Chetty (2006), in which he relates labor supply elasticities to the curvature of utility, to a schooling decision model. Specifically, we show that the income effect schedule of a UCT

and the price effect schedule of a CCT pin down the curvature of utility. To the best of our knowledge, no one has measured curvature using schooling elasticities. This is a useful exercise given the large variation in curvature estimates across different market settings. Interestingly, by showing the income and price effects are directly related to curvature of utility, we potentially invalidate the monotonic relationship the cash transfer literature places on these elasticities when comparing CCTs to UCTs. The current view is that if the income effect is similar in magnitude to the price effect, one should offer a UCT over a CCT given the effects on school enrollment are similar, while a UCT has the additional advantage of transferring to those who find it too costly to comply (see, for example, Baird et al, 2011; de Brauw and Hoddinott, 2011; Baird et al, 2013). However, we show that the larger the income effect, the greater the concavity of utility, potentially leading to a larger targeting benefit of CCTs.

Fourth, we contribute to the large literature on estimating income and price effects for cash transfer programs (see, for example, Schultz, 2004; de Janvry et al, 2006; Filmer and Schady, 2011; Edmonds, 2006; Edmonds and Schady, 2012). Importantly, our novel identification strategy allows us to estimate these elasticities in the same setting using a reduced-form specification.⁹ While Todd and Wolpin (2006) and de Brauw and Hoddinott (2011) estimate both behavioral responses in the setting of Progresa, the former uses a model-based simulation exercise to do so, while the latter can only pin down income effects relative to price effects.

Finally, we add to the literature that investigates alternative reforms of the Progresa program. Todd and Wolpin (2006) and Attanasio et al (2011) both examine the effects of counter-factual policies on school enrollment. Our paper differs in that we investigate the optimal allocation of the budget towards a CCT vs. UCT so as to best alleviate current poverty. In doing so we highlight the role conditions can play in improving the targeting of cash transfers.

III THEORETICAL FRAMEWORK

III.A Parents' Problem

In this section we develop a framework to highlight the targeting benefit of imposing conditions to send children to school on cash transfers. In this framework, we abstract away from conventional motives to impose conditions (e.g., parental under-investment

9. Other papers that have been able to estimate behavioral responses to CCTs and UCTs in the same setting include Bourguignon et al (2003) who use a model-based simulation exercise to predict behavioral responses in Brazil; Schady and Araujo (2008) who use implementation glitches in Ecuador; Baird et al (2011) and Benhassine et al (2015) who conduct randomized experiments in Malawi and Morocco, respectively; and Akresh et al (2013) who conduct a randomized experiment in Burkina Faso.

motives and/or political economy motives) so as to focus solely on the targeting consequences of CCTs relative to UCTs. We consider a two-generation world where, in the first generation, all households consist of a parent and child, and parents must decide whether to send their child to school or to work. In the second generation, children are adults and earn incomes that depend on whether they went to school as a child. The parents' objective is to maximize household utility over the two generations. We assume, realistically, that parents cannot borrow against the future generation's income (i.e., their child's future earnings). We will show in subsection III.D that this assumption is crucial to warrant any of the budget being allocated towards a CCT based on targeting arguments alone.

We introduce two forms of heterogeneity. The first is with respect to parental income which we initially assume is endowed (we relax this assumption in subsection III.D). The second form of heterogeneity governs the returns to schooling. While there are many potential ways to interpret this heterogeneity, we prefer to think of it as idiosyncratic child ability and/or parents' altruistic preferences over the future success of their children. We introduce this second form of heterogeneity for realism as it is likely an important form of heterogeneity affecting schooling decisions. Without a second form of heterogeneity, we would be able to observe a cut-off parental income level such that all parents below this cut-off would not send to school and all parents above this cut-off would send to school. However, in reality we observe shares of parents at each income level sending to school, suggesting there are additional relevant forms of heterogeneity. The parents' problem can be written as follows:

$$\begin{aligned} \max_{s \in \{0,1\}} \quad & u(c) + v(\mu, s) \\ \text{s.t.} \quad & c = y + y_c(1 - s) \end{aligned}$$

where $u(c)$ denotes household utility over consumption c in the first generation (where $u_c > 0, u_{cc} \leq 0$), y denotes heterogeneous parental income, s denotes whether parents decide to send to school or not, and y_c denotes child income (assumed constant for simplicity). Children only earn y_c if they do not go to school (i.e., if $s = 0$). Thus, schooling is a lumpy investment decision for parents. $v(\mu, s)$ represents household utility in the second generation and is a function of whether the child went to school in the first generation as well as the heterogeneous return to schooling μ . For example, if we view μ as idiosyncratic child ability then we could write $v(\mu, s) = w(y_2(\mu, s))$, where $w(y_2)$ denotes the utility children receive in the second generation from earning income y_2 , where their income is a function of their ability μ and whether they went to school as a child. Alternatively, if we view μ as heterogeneity in altruistic preferences, we could

write $v(\mu, s) = w(y_2(s))(1 + \mu)$, i.e., the child receives direct utility $w(y_2(s))$ and parents receive indirect utility from their child's happiness $\mu w(y_2(s))$. For the remainder of the paper, we will refer to μ as child ability.

We make the following functional form assumptions on $v(\mu, s)$: first, v is increasing in both child ability and schooling, $v_1(\mu, s) > 0$ and $v(\mu, 1) > v(\mu, 0)$. Second, those with higher ability benefit more from going to school, $\frac{\partial v(\mu, 1)}{\partial \mu} > \frac{\partial v(\mu, 0)}{\partial \mu}$. This allows us to derive the following lemma:

Lemma 1. *Schooling, s^* , is weakly increasing in ability type μ .*

Proof. A household will send to school iff $u(y) + v(\mu, 1) \geq u(y + y_c) + v(\mu, 0)$. Because we assume $v_1(\mu, 1) > v_1(\mu, 0)$, if some household with type μ_1 sends to school, any household with type $\mu_2 > \mu_1$ will send to school (holding parent and child income constant). Alternatively, if some household with child ability μ_3 does not send to school, any household with child ability $\mu_4 < \mu_3$ will also not send to school (holding parent and child income constant). \square

This implies that for a given parental income y , there exists a cut-off ability $\tilde{\mu}(y)$ such that a household with $\mu > \tilde{\mu}(y)$ will send to school and a household with $\mu < \tilde{\mu}(y)$ will not. In addition, we make the following assumption:

Assumption 1. $F(\mu|y)$ FOSD $F(\mu|y') \forall y > y'$.

Loosely speaking, Assumption 1 implies child ability and parental income are weakly positively correlated. We can then derive the following proposition:

Lemma 2. *If Assumption 1 holds, $\mathbb{E}[y|s^* = 0] \leq \mathbb{E}[y|s^* = 1]$.*

Proof. See Appendix A.1. \square

Assumption 1 ensures that parents who send their child to school earn more on average than parents who do not send their child to school (i.e., Assumption 1 rules out cases in which parental income and child ability are so negatively correlated such that the parents who send to school earn less on average than the parents who do not send to school). We make Assumption 1 not only because it is realistic, but also because it highlights that the existence of the targeting benefit does not rely on a negative correlation between child ability and parental income.

III.B Social Planner Problem

We consider a utilitarian social planner with a fixed budget to distribute to eligible households in the first generation. For simplicity, we assume eligibility has been predetermined (say via proxy-means testing and/or geographical targeting). In reality, however, the set of eligible households is also a choice variable for the planner. Given that we abstract from this decision, one could view our planner’s problem as an interior problem; i.e., once the planner has decided on the pool of eligible households, how should she allocate a budget to these households? It is worth noting that the majority of CCT and UCT programs are targeted to a poorer subset of the population as opposed to being offered universally, and, as suggested by Hanna and Olken (2018), targeted transfers result in substantially higher welfare gains relative to universal programs.¹⁰

Next, we realistically assume the planner cannot condition transfers on parental income given the difficulty in verifying incomes in developing countries (especially among poorer households who predominately work in the informal sector); she also cannot observe child abilities. However, the planner can observe schooling decisions. Thus, the planner has two tools at her disposal: a constant unconditional cash transfer that all eligible households receive, and a constant conditional cash transfer that only the eligible households who send their child to school receive.¹¹ Finally, we assume that the planner places the same value on the return to schooling as parents (i.e., parents do not under-value the return to schooling, nor are there positive externalities associated with increased enrollment). We make this assumption so as to purely focus on the targeting consequences of offering a CCT over a UCT, and, moreover, to investigate whether spending on a CCT can ever be justified on targeting grounds. However, as both a theoretical and empirical extension, we will consider a scenario where parents do not invest in their children’s schooling at the socially optimal level. In this extension, CCTs have an added benefit over UCTs in that they induce more households to send their children to school, thus helping correct for this underinvestment in schooling (see Appendix A.5 for the theoretical extension and Section VI.E for the empirical extension). With all this in mind, we can write the social planner’s problem as follows:

10. As a theoretical exercise, we investigate how increasing parental income inequality among the set of eligible households affects the planner’s problem, thus allowing us to conjecture how the targeting benefit of CCTs is affected by moving to a universal transfer scheme.

11. In Appendix A.4 we consider an extension where the planner can verify incomes, and thus can offer conditional and unconditional grants that vary with income: $t_u(y)$ and $t_c(y)$. In this situation, it will always be optimal for the planner to spend some of her budget of a CCT based on solely on targeting arguments. This is because, conditional on a parental income level, y , the households who send to school will have lower consumption, thus making it beneficial to direct more money to them.

$$\begin{aligned}
W &= \max_{t_u, t_c} \int_Y \int_M u(c^*) + v(\mu, s^*) f(\mu, y) d\mu dy \\
&\text{s.t. } t_u + t_c \int_Y \int_M \mathbb{1}(s^* = 1) f(\mu, y) d\mu dy \leq R \\
c^* &= y + (1 - s^*)y_c + t_u + t_c s^* \\
t_u &\geq 0, t_c \geq 0
\end{aligned}$$

where Y denotes the set of parental incomes for households that are eligible to receive the transfers, M denotes the set of household specific child abilities, R denotes the planner's budget, and t_u and t_c denote the unconditional and conditional cash transfers, respectively. Substituting in the budget constraint and noting that households with income y and $\mu < \tilde{\mu}(y, t_c, t_u)$ do not send to school, the planner's first order condition with respect to t_c can be expressed as follows:

$$\begin{aligned}
&\left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y \int_{\tilde{\mu}(y, t_c, t_u)} u_c(y + t_u(t_c) + t_c) f(\mu, y) d\mu dy + \\
&\frac{\partial t_u}{\partial t_c} \int_Y \int^{\tilde{\mu}(y, t_c, t_u)} u_c(y + y_c + t_u(t_c)) f(\mu, y) d\mu dy - \\
&\int_Y \left(\frac{\partial \tilde{\mu}}{\partial t_c} + \frac{\partial \tilde{\mu}}{\partial t_u} \frac{\partial t_u}{\partial t_c}\right) (u(y + t_u(t_c) + t_c) + v(\tilde{\mu}, 1) - u(y + y_c + t_u(t_c)) - v(\tilde{\mu}, 0)) f(\tilde{\mu}|y) f(y) dy = 0
\end{aligned}$$

where $\tilde{\mu}(y, t_c, t_u)$ is implicitly defined by the following indifference condition:

$$u(y + t_u + t_c) + v(\tilde{\mu}, 1) = u(y + y_c + t_u) + v(\tilde{\mu}, 0) \quad (1)$$

and where $\frac{\partial t_u}{\partial t_c}$ is implicitly defined as follows:

$$\frac{\partial t_u}{\partial t_c} = - \underbrace{\int_Y \int_{\tilde{\mu}} f(\mu, y) d\mu dy}_{\text{mechanical effect}} + t_c \underbrace{\int_Y \frac{\partial \tilde{\mu}}{\partial t_c} + \frac{\partial \tilde{\mu}}{\partial t_u} \frac{\partial t_u}{\partial t_c} f(\tilde{\mu}|y) f(y) dy}_{\text{behavioral effect}} < 0 \quad (2)$$

Noting that the change in lifetime utility experienced by the households who are induced to send to school is zero (by the envelope condition), the first order condition w.r.t. t_c simplifies to:

$$\begin{aligned}
& \underbrace{\left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y \int_{\tilde{\mu}(y,t_c,t_u)} u_c(y + t_u(t_c) + t_c) f(\mu, y) d\mu dy +}_{\text{Targeting Benefit (TB)} > 0} \\
& \underbrace{\frac{\partial t_u}{\partial t_c} \int_Y \int^{\tilde{\mu}(y,t_c,t_u)} u_c(y + y_c + t_u(t_c)) f(\mu, y) d\mu dy}_{\text{- Exclusion Cost (EC)} < 0} = 0
\end{aligned} \tag{3}$$

The first term in Equation (3), what we denote as the targeting benefit, represents the net social welfare gain the schooling households experience when we increase t_c . Specifically, $\left(1 + \frac{\partial t_u}{\partial t_c}\right)$ represents the net increase in transfers the schooling households receive when we increase t_c by one dollar: they receive a dollar more in t_c but lose $\frac{\partial t_u}{\partial t_c}$ dollars in t_u (as the planner must satisfy her budget constraint). We multiply this term by the aggregate marginal welfare gain each schooling household experiences when we increase their net transfer. The second term in Equation (3), denoted the exclusion cost, represents the social welfare loss from decreasing t_u by $\frac{\partial t_u}{\partial t_c}$ dollars for those not sending to school, i.e., those excluded from the program. Lastly, the expression for $\frac{\partial t_u}{\partial t_c}$, Equation (2), contains two components: the mechanical effect and the behavioral effect. The mechanical effect captures the change in the UCT required to satisfy the budget constraint when we increase the CCT, holding household schooling decisions constant. The behavioral effect captures the change in the UCT required to satisfy the budget constraint due to additional households now sending their child to school and thus receiving the CCT. Both the mechanical and behavioral effects are negative, hence $\frac{\partial t_u}{\partial t_c} < 0$.¹²

In order for the planner to allocate some of the budget towards a CCT, the targeting benefit must outweigh the exclusion cost at $t_c = 0, t_u = R$. The current consensus in the literature is that this could never occur because the schooling households have higher parental incomes (on average). However, because the schooling households forgo a discrete amount of child income, y_c , we will be able to show that, at least in theory, the targeting benefit can dominate. In these situations, spending some (or even all) of the budget on a CCT will actually improve the targeting of transfers to low consumption (high marginal utility of consumption) households.

III.C Results

We proceed to derive our key theoretical results. First we will show that if utility in the first generation is concave in consumption ($u_{cc} < 0$), it may be optimal for the planner to spend some or even all of her budget on a CCT. To understand why, consider a

12. Appendix A.3 shows that $\frac{\partial t_u}{\partial t_c} < 0$.

set of households with constant child ability, $\bar{\mu}$, but with heterogeneous parental incomes. There exists a parental income cut-off level \tilde{y} s.t. parents with incomes $y \geq \tilde{y}$ will send their child to school, and parents with incomes $y < \tilde{y}$ will not send their child to school, where \tilde{y} is implicitly defined as follows: $u(\tilde{y} + t_u + t_c) + v(\bar{\mu}, 1) = u(\tilde{y} + y_c + t_u) + v(\bar{\mu}, 0)$. As such, the household at \tilde{y} experiences a jump up in marginal utility of consumption due to the discrete loss in child income they experience from sending their child to school. This jump is illustrated in Figure II below.

In this world, by offering a pure CCT (i.e., allocating all of the budget towards t_c) over a pure UCT (i.e., allocating all of the budget towards t_u), the planner can transfer more to the schooling households (who potentially have higher marginal utility of consumption, on average). However, this comes at the cost of missing out on transferring to those households who find it too costly to send their child to school (Figure III illustrates the effect of a pure UCT and a pure CCT on marginal utility of consumption, respectively).¹³ Depending on the size of jump up in marginal utility (which in turn will depend on the cost of schooling, y_c , and the concavity of utility) and the distribution of parental incomes, it may be the case that the schooling households have, on average, higher marginal utilities of consumption, thus making it beneficial to target transfers to them. This leads to our main proposition.

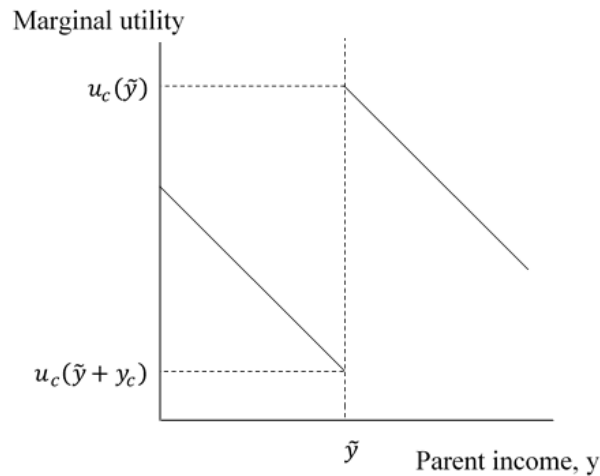


FIGURE II: DISCONTINUITY IN MARGINAL UTILITY OF CONSUMPTION (CONSTANT μ)

13. Moreover, a pure CCT induces a larger behavioral response (i.e., a greater increase in enrollment) compared to a pure UCT (illustrated by $\tilde{y} - \tilde{y}_{cct} > \tilde{y} - \tilde{y}_{uct}$ in Figure III). Inducing more marginal households to send to school is costly in our model given that it takes resources to make them send to school, but the change in their utility from doing so is negligible by the envelope condition.

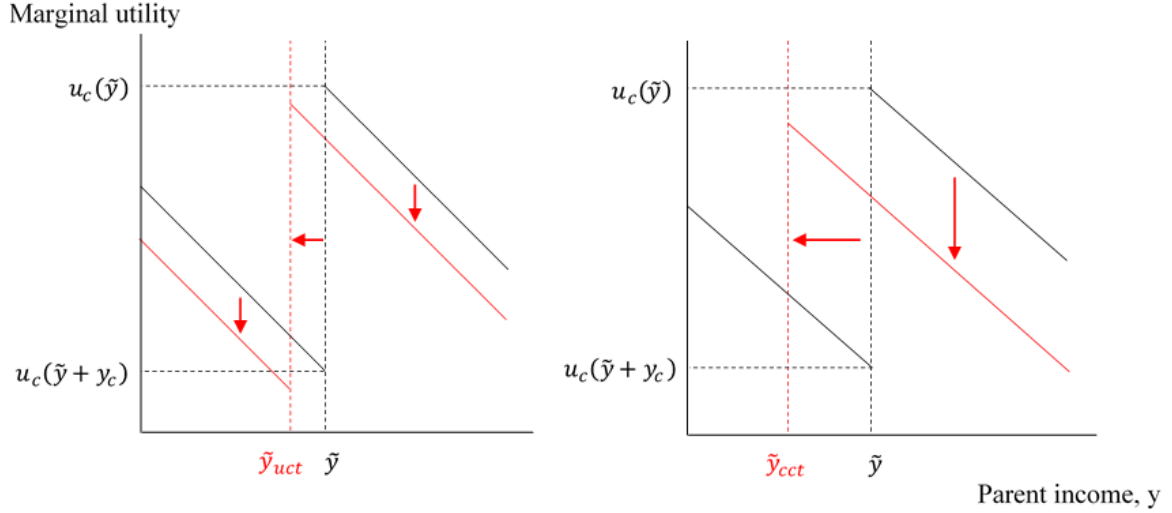


FIGURE III: PURE UCT (LEFT) VS PURE CCT (RIGHT)

Proposition 3. *If utility is concave in consumption, a pure UCT may not be optimal. In particular, if under a pure UCT, average marginal utility of consumption is greater for schooling relative to non-schooling households, a pure UCT is not optimal (i.e., $t_c^* > 0$).*

Proof. See Appendix A.6. □

Proposition 3 implies that based on targeting arguments alone, allocating some (or potentially all) of the budget to a CCT over a UCT can be optimal. We now proceed to investigate what factors affect the size of the targeting benefit relative to the exclusion cost, and, in turn, affect the optimal spending on a CCT.

First, we consider what happens if utility is linear in consumption. In this setting, one might initially think the planner is indifferent between a CCT and UCT given marginal utilities across households are equal. However, the planner must also take into account that imposing conditions distorts the enrollment decisions of marginal parents. This distortion is costly in our framework given we abstract from any benefits associated with increased enrollment. Thus, based on targeting arguments alone, a pure UCT will be optimal if utility is linear in consumption.

Proposition 4. *If utility is linear in consumption, a pure UCT is optimal.*

Proof. See Appendix A.7. □

Next, we consider how the targeting benefit is impacted by the distribution of parent income, $f(y)$. We show that as parents sending their children to school become poorer (in a stochastic dominance sense), the targeting benefit grows relative to the exclusion

cost, so that the budget share allocated to the CCT increases.

Proposition 5. *Suppose $t_c^* > 0$ under the CDF $F^{(1)}(y, \mu)$. Consider $F^{(2)}(y, \mu)$ s.t. $F^{(1)}(y|s^* = 1)$ (weakly) FOSD $F^{(2)}(y|s^* = 1)$ and $F^{(2)}(y|s^* = 0)$ (weakly) FOSD $F^{(1)}(y|s^* = 0)$, with one holding strictly. Further suppose that average behavioral responses and the share of children enrolled are the same under both $F^{(1)}$ and $F^{(2)}$ at t_c^* . Then under $F^{(2)}(y, \mu)$, the optimal local reform starting from t_c^* is to increase t_c .*

Proof. See Appendix A.8. □

Proposition 5 states as we shift the distribution of parent incomes for the schooling households to the left (i.e., we make the schooling households poorer), the targeting benefit becomes more important, making it optimal (locally) to increase the CCT (and, in turn, reduce the UCT). Shifting the distribution of parent incomes for the schooling households is useful to think about, as it is closely related to the degree of parental income inequality. Because schooling is a normal good (i.e., $\frac{\partial s^*}{\partial y} \geq 0$), if we reduce the parental incomes of the schooling households, loosely speaking, we lower the degree of parental income inequality. Conversely, as we increase the parental incomes of the schooling households, we increase the degree of parental income inequality. Then, by Proposition 5, as we increase (decrease) parental income inequality, we would expect the targeting benefit to fall (rise) relative to the exclusion cost, and, hence, the share of the budget allocated towards the CCT to also fall (rise). From this we conjecture that moving from our current set-up where the planner allocates money to a predetermined set of poor households (note, this is the typical - most CCT and UCT programs are offered only to a poorer subset of the population), to a scenario where the planner allocates money to the entire population (i.e., a universal CCT or UCT), would work to severely reduce the targeting benefit relative to the exclusion cost.

Finally, we consider how changing the size of child incomes affects the optimal CCT. As y_c increases (holding all else equal), the discrete loss in household consumption that results from sending a child to school increases, thus increasing the targeting benefit. However, making a formal statement as to how the optimal CCT changes with y_c proves difficult as changing y_c changes the set of enrolled households. Increasing y_c will result in fewer households sending their child to school, with this reduction likely coming from poorer (lower parental income) households. This will work in favor of reducing the targeting benefit, making a comparative static in y_c unfeasible.

III.D Extensions

We now consider six extensions to our baseline model and investigate if these extensions can overturn our main result, Proposition 3, that it can be optimal to spend on a CCT over a UCT based solely on targeting arguments. Our first extension allows parents to freely borrow across generations (see Appendix A.9 for model set-up and results). While we do not think this extension is realistic, we believe it is worthwhile to consider as it highlights that the lack of inter-generational borrowing markets is crucial for the targeting benefit to be relevant (i.e., for Proposition 3 to hold). With perfect borrowing, households perfectly smooth their consumption across generations. Schooling households will therefore have higher consumption in each generation given that they are higher lifetime utility households (and, hence, higher lifetime consumption households). Thus, with perfect borrowing, it cannot be optimal to target transfers to schooling households.

While it's unrealistic to think parents can borrow against their children's future income, it may be realistic (depending on the context) to assume parents have access to financial markets and can borrow against their own future income (within the first generation). We augment our baseline model to include multiple periods within the first generation and allow parents to borrow freely across these periods (but not across generations). In the first period, parents choose whether or not to send their child to school and how much to borrow against their own income in the second period of the first generation (see Appendix A.10 for model set-up and results).¹⁴ While being able to borrow allows parents to smooth consumption across periods within the first generation, they still cannot smooth consumption across generations. Hence, sending to school still induces a discrete loss of household consumption (although this loss is now smaller as parents can spread the loss over two periods). Consequently, we are able to show it can still be optimal to allocate some of the budget towards a CCT.

Related to the above extension, we could augment our model to incorporate additional investments parents can make such as agricultural investments or investments in child health. Assuming these alternative investments are continuous, we will still be able to show that sending a child to school results in a discrete loss in consumption today, and thus show that it can be optimal to allocate some of the budget to a CCT. The intuition is as follows: consider a household deciding how much to invest in the alternative investment and whether to send their child to school. If they choose to not send their child to school, they'll invest more in the alternative investment due to the boost in consumption they experience from child income. However, they will not do so in a one-to-one fashion, i.e., they will not increase their alternative investment one-for-one with their increase

14. Note, this order can easily be reversed so that parents can save in the first period, and send their child to school in the second period.

in household income.¹⁵ Thus, sending a child to school will result in a discrete loss of consumption.

Third, we incorporate parental labor supply decisions into our model so that parental income is no longer endowed. Rather, parents are endowed with a heterogeneous productivity level, and choose how much labor to supply during the first generation along with whether or not to send their child to school (see Appendix A.11 for model set-up and results). We show that parents partially offset the cost of schooling by increasing their labor supply. However, they do not do so in a one-to-one fashion, i.e., they do not fully offset the loss in child income. Hence, sending a child to school still results in a discrete loss of consumption; thus, we are still able to show it can be optimal for the planner to allocate some of her budget towards a CCT.

Fourth, we consider an extension where the planner has money to redistribute over multiple generations. We do so to highlight that (1) the targeting trade-off the planner faces in the first generation is unaffected by this extension (i.e., the targeting benefit relative to the exclusion cost still appears in the planner's first-order-condition), and (2) that a dynamic transfer scheme mitigates the concern that we are on net transferring more to higher lifetime utility households. First, because households cannot move resources across generations, when choosing how to spend money in the first generation, the planner is still concerned about directing money to those households with the lowest consumption in the first generation. Hence, the targeting trade-off in the first generation is unaffected by this extension. Second, assuming returns to schooling are sufficiently high, children who went to school in the first generation do not meet eligibility criteria for transfers when they are adults in the second generation. Thus, on net, it can easily be the case that the households that do not send to school in the first generation receive more over the lifetime (see Appendix A.12 for model set-up and results). In essence, the planner is alleviating the inefficiencies caused by the lack of intergenerational borrowing as she is giving money to schooling households in the first generation (i.e., when they are at their poorest), and giving money to non-schooling households in the second generation (i.e., when they are at their poorest).

Fifth, we investigate how changing the planner's social welfare function affects the size of the targeting benefit relative to the exclusion cost. We show that Proposition 3 holds for all continuous social welfare functions (see Appendix A.13). Essentially, with all continuous social welfare functions, the planner will always place some value on redistributing to those households who value a dollar today the most, who may in turn be the schooling households. Note, however, if we consider the non-continuous Rawlsian social welfare function where the planner only cares about redistributing to the lowest lifetime utility household, it will never be optimal to allocate any of the budget towards

15. We are implicitly assuming diminishing returns to the alternative investment.

a CCT, as the lowest lifetime utility household must be a household who does not send to school.¹⁶

Finally, we illustrate how our model can be extended to include multiple, binary schooling decisions over a set number of T years (see Appendix A.14 for model set-up and results). The planner has a fixed budget which she allocates between a constant, annual CCT and a constant, annual UCT for the next T years so as to maximize the total lifetime utility of parents. While the intuition of our simpler model is unchanged in this more complicated setting (e.g., within any given year, sending to school still results in a discrete loss of household consumption), such an extension is useful to show how to evaluate the planner's FOC in a more realistic setting. We show that the planner's FOC is simply a sum of the annual targeting benefits and annual exclusion costs over the T years.

IV SUFFICIENT STATISTICS FOR THE TARGETING TRADE-OFF

In this section we develop a way to calculate the size of the targeting benefit relative to the exclusion cost from empirically observable objects (sufficient statistics). This method will allow us to determine the optimal CCT/UCT mix based solely on targeting grounds and, therefore, will be useful in determining whether the targeting benefit is a quantitatively important advantage of CCTs. To begin, we re-write the targeting benefit (TB), relative to the exclusion cost (EC) as follows:

$$\frac{TB}{EC} = - \frac{\left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y u_c(y + t_u + t_c) S(X) f(y) dy}{\frac{\partial t_u}{\partial t_c} \int_Y u_c(y + y_c + t_u) (1 - S(X)) f(y) dy} \quad (4)$$

where $X = \{y, t_c, t_u\}$ and $S(X) = \int_{\tilde{\mu}(X)} f(\mu, y) d\mu$ denotes the proportion of households with parent income y sending their child to school under transfer schedule (t_c, t_u) (recall that $\tilde{\mu}(X)$ denotes the household with parental income y just indifferent between sending their child to school under transfer schedule (t_c, t_u) ; see Equation (1)). It can be seen from Equation (4) that in order to determine the relative size of the targeting benefit one needs to know the function $u_c(c)$, i.e., one needs to know the curvature of utility. This is intuitive, as this ratio essentially captures the extent to which the schooling households

16. Note, we are making two relatively innocuous assumptions: (1) the support of $f(y, \mu)$ is a lattice so that $f(y_{min}, \mu_{min}) \neq 0$, and (2) there exist some households that do not send to school. These two assumptions guarantee that the lowest lifetime utility household is a household not sending their child to school.

value receiving an extra dollar relative to the non-schooling households. Thus, we first derive a method to pin down the curvature of utility from observable quantities.

Following a similar procedure to Chetty (2006), we will show how two behavioral elasticities, the income effect of a UCT and the price effect of a CCT, allow us to pin down the curvature of utility of consumption. To do so, we first define the income and price effects as follows:

$$I(X) = \frac{\partial S(X)}{\partial t_u} = -\frac{\partial \tilde{\mu}(X)}{\partial t_u} f(\tilde{\mu}(X)|y)$$

$$P(X) = \frac{\partial S(X)}{\partial t_c} = -\frac{\partial \tilde{\mu}(X)}{\partial t_c} f(\tilde{\mu}(X)|y)$$

where $I(X)$ measures the increase in the share of parents (with income y) sending their child to school as we increase the unconditional transfer, and $P(X)$ measures the increase in the share of parents (with income y) sending their child to school when we increase the conditional transfer. Implicitly differentiating our indifference condition (Equation (1)) with respect to t_u and t_c , we obtain explicit formulas for the income and price effects:

$$I(X) = \frac{u_c(y + t_u + t_c) - u_c(y + y_c + t_u)}{v_1(\tilde{\mu}, 1) - v_1(\tilde{\mu}, 0)} f(\tilde{\mu}(X)|y) \quad (5)$$

$$P(X) = \frac{u_c(y + t_u + t_c)}{v_1(\tilde{\mu}, 1) - v_1(\tilde{\mu}, 0)} f(\tilde{\mu}(X)|y) \quad (6)$$

Taking the ratio of Equations (5) and (6) we get:

$$\frac{I(X)}{P(X)} = 1 - \frac{u_c(y + y_c + t_u)}{u_c(y + t_u + t_c)} \quad (7)$$

From Equation (7) we can see that the ratio of the income effect to the price effect is proportional to the discontinuity in marginal utility of consumption that results from sending a child to school. Notably, for a given transfer schedule (t_c, t_u) , this relationship holds for all parental income levels y . Therefore, if we had sufficient data, we could estimate I and P for all parental income levels under the observed transfer schedule and recover marginal utility of consumption for all consumption levels. However, in practice, given finite data and limited power, it's often convenient to make a functional form assumption on $u(c)$ and simply use moments of the distribution of $\frac{I(X)}{P(X)}$ to calibrate the parameters of the chosen utility function. For example, if we assume utility of consumption is CRRA with coefficient of constant relative risk aversion γ (i.e., $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$), one only needs to observe the average of the ratio of income to price effects (under the observed transfer schedule) to pin down marginal utility of consumption as γ solves:

$$\mathbb{E}_X \left[\frac{I(X)}{P(X)} \right] = 1 - \mathbb{E}_X \left[\left(\frac{y + y_c + t_u}{y + t_u + t_c} \right)^{-\gamma} \right] \quad (8)$$

Intuitively, it useful to understand why the income effect relative to the price effect allows us to pin down curvature, and, moreover, why a higher value of this ratio implies a greater degree of curvature. A large income effect implies that the share of households sending to school increases sharply as we increase the UCT. This may be the result of marginal utility of consumption decreasing quickly so that the opportunity cost of sending a child to school is decreasing quickly. However, we may also observe a large income effect simply because there is a large mass of indifferent households and/or because the return to schooling is relatively flat w.r.t. child ability (this can be seen by the terms $f(\tilde{\mu}|y)$ and $v_1(\tilde{\mu}, 1) - v_1(\tilde{\mu}, 0)$ in Equation (5), respectively). Fortunately, however, if there is a large mass of indifferent households, or if the return to schooling is flat, the price effect too will be large (i.e., the terms $f(\tilde{\mu}|y)$ and $v_1(\tilde{\mu}, 1) - v_1(\tilde{\mu}, 0)$ enter Equation (6) in the exact same manner). Thus, taking the ratio of the income effect to price effect will allow us to net out the “density component” and the “return to schooling component” from the income effect. Hence, a large ratio of the income effect to the price effect implies that marginal utility of consumption is decreasing quickly, i.e., there is a high degree of curvature in utility of consumption.

Once we know the curvature of utility (i.e., we know the function $u_c(c)$), the only unknown quantity in Equation (4) is $\frac{\partial t_u}{\partial t_c}$, i.e., how the unconditional cash transfer changes as we increase the conditional cash transfer. Fortunately, we can show that this derivative can be expressed in terms of the average price effect, the average income effect, and the average share enrolled. Taking the derivative of the planner’s budget constraint w.r.t. t_c yields the following implicit formula of $\frac{\partial t_u}{\partial t_c}$:

$$\begin{aligned} \frac{\partial t_u}{\partial t_c} &= - \int_Y \int_{\tilde{\mu}} f(\mu, y) d\mu dy + t_c \int_Y \left(\frac{\partial \tilde{\mu}}{\partial t_c} + \frac{\partial \tilde{\mu}}{\partial t_u} \frac{\partial t_u}{\partial t_c} \right) f(\tilde{\mu}|y) f(y) dy \\ &= -\bar{S} - t_c \left(\bar{P} + \bar{I} \frac{\partial t_u}{\partial t_c} \right) \end{aligned}$$

Rearranging yields:

$$\frac{\partial t_u}{\partial t_c} = - \frac{\bar{S} + t_c \bar{P}}{1 + t_c \bar{I}}$$

where $\bar{P} = \int_Y P(y, t_c, t_u) f(y) dy$ etc. This leads us to the following result: under the current transfer schedule, if we can observe the income effect schedule, $I(X)$, the price effect schedule, $P(X)$, the enrollment schedule, $S(X)$, the distribution of parental incomes $f(y)$, and size of forgone child incomes, y_c , we can evaluate the size of the targeting benefit

to the exclusion cost. It is worth mentioning that calculating $\frac{TB}{EC}$ under the observed transfer schedule is equivalent to calculating the optimal local reform (based solely on the targeting trade-off). For example, if we observe a ratio greater than 1, $\frac{TB}{EC} > 1$, this suggests it is optimal to increase the CCT and decrease the UCT from their observed values. Likewise, if this ratio were less than 1, this suggests it is optimal to decrease the CCT and increase the UCT from their observed values.

Finally, the optimal CCT (based on the targeting grounds alone) satisfies $\frac{TB}{EC}(t_c^*, t_u(t_c^*)) = 1$, i.e., the targeting benefit just offsets the exclusion cost. In order to determine the optimal CCT, we need to observe income and price effects at all values of the transfer schedule (t_c, t_u) , not just at the observed schedule.

V ESTIMATING INCOME AND PRICE EFFECTS IN RURAL MEXICO AT THE TIME OF PROGRESA

The remainder of the paper will focus on estimating the importance of the targeting benefit in the context of Progresa, a large CCT program in rural Mexico. In this section, we will focus on estimating the quantities necessary to recover curvature of utility - the income and price effect schedules - using evaluation data from Progresa. We will assume that the utility function is CRRA ($u(c) = \frac{c^{1-\gamma}}{1-\gamma}$) and, therefore, will use the average ratio of income to price effects to pin down the CRRA coefficient γ (see Equation (8) in Section IV above). First, we will briefly discuss the Progresa program. Second, we will describe key facts about child incomes, child labor supply, and evidence that schooling is a discrete, costly investment that results in lower household consumption. Third, we will discuss our identification strategy, focusing on how we will identify income effects from a pure CCT program. Finally, we will estimate the income and price effects and recover the implied curvature of utility.

V.A Progresa: Background

Progresa was one of the first CCT programs with the objective to alleviate current poverty while at the same increase the human capital in children so as to reduce the transmission of poverty (see Parker and Todd (2018) for a thorough review on the literature studying the Progresa program). The program was set up to transfer cash to poor households under the condition that they invest in the human capital of their children. There were two components of Progresa. The first component consisted of nutritional subsidies paid to mothers who register their children for growth and development check-ups, vaccinate their children, and attend courses on hygiene, nutrition and contraception.

The second component consisted of education grants paid to mothers conditional on their school-age children attending school on a regular basis. Specifically, mothers would receive transfers every two months if their children were enrolled in grades 3-9 and attended 85% or more days of school. The education grants were the largest component of the program and will be the part of the program we focus on. For example, the nutritional component (in 1998) was 100 pesos per month (or around \$10 USD) which corresponded to around 8% of the beneficiaries income, while the average education grant per household with children was around 348 pesos per month (Attanasio et al, 2011). See Table I for a summary of the Progresa grant schedules.

Progresa was first targeted at the locality level (targeted localities had both a high marginality index and adequate health and schooling infrastructure). Within each locality, individual households were targeted via proxy-means testing and then split into two groups: poor and non-poor. Poor households were eligible for Progresa transfers while non-poor households were ineligible.¹⁷ The program was initially implemented as follows: 506 of the targeted localities were randomly chosen across 7 states, and of these, 320 were randomly chosen to be offered the transfers immediately, starting May 1998. We refer to these 320 localities as the treated localities. The other 186 localities, the control localities, started the program 2 years later. Thus, eligible households in the treated localities started receiving transfers in May 1998, while eligible households in control localities did not start receiving transfers until May 2000.

Prior to the start of the program, a comprehensive survey was carried out in September 1997, and then supplemented with an additional survey in March 1998. These two surveys constitute the baseline survey. We will use these two surveys along with two surveys taken in November 1998 and November 1999 for our analysis. These surveys cover all households in the evaluation sample (i.e., the 506 localities) and contain extensive household information as well as information on each child including age, gender, education, labor supply, earnings, and school enrollment. Importantly, we know which households are considered eligible for the Progresa grants in both treated and control localities. These surveys cover approximately 24,000 households over 1997, 1998, and 1999.

17. See Skoufias et al (1999) for more information on targeting of localities and households.

TABLE I: PROGRESA TRANSFER AMOUNTS (BIMONTHLY, IN PESOS)

	1998, sem. 1	1998, sem. 2	1999, sem. 1	1999, sem. 2
Type of benefit				
Nutrition support	190	200	230	250
Primary school				
3	130	140	150	160
4	150	160	180	190
5	190	200	230	250
6	260	270	300	330
Junior secondary: girls				
7	400	410	470	500
8	440	470	520	560
9	480	510	570	610
Junior secondary: boys				
7	380	400	440	480
8	400	400	470	500
9	420	440	490	530
Maximum support	1170	1250	1390	1500

Source: Attanasio et al, 2011. Sem. denotes semester.

V.B Key Facts about Child Labor Supply and Income

As will be discussed in our identification strategy below, we estimate income and price effects for children of secondary school age only (children aged 12-15 years, inclusive), as nearly all children below the age of 12 attend school. Before doing so, we first investigate the extent to which children in this age range work, their earnings relative to their parents, and evidence that schooling induces a discrete loss in household consumption. Table II summarizes the labor supply statistics of children aged 12-15 years in 1997 (pre-Progres grants). It can be seen that of those children not in school, 63% of boys report having a job, while 20% of girls report having a job. Conversely, of those children who report they attend school, only 7.5% of boys and 2.9% of girls report having a job. Further, when we look at hours worked for those children with a job, those not attending school work an average of over 40 hours per week, whereas those attending school report an average of less than 15 hours per-week. Thus, it appears that going to school places a major constraint both on labor market participation and on hours worked in the labor market for children.

TABLE II: LABOR SUPPLY STATISTICS OF CHILDREN AND PARENTS, 1997

Parents with children aged 12-15	Mean	Std. Dev.
Has job (father)	0.97	0.18
Has job (mother)	0.14	0.34
Weekly income (father, inc>0)	174.2	129.2
Weekly income (mother, inc>0)	137.2	141.0
Boys aged 12-15	Mean	Std. Dev.
Enrolled	0.69	0.46
Has job (not enrolled)	0.63	0.48
Has job (enrolled)	0.075	0.26
Hours worked per week (not enrolled, has job)	42.4	13.3
Hours worked per week (enrolled, has job)	14.2	5.92
Weekly income (inc > 0)	132.5	90.6
Weekly income (inc > 0, not enrolled)	143.7	90.2
Offered weekly transfer (1998, semester 1)	38.8	17.6
Girls aged 12-15	Mean	Std. Dev.
Enrolled	0.62	0.49
Has job (not enrolled)	0.20	0.40
Has job (enrolled)	0.029	0.17
Hours worked per week (not enrolled, has job)	45.6	16.8
Hours worked per week (enrolled, has job)	14.6	6.85
Weekly income (inc > 0)	139.0	100.4
Weekly income (inc > 0, not enrolled)	146.5	99.6
Offered weekly transfer (1998, semester 1)	42.6	18.7

Note: Sample: all households in 1997 with two parents where one parent reports to be the head and with at least one child aged 12-15 years. Has job takes value 1 if individual reported to work in a paid or unpaid job last week, or reported to have a job but did not work last week. Weekly income is summarized for those reporting positive incomes (we remove those reporting positive incomes outside of the 1st – 99th percentiles). All incomes are inflation adjusted to be in 1998 values (inflation is calculated as the median percentage change in reported annual incomes by state).

One may wonder what the 37% of boys and 80% of girls who report not being in school and not having a job do with their time. As noted in Parker and Skoufias (2001), domestic activities and some other unpaid activities are not included in the survey definition of having a job. Fortunately, however, ProgresA carried out an additional survey, the June 1999 Time-Use Survey, where individuals were asked about how they allocated their time in the previous day, allowing one to investigate the extent that schooling and *all forms* of work are substitutes. Using this survey, Parker and Skoufias (2001) find that conditional on going to school, girls and boys aged 12-17 spend on average 6.3 hours a day at school (prior to grants).¹⁸ Moreover, they find that a) there is no change in hours spent at school after the introduction of ProgresA grants (conditional on going to school), b) there is no

18. In addition, children spend time doing homework and traveling to and from school. Parker and Skoufias (2001) find that almost all children in school report doing homework, and that the average number of hours spent on homework is 1 hour per-day.

change in daily hours of leisure after the introduction of Progresa grants for boys, and c) there is only a very small reduction in daily hours of leisure (≈ 0.26 hours) after the introduction of Progresa grants for girls. This implies that children who were induced to go to school as a result of Progresa grants (almost) perfectly substituted work hours for school hours, thus suggesting that schooling and work (where work includes both market and non-market work) are in fact substitutes.

Moving to earnings, boys (girls) with jobs earn an average of 133 (139) pesos per-week. Moreover, these earnings are substantial relative to parent earnings. Specifically, children aged 12-15 with jobs earn around 80% as much as fathers do, on average.¹⁹ Thus, combining these earning statistics with the fact that schooling directly competes with working, it appears that parents forgo a large, discrete loss in consumption.

V.C Identification Strategy

Our goal is to estimate income and price effects for annual enrollment, i.e., determine how the share of households sending their child to school for a year changes as we change the UCT and the CCT, respectively. In order to recover these derivatives, we first need to discuss where our source of identifying variation in UCTs comes from given Progresa only offered CCTs.²⁰

To obtain variation in UCTs, we will exploit the fact that enrollment of children below the age of 12 is nearly 100%. Figure IV plots the percentage of all children enrolled by age and gender in 1997.²¹ Just over 98% of children are enrolled below the age of 12. This is consistent with findings of previous studies, for example, Attanasio et al (2011) find limited to zero effects of the Progresa grants on enrollment for children under the age of 12 (see Table 2 of their paper), while Todd and Wolpin (2006) state “Because attendance, in the absence of any subsidy, is almost universal through the elementary school ages, subsidizing attendance at the lower grade levels, as under the existing program, is essentially an income transfer”. Thus, it seems reasonable to assume that for children younger than 12 years of age, the conditionalities of the transfer are not binding. We therefore view transfers to children under this age as unconditional transfers

19. It is worth noting that while nearly all fathers report having a job, only 14% of mothers report having a job. We suspect that the majority of mothers spend a substantial amount of their time doing domestic work. Supporting this, Parker and Skoufias (2000), find that women have similar amounts of leisure time as men on average.

20. De Brauw and Hoddinott (2011) exploit the fact that a small number of beneficiaries who received transfers did not receive forms needed to monitor the attendance of their children at school, and hence view transfers to these beneficiaries as unconditional. However, their methodology allows them to only compare enrollment differences between treatment households (i.e., those receiving transfers with forms and those receiving transfers without forms), hence only allowing them to identify income effects relative to price effects.

21. Note, enrollment takes value 1 if it is reported that a child is currently attending school. Thus, one may prefer to view this measure as attendance rather than enrollment.

to the household.

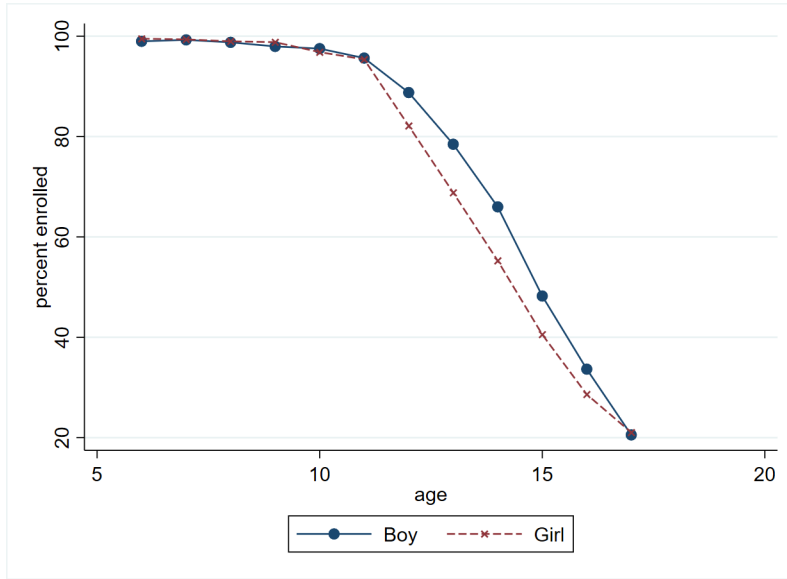


FIGURE IV: ENROLLMENT BY AGE AND GENDER, 1997

Therefore, for children aged 12 and above, we observe both variation in CCTs and UCTs. Variation in CCTs comes from (a) random assignment of CCTs to these children based on whether they live in treatment or control localities, (b) variation in grade, and (c) variation in gender (as the transfer schedules vary with grade and gender of the child - see Table I). Variation in UCTs comes from (a) random assignment of CCTs to younger siblings based on whether they live in treatment or control localities, and (b) variation in sibling composition (e.g., number of siblings, grades of siblings etc.). Thus, controlling for the direct effects that sibling composition, grade, and, gender have on enrollment, we can identify the effects that CCTs and UCTs have on enrollment.

Table III illustrates our identification strategy. For simplicity, assume that income and price effects are constant for all households and consider four groups of 13 year-old children from eligible (poor) households: (1) 13 year-old children living in treatment localities with a 7 year-old sibling, T_7 ; (2) 13 year-old children living in control localities with a 7 year-old sibling, C_7 ; (3) 13 year-old children living in treatment localities with an 8 year-old sibling, T_8 ; (4) 13 year-old children living in control localities with an 8 year-old sibling, C_8 . The reason we distinguish between 7 and 8 year-old siblings is that typically 7 year-old children are enrolled in grade two and, thus, do not receive Progresá education grants, while 8 year-old children are enrolled in grade three and, thus, do receive Progresá education grants. Assume everything else between these four groups is equal. Denote the CCT offered to the 13 year-old children in the treatment localities as t_c and the “UCT” offered to the 8 year-old children in the treatment localities as t_u . The price effect (scaled by the size of the CCT, t_c) will be given by the difference in enrollment between treatment and control 13 year-old children with the 7 year-old sibling before

and after the introduction of the grants, i.e., $Pt_c = (b - a) - (d - c)$. The price plus income effect will be given by the difference in enrollment between treatment and control 13 year-old children with the 8 year-old sibling before and after the introduction of the grants, i.e., $Pt_c + It_u = (f - e) - (h - g)$. Thus, we can identify the income effect through the following difference-in-difference-in-difference: $I = \frac{1}{t_u} ((f - e) - (h - g) - Pt_c)$. The key assumption here is that how parents' enrollment decision for their 13 year-old child changes with the CCT is unaffected by the age of the child's younger sibling (i.e., the price effect is not a function of sibling ages).

TABLE III: IDENTIFICATION STRATEGY: CONSTANT INCOME AND PRICE EFFECTS

	Share Enrolled (before grants)	Share Enrolled (after grants)
7 year-old sibling		
Treatment: T_7	a	b
Control: C_7	c	d
8 year-old sibling		
Treatment: T_8	e	f
Control: C_8	g	h

V.D Constant Price and Income Effects

We proceed to estimate income and price effects assuming they are constant (we relax this assumption later). Our sample consists of an unbalanced panel of children aged 12-15 years (inclusive) living in eligible households (in both treatment and control localities) across the three survey years, 1997, 1998, and 1999. This corresponds roughly to the age range of children attending junior secondary school (grades 7-9). We further restrict our sample to include children in two-parent households where one parent reports to be the household head.²² First we estimate a standard difference-in-difference regression to obtain the overall treatment effect of the Progresa program on enrollment:

$$Enroll_{it} = a_0 + a_1 treat_i + a_2 \mathbb{1}(year > 1997)_t + a_3 (treat_i \times \mathbb{1}(year > 1997)_t) + v_{it} \quad (9)$$

where $Enroll_{it}$ takes value 1 if child i in year t is reported to be currently attending school; and $treat_i$ takes value 1 if child i lives in a treatment locality. Results of Regression (9) are presented in column (1) of Table IV. We find that eligible children aged 12-15 years in treatment localities are 8.83 percentage points more likely to be enrolled in 1998

²². We make this restriction so that child schooling decisions are more likely made by parents than other members of a family. This is important as incentives to educate children may be different for more distant family members. This drops our sample from around 29500 12-15 year-old children to around 21500 12-15 year-old children over the three years.

and 1999 relative to eligible children aged 12-15 years in control localities.²³

TABLE IV: OVERALL TREATMENT EFFECT

	(1)
treat \times $\mathbb{1}(\text{year} > 1997)$	0.0883*** (0.015)
$\mathbb{E}[\text{Enroll} \text{control}, \text{year} > 97]$	0.683
$\mathbb{E}[\text{Enroll} \text{treat}, \text{year} > 97]$	0.767
Observations	21570
R-squared	0.0142

Note: Standard errors are clustered at the locality level and are presented in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: Enroll_{it} . Sample: unbalanced panel of children aged 12-15 years in 1997, 1998, and 1999 from eligible (poor) households with two parents and one parent reports to be the head of the household. $\text{treat} \times \mathbb{1}(\text{year} > 1997)$ takes value one if child i lives in a treated locality and the survey year is after 1997. Regressors not shown: treat and $\mathbb{1}(\text{year} > 1997)$.

Next, we estimate constant income and price effects for children aged 12-15 years. Specifically we estimate the following regression:

$$\text{Enroll}_{it} = b_0 + b_1 t_{c,it} + b_2 t_{u,it} + \beta Z_{it} + \delta_t + \eta_i + e_{it} \quad (10)$$

where $t_{c,it}$ denotes the per-capita CCT offered to child i in year t (in pesos, per-week), and $t_{u,it}$ denotes the per-capita transfers offered to child i 's siblings under the age of 12 in year t (we focus on per-capita transfers as we think what matters to a household is per-capita consumption).²⁴ Specifically, $t_{u,it}$ is calculated as follows:

$$t_{u,it} = \sum_j t_{c,jt} \mathbb{1}(\text{age}_{jt} < 12)$$

where j denotes the j^{th} sibling of child i , and where $t_{c,jt}$ denotes the per-capita conditional transfer offered to sibling j at time t .²⁵ Z_{it} denotes a vector of child i characteristics in year t . In our baseline specification, Z_{it} includes indicators for child i 's highest grade completed, indicators for child i 's age (in years), child i 's age interacted with highest grade completed, and father's weekly, per-capita income (adjusted for inflation).²⁶ Finally δ_t and η_i denote year and child fixed effects, respectively. Note, by including a child

23. When we allow the treatment effects to vary by year, we cannot reject the null that the treatment effects are the same in 1998 and 1999.

24. We measure family size as both parents plus all children aged 18 and below in the household.

25. Note, we adjust transfers for inflation by using the 1998 semester 1 transfer schedule to determine the CCT offered to each child in both 1998 and 1999 (the reason transfers increased over the two years was to compensate for inflation).

26. The majority of mothers in our sample do not report earning an income; hence we only include father's income in the regression.

fixed effect we control for child gender, area characteristics, and other fixed child and family characteristics.

One issue with estimating Equation (10) is that father’s income is likely measured with error. Highlighting this, the correlation between individual reported incomes across 1997-1998 and 1998-1999 is only around 0.4. Thus, estimating Equation (10) via OLS will likely result in attenuation bias. Moreover, there is a potential endogeneity issue with including father’s income as parents may adjust their labor supply in response to schooling decisions (although Parker and Skoufias (2000) find no evidence that parents adjust their labor supply in response to receiving grants, suggesting endogeneity may not be an issue). Therefore, we instrument for father i ’s income in year t with the mean hourly wage in father i ’s locality in year t , excluding father’s i wage. The assumption for the exclusion restrictions are that a) the reporting errors in incomes are independent within localities, and b) an individual’s labor supply decision has negligible effects on wages of others in the locality. Our first stage regression is presented in Appendix B.1.

Results from our OLS and IV estimation of Equation (10) are presented in columns (1) and (2) of Table V, respectively. Price and income effects are given by $P = \frac{\partial \mathbb{E}[Enroll_{it}|t_c, it, t_u, it, Z_{it}]}{\partial t_c, it} = b_1$ and $I = \frac{\partial \mathbb{E}[Enroll_{it}|t_c, it, t_u, it, Z_{it}]}{\partial t_u, it} = b_2$. Focusing on column (2), we find a positive price effect significant at the 1% level and a positive income effect significant at the 10% level.²⁷ Notably, the presence of a positive income effect suggests utility is not linear in consumption. Specifically our constant price effect is equal to 0.0082 and our constant income effect is equal to 0.0046. This means that as the conditional cash transfer increases by 1 peso (per-capita, per-week), enrollment of children aged 12-15 years increases by 0.82 percentage points. Likewise, as the unconditional cash transfer increases by 1 peso (per-capita, per-week), enrollment of children aged 12-15 years increases by 0.46 percentage points. Given the average t_c offered to children in treatment localities is 6.86 pesos (per-capita, per-week), and the average t_u offered to children is 2.90 pesos (per-capita, per-week), enrollment in treatment localities should be 6.96 percentage points higher relative to control localities after 1997 (as $0.82 \times 6.86 + 0.46 \times 2.90 = 6.96$). This estimate is lower than that in column (1) of Table IV (where we find enrollment increases by 8.83 percentage points), perhaps due to the fact we have assumed income and price effects are constant. Taking into account the exchange rate in 1999 (10 pesos \approx 1 USD) and the average family size of 6.61 members, a 1 USD increase in per-week, per-household conditional transfers leads to a 1.24 percentage point increase in enrollment of children aged 12-15, while a 1 USD increase in per-week, per-household unconditional cash transfers leads to a 0.70 percentage point increase in enrollment of children aged 12-15.

It is worth mentioning that our estimates for the price and income effects are robust

27. Given we have substantially less variation in t_u compared to t_c , it is not surprising that our income effect is not as precisely identified as our price effect.

across our OLS and IV specifications. Our point estimate on father’s income is substantially larger under our IV specification, suggesting that our OLS specification suffers from attenuation bias, as suspected. While our coefficient on father’s income is not significant in either specification, its point estimate in our IV specification is of a similar magnitude to the point estimate on t_u . This seems sensible as one may expect parents to react to an increase in unconditional grants and an increase in parental income in a similar manner. Finally, in column (3) of Table V we show that our results are robust to including controls for sibling ages and grades.²⁸

TABLE V: CONSTANT INCOME AND PRICE EFFECTS

	(1) OLS	(2) IV	(3) IV	(4) IV
tc	0.00838*** (0.002)	0.00820*** (0.002)	0.00823*** (0.002)	0.00814*** (0.002)
tu	0.00372* (0.002)	0.00462* (0.003)	0.00487* (0.003)	0.00485* (0.003)
y^f	-0.000236 (0.000)	0.00379 (0.004)	0.00362 (0.004)	0.00362 (0.004)
tc ⁽¹²⁺⁾				0.000296 (0.002)
$\mathbb{E}[\text{Enroll} \text{control}, \text{year} > 97]$	0.683	0.693	0.693	0.693
$\mathbb{E}[\text{Enroll} \text{treat}, \text{year} > 97]$	0.767	0.778	0.778	0.778
$\mathbb{E}[\text{tc} \text{treat}, \text{year} > 97]$	6.897	6.863	6.863	6.863
$\mathbb{E}[\text{tu} \text{treat}, \text{year} > 97]$	2.867	2.896	2.896	2.896
$\mathbb{E}[\text{famsize}]$	6.609	6.609	6.609	6.609
Observations	21321	17432	17432	17432
R-squared	0.153	0.116	0.121	0.121

Note: Standard errors are clustered at the locality level and are presented in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: Enroll_{it} . Sample: unbalanced panel of children aged 12-15 years in 1997-1999 from eligible (poor) households with two parents and one parent reports to be the head of the household. tc denotes the weekly, per-capita transfer offered to child i in year t . tu denotes the weekly, per-capita transfer offered to child i 's siblings under the age of 12. tc⁽¹²⁺⁾ denotes weekly, per-capita transfer offered to child i 's siblings aged 12-15 years. y^f denotes father’s weekly, per-capita income. Regressors not shown: child age dummies, highest grade completed dummies, age interacted with highest grade completed, year dummies, and a child fixed effect. In addition, column (3) includes controls for number of siblings in various age brackets and grade brackets. Finally, in columns (2)-(4) we instrument for father’s income with the average wage in the locality (excluding the wage of the father in question).

Finally, because a child’s siblings aged 12 and above are also offered transfers, we estimate the following regression to investigate the effect of transfers to these siblings on child i 's enrollment:

28. We include the number of siblings in age brackets [0-5], [6-7], [8-9], [10-11], [12-13], [14-15], and [16-18] and the number of siblings with highest grade completed in grade brackets [0-6], [7-9], and ≥ 10 .

$$Enroll_{it} = b_0 + b_1 t_{c,it} + b_2 t_{u,it} + b_3 t_{c,it}^{(12+)} + \beta Z_{it} + \delta_t + \eta_i + e_{it} \quad (11)$$

where

$$t_{c,it}^{(12+)} = \sum_j t_{c,jt} \mathbb{1}(age_{jt} \in [12, 15])$$

Results of Regression (11) are presented in column (4) of Table V. It appears that transfers to siblings aged 12-15 have no significant effect on child i 's enrollment, with this coefficient being highly insignificant and an order of magnitude lower than the coefficients on t_c and t_u . Our estimates of constant income and price effects are unaffected by the inclusion of this variable. It is worth mentioning that the effect transfers to siblings aged 12-15 years have on a child's enrollment is ambiguous given that enrollment decisions for these siblings are not binding. For example, if these siblings were going to go to school regardless of the transfer, then offering them a transfer should be a positive income shock to the household and therefore have a positive effect on child i 's enrollment. However, if these transfers push child i 's siblings into going to school, assuming forgone child income is greater than the transfer amount (which is typically the case in this setting, see Table II), this will be a negative income shock to the household.²⁹

V.E Heterogeneous Income and Price Effects

In the above regressions we assumed income and price effects were constant. However, as shown in Section IV it is likely these effects are heterogeneous with respect to parental income and the size of the transfers. Allowing for heterogeneity in these effects, we estimate the following regression:

$$Enroll_{it} = b_0 + b_1 t_{c,it} + b_2 t_{u,it} + b_3 t_{c,it} y_{it}^f + b_4 t_{u,it} y_{it}^f + b_5 t_{c,it}^2 + b_6 t_{u,it}^2 + b_7 t_{c,it} t_{u,it} + \beta Z_{it} + \delta_t + \eta_i + e_{it} \quad (12)$$

where $t_{c,it}$, $t_{u,it}$ denote the offered, per-capita conditional and unconditional cash transfers, respectively; y_{it}^f denotes father's per-capita income; δ_t denotes year fixed effects; and η_i denotes child fixed effects. As in Equation (10), Z_{it} includes the same controls as in Equation (10). Finally, we again instrument for father's income using the average locality

29. See Ferreira et al (2018) for a more in-depth analysis into how conditional transfers to siblings can have ambiguous effects on a child's enrollment.

wage (excluding the wage of the father in question).³⁰ In this specification, the price and income effects are given by:

$$P(X_{it}) = b_1 + b_3 y_{it}^f + 2b_5 t_{c,it} + b_7 t_{u,it}$$

$$I(X_{it}) = b_2 + b_4 y_{it}^f + 2b_6 t_{u,it} + b_7 t_{c,it}$$

where $X_{it} = \{y_{it}, t_{c,it}, t_{u,it}\}$. Regression results of Equation (12) are presented in Table VI below. We find that both income and price effects are decreasing in parental income, however this interaction term is only significant for price effects (given we have substantially less variation in unconditional grants, it is not surprising that our power to identify heterogeneous effects is limited). Moreover we find evidence to suggest that income effects are decreasing in the size of the unconditional grant indicated by the negative coefficient on t_u^2 . The coefficients on t_c^2 and $t_c t_u$ are both economically and statistically insignificant. Using this specification, we calculate the implied increase in enrollment to be 8.1 percentage points in treatment villages relative to control villages. This is much closer to the observed increase in enrollment of 8.8 percentage points (see column (1) of Table IV).

30. We also instrument for $t_c \times y^f$ and $t_u \times y^f$ with the average locality wage (excluding the wage of the father in question) interacted with t_c and t_u , respectively.

TABLE VI: HETEROGENEOUS INCOME AND PRICE EFFECTS

	(1)
tc	0.0174*** (0.006)
tu	0.0182* (0.011)
tc \times y^f	-0.000272* (0.000)
tu \times y^f	-0.000281 (0.000)
tc ²	0.0000488 (0.000)
tu ²	-0.00118 (0.001)
tc \times tu	0.00000588 (0.001)
y^f	0.00399 (0.004)
Observations	17432
R-squared	0.124

Note: Standard errors are clustered at the locality level and are presented in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $Enroll_{it}$. Sample: unbalanced panel of children aged 12-15 years in 1997-1999 from eligible (poor) households with two parents and one parent reports to be the head of the household. y^f denotes father's weekly per-capita income; tc (tu) denotes the weekly, offered per-capita CCT (UCT). We instrument for father's per-capita income with average locality wage (excluding the wage of the father in question), and we instrument for $t_c \times y^f$ and $t_u \times y^f$ with the offered CCT and UCT interacted with the average locality wage (excluding the wage of the father in question). Regressors not shown: child age dummies, highest grade completed dummies, age interacted with highest grade completed, year dummies, and a child fixed effect.

We proceed to calculate moments of the income and price effect schedules (see Appendix B.2 for explicit calculation of moments). Moments are presented in Table VII below. We find a significant average price effect of 0.010, and an average income effect of 0.0037. We then estimate the average value of the ratio of the income effect to the price effect using the second-order Taylor series expansion given by Equation (13) below.³¹ We find the average of this ratio is equal to 0.21. Finally, in Appendix B.4 we allow for additional heterogeneity in price and income effects with respect to child age and child gender. We do not find statistical evidence to support heterogeneity in these additional factors. Our moment estimates are robust to this more flexible specification.

31. $\mathbb{E} \left[\frac{\hat{I}(X)}{\hat{P}(X)} \right]$ will be a biased estimate of $\mathbb{E} \left[\frac{I(X)}{P(X)} \right]$. Instead we evaluate the second-order Taylor series expansion of $\mathbb{E} \left[\frac{I(X)}{P(X)} \right]$. We adjust second-order moments of income and price effects used in this expansion for bias (see Appendix B.2 for calculation of second-order moments).

$$\mathbb{E} \left[\frac{I(X)}{P(X)} \right] \approx \mathbb{E}[I(X)] \left(\frac{1}{\mathbb{E}[P(X)]} + \frac{1}{\mathbb{E}[P(X)]^3} \text{Var}(P(X)) \right) - \frac{1}{\mathbb{E}[P(X)]^2} \text{Cov}(I(X), P(X)) \quad (13)$$

TABLE VII: MOMENTS OF THE INCOME AND PRICE EFFECT SCHEDULES

$\mathbb{E}[P(X)]$	0.0104*** (0.0030)
$\mathbb{E}[I(X)]$	0.0037 (0.0033)
$\mathbb{E} \left[\frac{I(X)}{P(X)} \right]$	0.21 (0.61)

Note: Moments are calculated for the treated, eligible sample of children aged 12-15 years in 1998 and 1999 (see Appendix B.2 for explicit calculation). Bootstrapped standard errors from 500 bootstrap draws are presented in parentheses.

V.F Coefficient of Constant Relative Risk Aversion

We now use our estimate of $E \left[\frac{I}{P} \right]$ to determine the implied coefficient of relative risk aversion, γ , using the following (discretized) version of Equation (8):

$$\mathbb{E} \left[\frac{I}{P} \right] = 1 - \frac{1}{N} \sum_{it=1}^N \left(\frac{y_{it} + t_{u,it} + y_{c,it}}{y_{it} + t_{u,it} + t_{c,it}} \right)^{-\gamma} \quad (14)$$

Before solving for γ , we need to address two issues. First, we need a measure of (per-capita) parental income, y . While almost all fathers in our sample report earning an income, very few mothers report earning an income. However, as shown in Parker and Skoufias (2000), women work similar hours to men in unpaid domestic activities. Given these activities generate valuable goods and services, they should be valued and included in our measure of parental income. Valuing such services is a very difficult task and would require one to make many assumptions. Instead, we naively assume women generate 80% as much as their husband's, where this ratio is taken from the observed ratio of earnings for the few women that do report an income (see Table II). Thus parental income is given by $y = 1.8y^f$. Given the naivety of such an assumption, we will show robustness of all our results to this assumption by a) varying mother's earnings from 0% to 100% of father's earnings, i.e., $y = (1 + \theta)y^f$ where $\theta \in [0, 1]$, and b) assuming all mother's earn a constant amount equal to 80% of mean father's income, i.e., $y = y^f + 0.8\bar{y}^f$. Our results will be very robust to these alternative specifications.

Second, we need a measure of potential (per-capita) child income, y_c . However, we do not observe potential incomes for those children in school. We therefore predict child income (for all children) using earnings data from children reporting earning an income. One may be concerned with selection bias in such a procedure; however, as shown in Attanasio et al (2011), there is no evidence of selection effects on children’s earnings, likely due to the fact that the jobs children perform in these rural villages are very low-skilled. Thus, our predicted measure of potential child income, y_c , is simply a function of child age, child gender, and area characteristics (see Appendix B.3 for details on how we estimate potential child incomes).

Using data from our sample of eligible, treated 12-15 year-old children for the years 1998 and 1999, and using our estimate of $\mathbb{E} \left[\frac{I}{P} \right]$ from Table VII above, we solve Equation (14) for γ . Our results are presented in Table VIII below. Our curvature estimate is $\gamma = 1.05$. Interestingly, this estimate is very similar to that estimated in Chetty (2006) who finds a value of $\gamma \approx 1$, i.e., $u(c) = \log(c)$ (similarly, Layard and Mayraz (2008) find $\gamma = 1.26$). When we vary mothers’ earnings from 0% to 100% as much as fathers’ earnings, we find that γ ranges from $[0.72, 1.13]$.³² If we instead assume mother’s income is equal to 80% of the mean value of father’s income, we find $\gamma = 1.15$.

TABLE VIII: CURVATURE OF UTILITY

$\mathbb{E} \left[\frac{I}{P} \right]$	0.21
γ	1.05

Note: γ is estimated using Equation (14) and data on our sample of treated, eligible children aged 12-15 years in 1998 and 1999. All monetary values used in these calculations are in pesos, per-capita, per-week and are inflation adjusted to 1998 values. Parental income is measured as 1.8 times father’s income. Finally, our estimate of $\mathbb{E} \left[\frac{I}{P} \right]$ are taken from Table VII above.

VI QUANTIFYING THE TARGETING TRADE-OFF FOR PROGRESA

In this section, we quantify the targeting trade-off for Progresa households. First, we estimate the size of the targeting benefit relative to the exclusion cost under the observed, average Progresa grants. Second, we calculate the optimal CCT/UCT mix based solely on this targeting trade-off. Finally, to conclude the section, we expand our framework beyond just a targeting framework and incorporate the possibility that parents in Progresa villages misperceive the returns to education and consequently under-invest in their children’s schooling. Using this augmented framework, we conduct a back-of-the-

³²In order to rationalize the observed behavioral response (i.e., in order to rationalize the observed $\mathbb{E} \left[\frac{I}{P} \right]$), as we increase parental income, we must increase curvature. Thus, γ is increasing as we increase mom’s share of income from 0% to 100% of dad’s income.

envelope calculation for the size of the targeting benefit relative to the enrollment benefit CCTs offer when parents under-value the return to schooling. This exercise is useful as it allows us to investigate whether our targeting benefit is comparable in magnitude to one of the standard benefits CCTs offer.

VI.A Size of the Targeting Benefit relative to the Exclusion Cost under Progresa Grants

First, we estimate the size of the targeting benefit relative to the exclusion cost under the observed, average Progresa grants: $\bar{t}_c = 6.86$ and $\bar{t}_u = 2.90$ pesos, per-capita, per-week. To do so, we evaluate the following discretized version of Equation (4) for our sample of treated, eligible children in 1998 and 1999:

$$\frac{TB}{EC} = \frac{\left(1 - \frac{\bar{S} + \bar{t}_c \bar{P}}{1 + \bar{t}_c \bar{I}}\right) \sum_{it=1}^N (y_{it} + \bar{t}_u + \bar{t}_c)^{-\gamma} S(y_{it}, \bar{t}_c, \bar{t}_u)}{\left(\frac{\bar{S} + \bar{t}_c \bar{P}}{1 + \bar{t}_c \bar{I}}\right) \sum_{it=1}^N (y_{it} + \bar{t}_u + y_{c,it})^{-\gamma} (1 - S(y_{it}, \bar{t}_c, \bar{t}_u))}$$

where N denotes our sample size; $S(y_{it}, \bar{t}_c, \bar{t}_u)$ denotes predicted enrollment for a household with parental income y_{it} under the average Progresa grants (where we predict $S(y_{it}, \bar{t}_c, \bar{t}_u)$ using our estimated coefficients from Equation (12)); $\bar{S} = \frac{1}{N} \sum_{it} S(y_{it}, \bar{t}_c, \bar{t}_u)$ denotes average enrollment under average grants and is equal to 0.78; γ is taken from Table VIII above, $\gamma = 1.05$; and, finally, average income and price effects are taken from Table XV above, $\bar{I} = 0.0037$, $\bar{P} = 0.010$.

Our estimate for $\frac{TB}{EC}$ is presented in column (1) of Table IX below. We find a ratio equal to 0.88, indicating the targeting benefit is substantial relative to the exclusion cost under the average Progresa grants. Moreover, given this ratio is close to 1, this is suggestive that the current transfers are close to the optimal transfers (based on targeting arguments alone), although t_c^* will be lower than \bar{t}_c and t_u^* will be higher than \bar{t}_u . Lastly, the size of the targeting benefit is robust to the magnitude of mothers' earnings: $\frac{TB}{EC}$ stays between 0.87 and 0.88 as we vary mom's earnings from 0% to 100% of father's earnings or if we set mom's earnings to equal 80% of average father's earnings.³³

33. There are two competing effects going on as we increase mom's income share from 0% to 100%: 1) the variance of parental income is increasing which works against the targeting benefit; 2) curvature is increasing which works in favor of the targeting benefit. The reason increasing curvature works in favor of the targeting benefit is because in this setting, it turns out that the lowest consumption households are the enrolled households. Thus, as we increase curvature, we care more about transferring towards the enrolled households.

TABLE IX: SIZE OF THE TARGETING TRADE-OFF

	(1) Observed	(2) Optimal
t_c	6.86	4.49
t_u	2.90	4.85
Share of budget to CCT	65%	41%
TB/EC	0.88	1

Note: t_c, t_u are in pesos, per-capita, per-week. Observed transfers (i.e., t_c and t_u reported in column (1)) are the average transfers offered to eligible, treated children aged 12-15 in 1998 and 1999. Optimal transfers (i.e., t_c and t_u in column (2)) are the transfers s.t. the targeting benefit just offsets the exclusion cost.

VI.B Optimal CCT Based on the Targeting Trade-Off

Next we calculate the optimal CCT/UCT mix based solely on the targeting trade-off. To do so, we first determine the size of the government budget, R . We approximate this budget to match the amount of Progresa spending on the children in our sample: $R = \bar{S}t_c + \bar{t}_u = 0.78 \times 6.86 + 2.90 = 8.25$, where R represents the per-child budget (in pesos, per-capita, per-week).³⁴ Then, using data on our treated, eligible sample of children aged 12-15 in 1998 and 1999, we determine the transfer schedule that satisfies $\frac{TB}{EC} = 1$. Specifically we solve the following equation for t_c and t_u :

$$\frac{\left(1 - \frac{\bar{S}(t_c, t_u) + t_c \bar{P}(t_c, t_u)}{1 + t_c \bar{I}(t_c, t_u)}\right) \sum_{it=1}^N (y_{it} + t_u + t_c)^{-\gamma} S(y_{it}, t_c, t_u)}{\left(\frac{\bar{S}(t_c, t_u) + t_c \bar{P}(t_c, t_u)}{1 + t_c \bar{I}(t_c, t_u)}\right) \sum_{it=1}^N (y_{it} + t_u + y_{c,it})^{-\gamma} (1 - S(y_{it}, t_c, t_u))} = 1$$

$$\text{s.t. } t_u = R - t_c \bar{S}(t_c, t_u)$$

where $S(y_{it}, t_c, t_u)$ denotes predicted enrollment of a household with parental income y_{it} under the transfer schedule (t_c, t_u) ; $\bar{S}(t_c, t_u) = \frac{1}{N} \sum_{it} S(y_{it}, t_c, t_u)$ denotes the average share of children enrolled under the transfer schedule (t_c, t_u) ; $\bar{P}(t_c, t_u) = \frac{1}{N} \sum_{it} P(y_{it}, t_c, t_u)$ denotes the average price effect under the schedule (t_c, t_u) ; and $\bar{I}(t_c, t_u) = \frac{1}{N} \sum_{it} I(y_{it}, t_c, t_u)$ denotes the average income effect under the schedule (t_c, t_u) . Note, to calculate $S(y_{it}, t_c, t_u)$, we use the following relationship:

$$S(y_{it}, t_c, t_u) = S(y_{it}, \bar{t}_c, \bar{t}_u) + \int_{\bar{t}_c}^{t_c} P(y_{it}, v, \bar{t}_u) dv + \int_{\bar{t}_u}^{t_u} I(y_{it}, t_c, v) dv$$

³⁴ For example, if an eligible child aged 12-15 is from a family of size 6, we would estimate that Progresa has $6 \times 8.25 = 50$ pesos to offer this child unconditionally each week.

Results for the optimal transfers are presented in column (2) of Table IX above. Our findings suggest that 41% of the Progresa budget should be allocated towards a CCT (compared to the observed proportion of 65%), with the optimal transfers equal to $t_c^* = 4.49, t_u^* = 4.85$ pesos, per-capita, per-week. Thus, based solely on targeting arguments, it is optimal to spend nearly half of the Progresa budget on a CCT over a UCT. Moreover, if we restrict the planner's choice set to a pure UCT or a pure CCT, we find that social welfare is higher under the pure CCT.

VI.C Discussion: Why is the Targeting Benefit so Large?

There are three key factors leading to a sizable targeting benefit in this setting. First, we estimate a reasonable degree of curvature in utility of consumption (driven by the fact that we find non-negligible income effects relative to price effects). Thus, marginal utility is decreasing reasonably quickly in consumption. Second, child incomes are high relative to parent incomes: our predicted measure of child income gives a mean child income for boys around 72% as large as father's income (see Table XIII in Appendix B.3 for average, predicted child incomes). Third, there is substantial overlap in the density of per-capita parental incomes for those sending their children to school and those not. Highlighting this, Figure V plots the density of per-capita parental income split by those sending their 13 year-old child to school and those not sending their 13 year-old children to school under zero transfers (note, figures look very similar when looking at 12, 14, and 15 year-old children).³⁵ Due to the large loss in child income, the set of households that send their teenage child to school have substantially lower consumption today relative to the households that do not send their teenage child to school. Figure VI plots the density of per-capita household income for those sending their 13 year-old child to school and those not, where household income is calculated as parent income plus predicted child income if the child is not in school. As a result, the planner will want to target transfers towards the households sending their teenage child to school as these households place a greater value on receiving an extra dollar today.

One may then ask, why does the planner not allocate all of her budget towards a CCT given those sending to school have substantially lower consumption? The reason is that the behavioral cost of offering a CCT is large in this setting (the average price effect is nearly three times as large as the average income effect). Consequently, once t_c is sizable, the opportunity cost of raising t_c further is very high in terms of reducing

35. To construct the density of parental incomes for those parents sending to school, we estimate a kernel density over the observed (per-capita) parental incomes y_{it} with weights $\frac{S(y_{it},0,0)}{S(0,0)}$, where $S(y_{it}, 0, 0)$ denotes predicted enrollment for child i in year t under zero transfers. Likewise, to construct the density of parental incomes for those parents not sending to school, we estimate a kernel density over the observed incomes y_{it} with weights $\frac{1-S(y_{it},0,0)}{1-S(0,0)}$.

the UCT, which is particularly costly for the low parental income households who do not send their child to school.

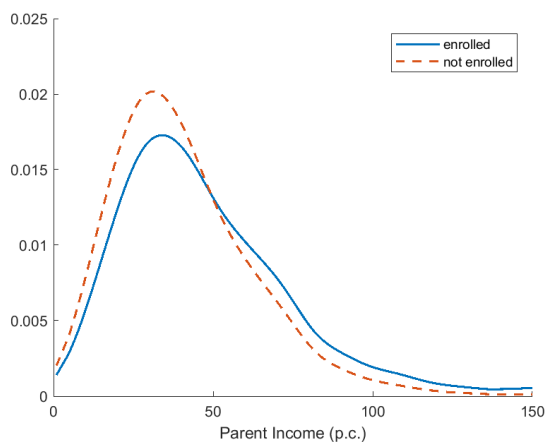


FIGURE V: DISTRIBUTION OF PARENT INCOME BY ENROLLMENT OF 13 YEAR-OLD CHILD

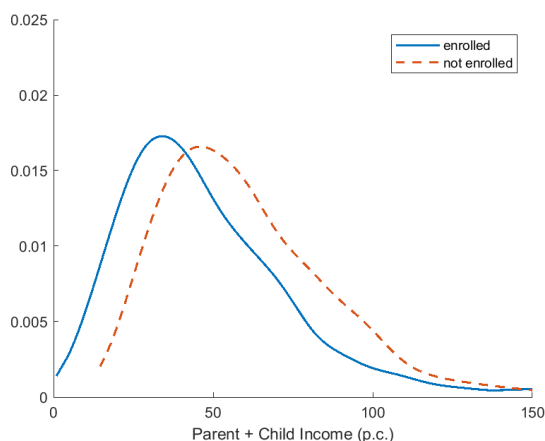


FIGURE VI: DISTRIBUTION OF HOUSEHOLD INCOME BY ENROLLMENT OF 13 YEAR-OLD CHILD

VI.D *Are we Over-Estimating the Targeting Benefit?*

One may be concerned that we are over-estimating the targeting benefit relative to the exclusion cost, perhaps because parental income is measured with error, and/or because we implicitly make the assumption that consumption is equivalent to household income (i.e., we assume households do not borrow or save). To address the first issue, we investigate to what extent the Progresa created poverty-score index overlaps for schooling and non-schooling households. This index was created by Progresa officials prior to the start of the program, and is constructed to predict whether household per-capita adult income (income of those above age 18) is above or below a poverty threshold (see Skoufias et

al (1999) for more details on how this score was created). Figure VII plots this poverty score for households with a 13 year-old child, split by whether their 13 year-old child is in school or not (figures look very similar for 12, 14, or 15 year-old children). As you can see, while the schooling households do have higher scores on average (i.e., higher per-capita household income), there is substantial overlap in these scores, consistent with our finding of substantial overlap in parental incomes by school enrollment.

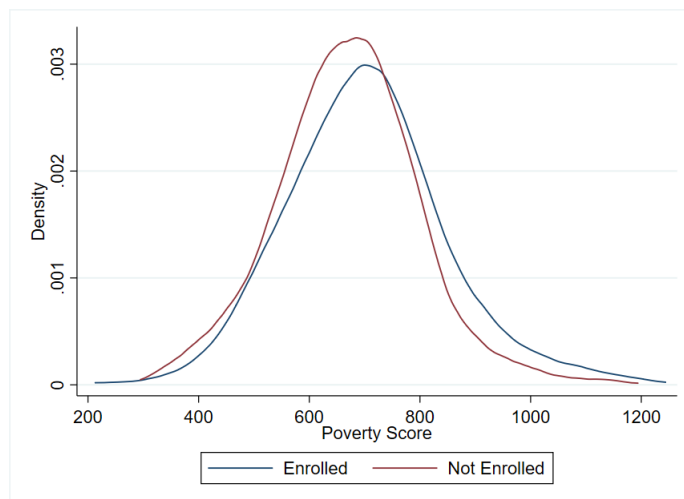


FIGURE VII: POVERTY SCORE BY SCHOOL ENROLLMENT OF 13 YEAR-OLD CHILD

Next we investigate to what extent households in these villages borrow and/or save. When asked in the November 1998 survey whether any household members borrowed from various sources (banks, friends, money lenders etc.), 97% of households replied no. When asked in the March 1998 survey, if given extra money, what would you spend it on, less than 2% of households selected debt repayment as their first choice, and less than 5% listed saving it as their first choice. Conversely, over 70% of households selected spending the additional money on food as their first choice. This is suggestive evidence that the majority of household income is consumed and is not saved or supplemented with borrowing.

However, one may still be concerned that there are ways for schooling households to help offset the child income they forgo (and/or for the non-schooling households to invest/save part of the child income they receive). If this is the case, we will overestimate household consumption for non-schooling households and underestimate household consumption for schooling households. To help address this issue, we show sensitivity of our results to varying the mean difference in parental income for schooling and non-schooling households, $\mathbb{E}[y|s = 1] - \mathbb{E}[y|s = 0]$. As we increase this difference, we increase the consumption levels of the schooling households relative to the non-schooling households. To increase $\mathbb{E}[y|s = 1] - \mathbb{E}[y|s = 0]$, we increase the rate at which enrollment increases with parental income, holding average enrollment fixed, i.e., we vary the derivative $\frac{\partial S_{it}}{\partial y_{it}}$

keeping \bar{S} constant.³⁶ Figure VIII plots the optimal share of the budget to a CCT vs. this mean difference in parental incomes. As expected, the optimal CCT is decreasing in this difference - as the schooling households get richer, we want to redistribute to them less. However, it is worth noting that even if the difference between average parental income for those sending to school and those not sending to school is twice as large as that observed, it would still be optimal to spend 20% of the budget on a CCT. Essentially, there is still a non-negligible fraction of very low parental income households sending their children to school who the planner is very concerned about directing money to.

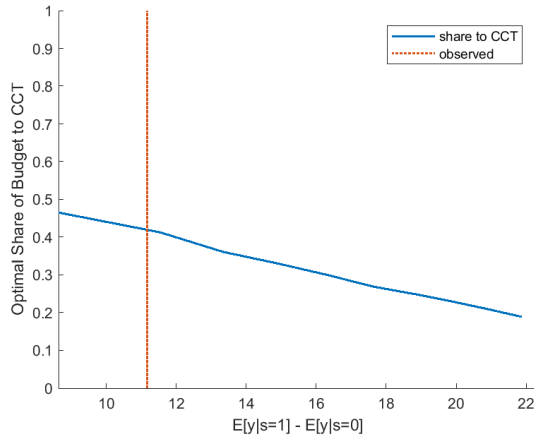


FIGURE VIII: VARYING PARENTAL INCOME INEQUALITY

Finally, in Appendix B.5 we show sensitivity of our results to varying our curvature estimate, γ , and varying the size of child incomes.

VI.E *Misperceived Returns to Schooling: Targeting Benefit vs. Enrollment Benefit*

Throughout this paper, we have assumed away any direct motives for the planner to want to increase school enrollment so as to focus solely on the targeting consequences of CCTs vs UCTs. However, a common argument for offering CCTs is that parents undervalue the return to schooling and, thus, under-invest in their children's education. In such a situation, CCTs offer an additional benefit over UCTs as they induce more parents to send their children to school (as price effects are larger than income effects); thus, CCTs help correct more for parental under-investment than UCTs. We will refer to this benefit as the enrollment benefit. One may therefore not only be interested in how large the targeting benefit of CCTs is relative to the exclusion cost, but also be interested in how large the targeting benefit is relative to the enrollment benefit. The goal of this

36. Increasing $\frac{\partial S_{it}}{\partial y_{it}}$ will increase the mean difference in parental incomes for schooling and non-schooling households as $\mathbb{E}[y|s=1] = \frac{1}{N} \sum_{it=1}^N \frac{S_{it}y_{it}}{\bar{S}}$ and $\mathbb{E}[y|s=0] = \frac{1}{N} \sum_{it=1}^N \frac{(1-S_{it})y_{it}}{1-\bar{S}}$.

subsection is to provide a back-of-the-envelope calculation as to how large the targeting benefit is relative to the enrollment benefit in Progresa villages.

Modeling the Enrollment Benefit. We begin by modeling the enrollment benefit so as to obtain a mathematical expression for the targeting benefit relative to the enrollment benefit, $\frac{TB}{EB}$. As before, denote parents' utility over consumption in generation two as $v(\mu, s)$; however, now denote the planner's utility over consumption in generation two as $w(\mu, s)$, where $w(\mu, 0) = v(\mu, 0) \forall \mu$ (i.e., the planner and parents agree on the return to not sending to school), but $w(\mu, 1) > v(\mu, 1) \forall \mu$ (i.e., the planner places a greater value on the return to schooling than parents for all child ability levels). Thus, the planner's first order condition is now (see Appendix A.5 for the planner's problem and derivation of the first order condition below):

$$\underbrace{\left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y u_c(y + t_u + t_c) S(y) f(y) dy}_{\text{Targeting Benefit (TB)}} + \underbrace{\frac{\partial t_u}{\partial t_c} \int_Y u_c(y + y_c + t_u) (1 - S(y)) f(y) dy}_{\text{-Exclusion Cost (EC)}} + \underbrace{\int_Y \left(P(y) + I(y) \frac{\partial t_u}{\partial t_c}\right) (w(\tilde{\mu}, 1) - v(\tilde{\mu}, 1)) f(y) dy}_{\text{Enrollment Benefit (EB)}} = 0$$

where the enrollment benefit captures the increase in enrollment induced by increasing the CCT by one dollar multiplied by the extent to which (marginal) parents misperceive the return to an additional year of education. From this first order condition, we obtain the following expression for the targeting benefit relative to the enrollment benefit:

$$\frac{TB}{EB} = \frac{\left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y u_c(y + t_u + t_c) S(y) f(y) dy}{\int_Y \left(P(y) + I(y) \frac{\partial t_u}{\partial t_c}\right) (w(\tilde{\mu}, 1) - v(\tilde{\mu}, 1)) f(y) dy} \quad (15)$$

Estimating the Targeting Benefit to the Enrollment Benefit. We now proceed with a back-of-the-envelope calculation to evaluate $\frac{TB}{EB}$ for Progresa villages. To do so, we first need to make some assumptions around the term $w(\tilde{\mu}, 1) - v(\tilde{\mu}, 1)$. This term captures the extent to which (marginal) parents misperceive the return to an additional year of education. For our sample of 12-15 year-old children in Progresa villages, we will make the following assumptions: (1) children will work for 50 years as adults (this corresponds to a retirement age of 62-65); (2) there is an annual discount rate is 0.05; (3) parents believe an additional year in school leads to a 2% gain in weekly income while the actual gain is 8% (where, for simplicity, we assume weekly income is equal to the average, per-capita,

weekly income of parents in our sample, \bar{y}). These estimates for the perceived and actual returns to an additional year of school are taken from Jensen (2010) where he estimates that in the Dominican Republic the return to an additional year of secondary school translates into an 8% increase in monthly income, while the perceived increase in monthly income is only 2% (note, few papers have investigated the difference between actual and perceived returns to schooling, hence, why we use estimates from the Dominican Republic as opposed to Mexico).³⁷ Finally, we assume that annual utility of consumption is CRRA with parameter $\gamma = 1.05$. Thus, we get the following expression for $w(\tilde{\mu}, 1) - v(\tilde{\mu}, 1)$:

$$w(\tilde{\mu}, 1) - v(\tilde{\mu}, 1) = \sum_{t=1}^{50} 0.95^t \frac{(1.08\bar{y})^{1-\gamma} - (1.02\bar{y})^{1-\gamma}}{1-\gamma}$$

We then proceed to evaluate the discretized version of Equation (15) under the average Progresca grants (for our sample of treated, eligible children in years 1998 and 1999):

$$\frac{TB}{EB} = \frac{\left(1 - \frac{\bar{S} + \bar{t}_c \bar{P}}{1 + \bar{t}_c \bar{I}}\right) \sum_{it=1}^N (y_{it} + \bar{t}_u + \bar{t}_c)^{-\gamma} S_{it}(\bar{t}_c, \bar{t}_u)}{N \left(\bar{P} - \bar{I} \left(\frac{\bar{S} + \bar{t}_c \bar{P}}{1 + \bar{t}_c \bar{I}}\right)\right) 17.5 \frac{(1.08\bar{y})^{1-\gamma} - (1.02\bar{y})^{1-\gamma}}{1-\gamma}} = 0.38$$

We find that under the average Progresca grants, the targeting benefit is 38% as large as the enrollment benefit. Alternatively, if the actual return to an additional year of schooling is only 4% (and parents still perceive it as 2%) then $\frac{TB}{EB} = 1.12$, whereas if the actual return to an additional year of schooling is 12%, then $\frac{TB}{EB} = 0.23$. Thus, the targeting benefit is comparable in magnitude to the enrollment benefit, and therefore should be taken into account when trading-off the costs and benefits of CCTs vs. UCTs.

VII CONCLUSION

In this paper, we argue that there exists an unexplored targeting benefit of imposing conditions on cash transfers to send children to school: by imposing conditions, a planner can direct more money to those households who forgo a discrete loss of child income. We argue that this benefit mitigates the adverse targeting effects that arise from imposing conditions on a normal investment. Ultimately, we argue that there exists a targeting trade-off when considering how to allocate a budget between a CCT and a UCT, and that it can be optimal to allocate some of a budget to a CCT based on targeting arguments alone.

We formalize this intuition by developing a theoretical framework that captures the

37. Note, Attanasio and Kaufmann (2014) investigate perceived returns (of both mothers and children) of education in Mexico. However, they are unable to compare these perceived returns to actual returns given differences between samples for which perceived and actual returns are calculated.

targeting effects of offering a CCT over a UCT. We show that a social planner faces a key trade-off: by increasing the share of the budget to a CCT, she increases the transfers received by schooling households who have forgone a discrete amount of child income (the targeting benefit); however, she reduces the transfers received by the non-schooling households who are, on average, lower parental income households (the exclusion cost). We show that, at least in theory, the targeting benefit can dominate the exclusion cost, thus making it optimal to allocate some of the budget to a CCT. This result goes against the current consensus that CCTs are unambiguously worse at targeting transfers within a set of beneficiary households.

We then attempt to quantify the importance of the targeting benefit in practice. To do so, we first express the size of the targeting benefit relative to the exclusion cost in terms of empirically observable quantities. We show that two of the relevant quantities are the income effect of a UCT and the price effect of a CCT. These two elasticities allow us to pin down the curvature of utility, thus allowing us to calculate the extent to which the households sending their children to school value receiving an extra dollar relative to the households not sending their children to school.

We then proceed to estimate income and price effects for secondary-school enrollment for a large CCT in rural Mexico, Progresa. We use “conditional” transfers to children’s siblings under the age of 12 to identify income effects as nearly 100% of children under the age of 12 are enrolled (prior to receiving any grants). Using these elasticities, we estimate that 41% of the Progresa budget should be allocated to a CCT over a UCT based on targeting arguments alone. This implies that the targeting benefit can be a quantitatively important benefit of CCTs. Three key empirical factors drive this finding for the Progresa setting: (1) forgone child incomes are large; (2) parental income differences between beneficiary households sending to school and beneficiary households not sending to school are small; and (3) curvature in utility of consumption is substantial (our income and price effect estimates imply a coefficient of constant relative risk aversion of 1.05). Thus, by allocating some of the budget towards a CCT, the planner can better target transfers towards households who place a higher value on receiving an extra dollar today.

Moving forward, we believe our results have several implications for the design of cash transfer programs. First, it is important to understand the magnitudes of the behavioral schooling elasticities of both CCTs and UCTs in a pilot study. These elasticities are important not only for evaluating how effective the program will be at increasing enrollment, but also for understanding the curvature of utility - a critical parameter for any optimal redistribution problem. Second, it is important to have accurate data on the cost of schooling (e.g., forgone child incomes) and parental incomes as both these variables are crucial for determining the magnitude of the targeting benefit relative to the exclusion cost. And finally, although we abstract from under-investment in this paper so as to focus

on the targeting trade-off, it is important to know the extent to which school enrollment levels are below socially optimal levels (e.g., it is important to know the extent to which parents undervalue the returns to schooling).

Combining all these factors together, Tables X and XI offer a simple heuristic to understand which situations CCTs are preferable to UCTs. First, if income effects are large relative to price effects, this implies that (a) there is a high degree of curvature in utility of consumption, and (b) both a UCT and CCT will have similar effects on enrollment. Thus, all that will matter for deciding on whether to offer a UCT vs. a CCT is which households have the lowest consumption today. If the households who send to school have much lower consumption, one should offer a CCT, whereas, if the households that do not send to school have much lower consumption, one should offer a UCT (see Table X). Second, if income effects are small relative to price effects, this implies that (a) utility is fairly linear in consumption, and (b) CCTs induce a much larger increase in enrollment than UCTs. If this is the case, all that will matter for deciding on whether to offer a UCT vs. a CCT is whether there is under-enrollment or not (e.g., whether parents undervalue the return to schooling and/or whether there are positive externalities associated with increased schooling). If there is no under-enrollment, one should offer a UCT so as to not distort parental enrollment decisions, whereas, if there is substantial under-enrollment, one should offer a CCT (see Table XI).

TABLE X: LARGE INCOME EFFECTS: $I \approx P$

	No under-enrollment	High under-enrollment
$cons_{sch} \ll cons_{no\ sch}$	CCT	CCT
$cons_{sch} \gg cons_{no\ sch}$	UCT	UCT

TABLE XI: SMALL INCOME EFFECTS: $I \ll P$

	No under-enrollment	High under-enrollment
$cons_{sch} \ll cons_{no\ sch}$	UCT	CCT
$cons_{sch} \gg cons_{no\ sch}$	UCT	CCT

Looking forward, we believe much of our analysis around the targeting benefit applies to settings beyond the CCT/UCT discussion in developing countries. In particular, governments redistributing towards college students (who may have a currently high marginal utility of consumption if borrowing markets are incomplete) may be justified by the same targeting principle. Similarly, child tax credits can be justified using the same sort of logic: having a child is a discrete decision that lowers per-capita consumption (as now income must be split among an additional person) so that directing funds towards these families is useful due to the targeting benefit. Hence, we expect analyzing targeting benefits in relation to discrete decisions may yield further insights into a variety of optimal

redistribution problems.

REFERENCES

- ALATAS, V., A. BANERJEE, R. HANNA, B. OLKEN, R. PURNAMASARI, AND M. WAI-POI (2016): “Self-Targeting: Evidence from a Field Experiment in Indonesia,” *Journal of Political Economy* vol. 124(3): 371-427. <http://dx.doi.org/10.1257/aer.102.4.1206>
- ALATAS, V., A. BANERJEE, R. HANNA, B. OLKEN, AND J. TOBIAS (2012): “Targeting the Poor: Evidence from a Field Experiment in Indonesia,” *American Economic Review* vol. 102 (4): 1206-1240. <http://dx.doi.org/10.1257/aer.102.4.1206>
- AKERLOF, G. (1970): “The Market for Lemons: Quality Uncertainty and the Market Mechanism,” *Quarterly Journal of Economics* vol. 84 (3): 488-500.
- ANGELUCCI, M. AND G. DE GIORGI (2009): “Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles’ Consumption,” *American Economic Review* vol. 99, 486-508.
- AKRESH, R., D. DE WALQUE, AND H. KAZIANGA (2013): “Cash Transfers and Child Schooling: Evidence from a Randomized Evaluation of the Role of Conditionality,” *World Bank Policy Research Working Paper* 6340.
- ATTANASIO O., AND K. KAUFMANN (2014): “Education choices and returns to schooling: Mothers’ and youths’ subjective expectations and their role by gender,” *Journal of Development Economics* vol. 109, 203-216.
- ATTANASIO O., C. MEGHIR, AND A. SANITAGO (2011): “Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA,” *Review of Economic Studies* vol. 79, 37-66.
- BAIRD, S., B. ÖZLER, AND C. MCINTOSH (2011): “Cash or Condition? Evidence from a Cash Transfer Experiment,” *Quarterly Journal of Economics* vol. 126, 1709-1753.
- BAIRD, S., F. FERREIRA, B. ÖZLER, AND M. WOOLCOCK (2013): “Relative Effectiveness of Conditional and Unconditional Cash Transfers for Schooling Outcomes in Developing Countries: A Systematic Review,” *The Campbell Collaboration*
- BANERJEE, A., R. HANNA, J. KYLE, B. OLKEN, AND S. SUMARTO (2018): “Tangible Information and Citizen Empowerment: Identification Cards and Food Subsidy Programs in Indonesia,” *Journal of Political Economy* vol. 126(2): 451-491.

- BARSKY, R., R. JUSTER, M. KIMBALL, AND M. SHAPIRO: “Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study.” *Quarterly Journal of Economics* vol. CXII (1997), 537-580.
- BASTAGLI, F., J. HAGEN-ZANKER, L. HARMAN, V. BARCA, G. STURGE, T. SCHMIDT, AND L. PELLERANO (2016): “Cash transfers: what does the evidence say? A rigorous review of programme impact and of the role of design and implementation features,” *Overseas Development Institute*
- BENHASSINE, N., F. DEVOTO, E. DUFLO, P. DUPAS, AND V. POULIQUEN (2015): “Turning a Shove into a Nudge? A ”Labeled Cash Transfer” for Education,” *American Economic Journal: Economic Policy* vol. 7(3), 86-125.
- BOURGUIGNON, F., F. FERREIRA, AND P. LEITE (2003): “Cash Transfers, Schooling, and Child Labor: Micro-Simulating Brazils Bolsa Escola Program” *World Bank Economic Review* vol. 17, 229-254.
- COADY, D., M. GROSH, AND J. HODDINOTT (2004): “Targeting of Transfers in Developing Countries: Review of Lessons and Experience,” *The International Bank for Reconstruction and Development, The World Bank*.
- COHEN, A., AND L. EINAV (2007): “Estimating Risk Preferences from Deductible Choice,” *American Economic Review*. vol. 97(3), 745-788.
- CHETTY, R. (2006): “A New Method of Estimating Risk Aversion,” *American Economic Review* vol. 96(5), 1821-1834.
- DE BRAUW, A. AND J. HODDINOTT (2011): “Must conditional cash transfer programs be conditioned to be effective? The impact of conditioning transfers on school enrollment in Mexico,” *Journal of Development Economics* vol. 96, 356-370.
- DE JANVRY, A. F. FINAN, E. SADOULET, AND R. VAKIS (2006): “Serve as Safety Nets in Keeping Children at School and from Working When Exposed to Shocks?,” *Journal of Development Economics* vol. 79, 349-373.
- EDMONDS, E. (2006): “Child Labor and Schooling Responses to Anticipated Income in South Africa,” *Journal of Development Economics* vol. 81, 386-414.
- EDMONDS, E., AND SCHADY, N. (2012): “Poverty Alleviation and Child Labor,” *American Economic Journal: Economic Policy* vol. 4(4), 100-124.
- FERREIRA, F., D. FILMER, AND N. SCHADY (2018): “Own and Sibling Effects of Conditional Cash Transfer Programs: Theory and Evidence from Cambodia,” *S. Bandyopadhyay (ed.) Research in Economic Inequality* vol. 25, 259-298.

- FILMER, D. AND N. SCHADY (2008): “Getting Girls into School: Evidence from a Scholarship Program in Cambodia,” *Economic Development and Cultural Change* vol. 56(2), 581-617.
- FILMER, D. AND N. SCHADY (2011): “Does More Cash in Conditional Cash Transfer Programs Always Lead to Larger Impacts on School Attendance?,” *Journal of Development Economics* vol. 96, 150-157.
- FISZBEIN, A. AND N. SCHADY (2009): “Conditional Cash Transfers: Reducing Present and Future Poverty,” *The International Bank for Reconstruction and Development / The World Bank*
- FREELAND, N. (2007): “Superfluous, Pernicious, Atrocious and Abominable? The Case Against Conditional Cash Transfers,” vol. 38, Number 3. *IDS Bulletin, Institute of Development Studies*
- GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): “Optimal Indirect and Capital Taxation,” *The Review of Economic Studies* vol. 70(3), 569-587.
- HANNA, R. AND B. OLKEN (2018): “Universal Basic Incomes vs. Targeted Transfers: Anti-Poverty Programs in Developing Countries,” *NBER Working Paper* 24939. <http://www.nber.org/papers/w24939>
- JENSEN, R. (2010): “The (Perceived) Returns to Education and the Demand for Schooling,” *Quarterly Journal of Economics* vol. 125(2), 515-548.
- KAPLOW (2005): “The Value of a Statistical Life and the Coefficient of Relative Risk Aversion,” *Journal of Risk and Uncertainty* vol. 31(1), 23-34.
- KOCHERLAKOTA, N. (2005): “Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation,” *Econometrica* vol. 73(5), 1587-1621.
- MEHRA, R. AND PRESCOTT E. (1985): “The Equity Premium: A Puzzle,” *Journal of Monetary Economics* vol. 15(2), 145-161.
- LAYARD, R., S. NICKELL, AND G. MAYRAZ (2008): “The Marginal Utility of Income,” *Journal of Public Economics* vol. 92, 1846-1857.
- PARKER, S. AND E. SKOUFIAS (2000): “The Impact of PROGRESA on Work, Leisure and Time Allocation,” *International Food Policy Research Institute*
- PARKER, S. AND E. SKOUFIAS (2001): “Conditional Cash Transfers and Their Impact on Child Work and Schooling: Evidence from the PROGRESA Program in Mexico,” *Economía* vol. 2(1), 45-96.

- PARKER, S. AND P. TODD (2017): “Conditional Cash Transfers: The Case of Progres/Oportunidades,” *Journal of Economic Literature* vol. 55(3), 866-915.
- RAVALLION, M. (2009): “How Relevant is Targeting to the Success of an Antipoverty Program? A Case Study for China,” *World Bank Research Observer* vol. 24(2): 205-231.
- RODRÍGUEZ-CASTELÁN, C. (2017): “Conditionality as Targeting? Participation and Distributional Effects of Conditional Cash Transfers,” *Poverty and Equity Global Practice Group*
- SCHADY, N., AND M. ARAUJO (2008): “Cash Transfers, Conditions, and School Enrollment in Ecuador,” *Economica* vol. 8, 4370.
- SCHULTZ, T. (2004): “School subsidies for the poor: evaluating the Mexican Progres poverty program,” *Journal of Development Economics* vol. 74(1), 199-250.
- SKOUFIAS, E., B. DAVIS, AND J. BEHRMAN (1999): “An Evaluation of the Selection of Beneficiary Households in the Education, Health, and Nutrition Program (PROGRESA) of Mexico,” *International Food Policy Research Institute*
- THORISSON, H. (1995): “Coupling Methods in Probability Theory,” *Scandinavian Journal of Statistics* vol. 22(2), 159-182.
- TODD, P. AND K. WOLPIN (2006): “Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility,” *American Economic Review* vol. 96(5), 1384-1417.

A APPENDIX

A.1 Proof of Lemma 2

Denote $\tilde{\mu}(y)$ s.t. parents with $\mu \geq \tilde{\mu}(y)$ send to school and vice versa. We can write the expected value of parental income for those sending to school as follows:

$$\mathbb{E}[y|s = 1] = \frac{1}{\int_Y 1 - F(\tilde{\mu}(y)|y)} \int_Y y(1 - F(\tilde{\mu}(y)|y)) f(y) dy = \mathbb{E}[y] + \frac{\text{Cov}(y, 1 - F(\tilde{\mu}(y)|y))}{\int_Y 1 - F(\tilde{\mu}(y)|y) f(y) dy}$$

This covariance term will be non-negative if $\frac{\partial(1-F(\tilde{\mu}(y)|y))}{\partial y} = -f(\tilde{\mu}(y)|y) \frac{\partial \tilde{\mu}(y)}{\partial y} + \frac{\partial(1-F(\tilde{\mu}(y)|y))}{\partial y} \geq 0$.³⁸ Simple implicit function theorem arguments can show that $\frac{\partial \tilde{\mu}(y)}{\partial y} \leq 0$ and our assumption that $F(\mu|y)$ FOSD $F(\mu|y')$ $\forall y > y'$ ensures that the second term is (weakly) positive. Hence, $\mathbb{E}[y|s = 1] \geq \mathbb{E}[y] \implies \mathbb{E}[y|s = 1] \geq \mathbb{E}[y|s = 0]$.

A.2 Proof: $t_c^* < y_c$

We will show that $t_c^* < y_c$ so that offering a conditional transfer doesn't necessarily result in all households choosing to send to school. Suppose not, i.e. $t_c = y_c$. Then everyone sends to school as $u(y + t_c + t_u) + v(\mu, 1) > u(y + y_c + t_u) + v(\mu, 0) \forall y, \mu$ as $v(\mu, 1) > v(\mu, 0)$ and $u(y + t_c + t_u) = u(y + y_c + t_u)$ as $t_c = y_c$. The planner's FOC is then $\left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y u_c(y + t_c + t_u) f(y) dy$. But $\frac{\partial t_u}{\partial t_c} = -1 - t_c \bar{P} - t_c \bar{I} \frac{\partial t_u}{\partial t_c} < -1$ as $\bar{P} = \int_Y -\frac{\partial \tilde{\mu}(y)}{\partial t_c} f(\tilde{\mu}|y) f(y) dy = \int_Y \frac{u_c(y+y_c+t_u)}{v_1(\tilde{\mu}, 1) - v_1(\tilde{\mu}, 0)} f(\tilde{\mu}|y) f(y) dy > 0$, $\bar{I} = \int_Y \frac{-\partial \tilde{\mu}}{\partial t_u} f(\tilde{\mu}|y) f(y) dy = \int_Y \frac{u_c(y+t_c+t_u) - u_c(y+y_c+t_u)}{v_1(\tilde{\mu}, 1) - v_1(\tilde{\mu}, 0)} f(\tilde{\mu}(y)|y) f(y) dy = 0$ when $t_c = y_c$. Thus, the planner's FOC is strictly negative when $t_c = y_c$.

A.3 Proof: $\frac{\partial t_u}{\partial t_c} < 0$

Taking the derivative of the planner's budget constraint w.r.t. t_c and rearranging we get: $\frac{\partial t_u}{\partial t_c} = \frac{-\int_Y \int_{\tilde{\mu}} f(\mu, y) d\mu dy + t_c \int_Y \frac{\partial \tilde{\mu}}{\partial t_c} f(\tilde{\mu}|y) f(y) dy}{1 - t_c \int_Y \frac{\partial \tilde{\mu}}{\partial t_u} f(\tilde{\mu}|y) f(y) dy}$. Since $\frac{\partial \tilde{\mu}}{\partial t_u} = -\frac{u_c(y+t_u+t_c) - u_c(y+y_c+t_u)}{v_\mu(\tilde{\mu}, 1) - v_\mu(\tilde{\mu}, 0)} \leq 0$ (for $t_c < y_c$) and $\frac{\partial \tilde{\mu}}{\partial t_c} = -\frac{u_c(y+t_u+t_c)}{v_\mu(\tilde{\mu}, 1) - v_\mu(\tilde{\mu}, 0)} < 0$, the numerator of $\frac{\partial t_u}{\partial t_c}$ is negative while the denominator is positive, hence $\frac{\partial t_u}{\partial t_c} < 0 \forall t_c < y_c$. We show above (in Appendix A.2) that $t_c^* < y_c$.

38. See Thorisson (1995) for a proof that the covariance of two increasing functions of a random variable is positive.

A.4 Planner can condition on parental income

We now assume the planner can condition on parental income y and schooling decisions but not child ability. Thus, the planner can now offer transfers $t_u(y)$ and $t_c(y)$ (we still assume the planner can only distribute money, i.e., $t_u(y) \geq 0$, $t_c(y) \geq 0 \forall y$). The planner solves the following:

$$\begin{aligned} & \max_{t_u(y), t_c(y)} \int_Y \int_{\tilde{\mu}} u(y + t_u(y) + t_c(y)) + v(\mu, 1) dF(\mu, y) \\ & \quad \int_Y \int_{\tilde{\mu}} u(y + y_c + t_u(y)) + v(\mu, 0) dF(\mu, y) \\ \text{s.t.} & \quad \int_Y t_u(y) dF(y) + \int_Y \int_{\tilde{\mu}} t_c(y) dF(\mu, y) \leq R \\ & \quad \& t_c(y) \geq 0 \forall y \\ & \quad \& t_u(y) \geq 0 \forall y \end{aligned}$$

We can write the Lagrangian as follows

$$\begin{aligned} & \int_Y \int_{\tilde{\mu}} u(y + t_u(y) + t_c(y)) + v(\mu, 1) dF(\mu, y) + \int_Y \int_{\tilde{\mu}} u(y + y_c + t_u(y)) + v(\mu, 0) dF(\mu, y) \\ & + \lambda \left(R - \int_Y t_u(y) dF(y) - \int_Y \int_{\tilde{\mu}} t_c(y) dF(\mu, y) \right) + \int_Y \delta_1(y) t_c(y) dy + \int_Y \delta_2(y) t_u(y) dy \end{aligned}$$

We now consider perturbing $t_u(y)$ by $d\tau$ on the interval $[y, y + \epsilon]$, and $t_c(y)$ by $d\tau$ on the interval $[y, y + \epsilon]$. The idea is that starting from an optima, the net effect of any small perturbation on the Planner's Lagrangian should be 0. Thus, the optimal transfer schedules $t_u(y)$ and $t_c(y)$ must satisfy the following conditions, respectively:

$$\epsilon d\tau (u_c(y + t_u(y) + t_c(y))S(y) + u_c(y + y_c + t_u(y))(1 - S(y)) - \lambda(1 + t_c(y)I(y)) + \delta_2(y)) = 0 \quad (16)$$

$$\epsilon d\tau (u_c(y + t_u(y) + t_c(y))S(y) - \lambda(S(y) + t_c(y)P(y)) + \delta_1(y)) = 0 \quad (17)$$

where $I(y) = \frac{\partial \tilde{\mu}(y)}{\partial t_u(y)} f(\tilde{\mu}|y)$ denotes the income effect (i.e, the increase in the share of parents with income y sending to school when we increase their unconditional cash transfer by \$1), $P(y) = \frac{\partial \tilde{\mu}(y)}{\partial t_c(y)} f(\tilde{\mu}|y)$ denotes the price effect (i.e, the increase in the share of parents with income y sending to school when we increase their conditional cash

transfer by \$1), and $S(y) = \int_{\tilde{\mu}} f(\tilde{\mu}|y)d\mu$ denotes the share of parents with income y sending to school.

We will now show that if utility is concave in consumption and $t_u(y) > 0$ for some parent income level y , then $t_c(y) > 0$. Suppose not, i.e., $t_u(y) > 0$ but $t_c(y) = 0$. From Equation (16) we get

$$u_c(y + t_u(y) + t_c(y))S(y) + u_c(y + y_c + t_u(y))(1 - S(y)) = \lambda \quad (18)$$

And from Equation (17) we get

$$u_c(y + t_u(y) + t_c(y))S(y) = \lambda S(y) - \delta_1(y) < \lambda S(y) \quad (19)$$

as $S(y) \leq 1$ and $\delta_1(y) > 0$ (as $t_c(y) = 0$). But $u_c(y + t_u(y) + t_c(y)) > u_c(y + y_c + t_u(y))$ (as we assume utility is concave), hence, by Equation (18) we get $u_c(y + t_u(y) + t_c(y)) > \lambda$, a contradiction with Equation (19). Thus, if utility is concave, it can *never* be optimal to offer a pure UCT.

A.5 Undervaluing the Return to Schooling

To highlight the importance of the targeting benefit, we assumed away any other benefits of imposing conditions, e.g., we assume that parents correctly infer the return to schooling. We now relax this assumption. As before, we assume parents' utility over consumption in generation two is given by $v(\mu, s)$; however, we now assume the planner's utility over consumption in generation two is given by $w(\mu, s)$, where $v(\mu, 0) = w(\mu, 0) \forall \mu$ (i.e., planner and parents agree on the return to not sending to school), but $w(\mu, 1) > v(\mu, 1) \forall \mu$ (i.e., the planner places a greater value on the return to schooling than parents for all child ability levels). The planner's problem can now be written as:

$$\begin{aligned} & \max_{t_u, t_c} \int_Y \int_{\tilde{\mu}(y, t_u, t_c)} u(y + t_u + t_c) + w(\mu, 1) f(\mu, y) d\mu dy + \\ & \int_Y \int_{\tilde{\mu}(y, t_u, t_c)} u(y + y_c + t_u) + w(\mu, 0) f(\mu, y) d\mu dy \\ & \text{s.t. } t_u + t_c \int_Y \int_{\tilde{\mu}(y, t_u, t_c)} f(\mu, y) d\mu dy \leq R \end{aligned}$$

where $\tilde{\mu}(y, t_u, t_c)$ denotes the indifferent household, defined implicitly as follows:

$$u(y + t_u + t_c) + v(\tilde{\mu}, 1) = u(y + y_c + t_u) + v(\tilde{\mu}, 0)$$

The planner's first order condition can now be expressed as

$$\begin{aligned} & \left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y \int_{\tilde{\mu}}^{\infty} u_c(y + t_u + t_c) f(\mu, y) d\mu dy + \frac{\partial t_u}{\partial t_c} \int_Y \int^{\tilde{\mu}} u_c(y + y_c + t_u) f(\mu, y) d\mu dy - \\ & \int_Y \left(\frac{\partial \tilde{\mu}}{\partial t_c} + \frac{\partial \tilde{\mu}}{\partial t_u} \frac{\partial t_u}{\partial t_c} \right) (u(y + t_u + t_c) + w(\tilde{\mu}, 1) - u(y + y_c + t_u) - w(\tilde{\mu}, 0)) f(\tilde{\mu}|y) f(y) dy = 0 \end{aligned}$$

Substituting in that $u(y + t_u + t_c) - u(y + y_c + t_u) = v(\tilde{\mu}, 0) - v(\tilde{\mu}, 1)$ and noting that $w(\tilde{\mu}, 0) - v(\tilde{\mu}, 0) = 0$, we get:

$$\begin{aligned} & \underbrace{\left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y \int_{\tilde{\mu}}^{\infty} u_c(y + t_u + t_c) f(\mu, y) d\mu dy}_{\text{Targeting Benefit (TB)}} + \underbrace{\frac{\partial t_u}{\partial t_c} \int_Y \int^{\tilde{\mu}} u_c(y + y_c + t_u) f(\mu, y) d\mu dy}_{\text{Exclusion Cost (EC)}} \\ & + \underbrace{\int_Y - \left(\frac{\partial \tilde{\mu}}{\partial t_c} + \frac{\partial \tilde{\mu}}{\partial t_u} \frac{\partial t_u}{\partial t_c} \right) (w(\tilde{\mu}, 1) - v(\tilde{\mu}, 1)) f(\tilde{\mu}|y) f(y) dy}_{\text{Enrollment Benefit (EB)}} = 0 \end{aligned} \tag{20}$$

Denote the share of households sending their child to school with parent income y as:

$$S(y) = \int_{\tilde{\mu}} f(\mu|y) d\mu$$

Note, we are suppressing the fact that this share is also a function of the transfer schedule, as $\tilde{\mu}$ is a function of the transfer schedule. Now denote the income effect as:

$$I(y) = \frac{\partial S(y)}{\partial t_u} = -\frac{\partial \tilde{\mu}}{\partial t_u} f(\tilde{\mu}|y)$$

And denote the price effect as:

$$P(y) = \frac{\partial S(y)}{\partial t_c} = -\frac{\partial \tilde{\mu}}{\partial t_c} f(\tilde{\mu}|y)$$

This allows us to rewrite Equation (20) as follows:

$$\begin{aligned}
& \underbrace{\left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y u_c(y + t_u + t_c) S(y) f(y) dy}_{\text{Targeting Benefit (TB)}} + \underbrace{\frac{\partial t_u}{\partial t_c} \int_Y u_c(y + y_c + t_u) (1 - S(y)) f(y) dy}_{\text{Exclusion Cost (EC)}} \\
& + \underbrace{\int_Y \left(P(y) + I(y) \frac{\partial t_u}{\partial t_c}\right) (w(\tilde{\mu}, 1) - v(\tilde{\mu}, 1)) f(\tilde{\mu}|y) f(y) dy}_{\text{Enrollment Benefit (EB)}} = 0
\end{aligned}$$

where we have a new term relative to the FOC given by Equation (3), which we call the “Enrollment Benefit”. A sufficient condition for this term to be positive is $\frac{\partial t_u}{\partial t_c} \geq -1$, i.e., as we increase t_c by a dollar, t_u does not decrease by more than a dollar (note we always assume this, as if $\frac{\partial t_u}{\partial t_c} < -1$, the targeting benefit will be negative).³⁹ The enrollment benefit captures the increase in enrollment experienced by increasing t_c by one dollar (this is captured by the term $P(y) + I(y) \frac{\partial t_u}{\partial t_c}$), multiplied by the extent to which the marginal households misperceive the return to schooling.

A.6 Proof of Proposition 3

Proof. First, let us show that if, under a pure UCT, the average marginal utility of consumption is greater for schooling households relative to non-schooling households, then $t_c^* > 0$. Under a pure UCT ($t_u = R, t_c = 0$), our FOC w.r.t. t_c is given by

$$(1 - \bar{S}) \int_Y u_c(y + R) S(y) f(y) dy - \bar{S} \int_Y u_c(y + y_c + R) (1 - S(y)) f(y) dy$$

where $S(y) = \int_{\tilde{\mu}(y)} f(\mu|y) d\mu$ denotes average enrollment for households with parent income y , and where $\bar{S} = \int_Y S(y) f(y) dy$ denotes average enrollment. Note, when $t_c = 0$ we get $\frac{\partial t_u}{\partial t_c} = -\bar{S}$. Diving through by $\frac{1}{(1-\bar{S})\bar{S}}$ we get

$$\frac{1}{\bar{S}} \int_Y u_c(y + R) S(y) f(y) dy - \frac{1}{1 - \bar{S}} \int_Y u_c(y + y_c + R) (1 - S(y)) f(y) dy$$

which is simply the average marginal utility of consumption of the schooling households less the average marginal utility of consumption of the non-schooling households. Thus, if this difference is positive, then the FOC w.r.t. t_c at $t_c = 0$ is positive meaning a

³⁹ $P(y) > I(y) \forall y$, and, by assumption, $w(\mu, 1) - v(\mu, 1) > 0 \forall \mu$. Hence if $\frac{\partial t_u}{\partial t_c} \geq -1$, $\left(P(y) + I(y) \frac{\partial t_u}{\partial t_c}\right) (w(\tilde{\mu}, 1) - v(\tilde{\mu}, 1)) > 0 \forall y$; hence, $EB > 0$. To see why $P(y) > I(y) \forall y$, see Equations (5) and (6).

pure UCT cannot be optimal.

Now, let us show via an example that it can be the case that the average marginal utility of consumption of the schooling households is greater than the average marginal utility of consumption of the non-schooling households under a pure UCT. Suppose that $u(c) = 10c^{1/2}$. As a simplification, assume away heterogeneity in μ , such that all individuals have $\mu = 10$. Further suppose $v(\mu, s)$ is as follows: $v(10, 1) = 10$ and $v(10, 0) = 0$. Lastly suppose $R = 3$, $y_c = 7$, $y \sim Unif[4, 8]$. In this example, it is easy to calculate that all individuals with income above 6 send their children to school when $t_u = R$, $t_c = 0$. Consider the FOC of the planner's problem with respect to t_c evaluated at $t_u = R$, $t_c = 0$, noting that $\frac{\partial t_u}{\partial t_c} = -1/2$.

$$\begin{aligned} & \left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_6^8 u_c(y + t_c + t_u) f(y) dy + \frac{\partial t_u}{\partial t_c} \int_4^6 u_c(y + y_c + t_u) f(y) dy \\ &= \frac{1}{2} \int_6^8 5(y + 0 + 3)^{-1/2} \frac{1}{4} dy - \frac{1}{2} \int_4^6 5(y + 7 + 3)^{-1/2} \frac{1}{4} dy \\ &= \frac{5}{8} \int_6^8 (y + 3)^{-1/2} dy - \frac{5}{8} \int_4^6 (y + 10)^{-1/2} dy > 0 \end{aligned}$$

Hence, we increase social welfare by increasing t_c , which means that a pure UCT cannot be optimal in this setting. \square

A.7 Proof of Proposition 4

We show if utility is linear in consumption, a pure UCT is optimal.

The Planner's FOC w.r.t. t_c when $u''(c) = 0$ is:

$$\int_Y \int_{\tilde{\mu}(y)}^{\infty} \left(1 + \frac{\partial t_u}{\partial t_c}\right) f(\mu, y) d\mu dy + \frac{\partial t_u}{\partial t_c} \int_Y \int_0^{\tilde{\mu}(y)} f(\mu, y) d\mu dy$$

where

$$\frac{\partial t_u}{\partial t_c} = -\bar{S} + t_c \int_Y \frac{\partial \tilde{\mu}}{\partial t_c} + \frac{\partial \tilde{\mu}}{\partial t_u} \frac{\partial t_u}{\partial t_c} f(\tilde{\mu}|y) f(y) dy$$

where $\bar{S} = \int_Y \int_{\tilde{\mu}(y)}^{\infty} f(\mu, y) d\mu dy$. Since $v(\tilde{\mu}, 1) - y_c + t_c = v(\tilde{\mu}, 0)$, we know $\frac{\partial \tilde{\mu}}{\partial t_u} = 0$ and $\frac{\partial \tilde{\mu}}{\partial t_c} = \frac{-1}{v_1(\tilde{\mu}, 1) - v_1(\tilde{\mu}, 0)} < 0$ as we assume $v_1(\mu, 1) - v_1(\mu, 0) > 0 \forall \mu$, i.e., more able children benefit from schooling more. Denoting $-\frac{\partial \tilde{\mu}}{\partial t_c} f(\tilde{\mu}|y) = P(y) > 0$ and $\int_Y P(y) f(y) dy = \bar{P}$, we can simplify the following expression

$$\frac{\partial t_u}{\partial t_c} = -\bar{S} - t_c \bar{P} < 0$$

Thus our FOC simplifies to

$$\bar{S} (1 - \bar{S} - t_c \bar{P}) + (1 - \bar{S}) (-\bar{S} - t_c \bar{P}) = -t_c \bar{P}$$

Thus our FOC is 0 at $t_c = 0$ and < 0 for $t_c > 0$. Thus, the optimal level of t_c is 0.

A.8 Proof of Proposition 5

Proof. First, let us denote $\mathbb{P}(s^* = 1|y) = \int_{\bar{\mu}(y)} f(\mu|y)d\mu = S(y)$ and $\mathbb{P}(s^* = 1) = \int_Y S(y)f(y)dy = \bar{S}$. Second, note that by Bayes' Theorem, $f(y|s^* = 1) = \frac{f(y)S(y)}{\bar{S}}$ and $f(y|s^* = 0) = \frac{f(y)(1-S(y))}{1-\bar{S}}$. Evaluating the inner integral of Equation (3), we can rewrite the FOC as:

$$\left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y u_c(y + t_u + t_c) S(y) dF(y) + \frac{\partial t_u}{\partial t_c} \int_Y u_c(y + y_c + t_u) (1 - S(y)) dF(y) = 0$$

Multiplying the first term by \bar{S}/\bar{S} and multiplying the second term by $(1 - \bar{S})/(1 - \bar{S})$ we get:

$$\underbrace{\left(1 + \frac{\partial t_u}{\partial t_c}\right) \bar{S} \int_Y u_c(y + t_u + t_c) dF(y|s^* = 1)}_{\text{Targeting Benefit, TB}} + \underbrace{\frac{\partial t_u}{\partial t_c} (1 - \bar{S}) \int_Y u_c(y + y_c + t_u) dF(y|s^* = 0)}_{\text{-Exclusion Cost, EC}} = 0$$

We know t_c^* is optimal under $F^{(1)}$ implying $1 + \frac{\partial t_u}{\partial t_c}(t_c^*, F^{(1)}) > 0$ (as $FOC(t_c^*, f^{(1)}) = 0$ and $\frac{\partial t_u}{\partial t_c}(t_c^*, F^{(1)}) < 0$). Moreover, since we assume the same behavioral responses and average enrollments at t_c^* under the two densities $F^{(1)}, F^{(2)}$, this implies $\frac{\partial t_u}{\partial t_c}(t_c^*, F^{(1)}) = \frac{\partial t_u}{\partial t_c}(t_c^*, F^{(2)})$. Since $F^{(1)}(y|s^* = 1)$ (weakly) $FOSD F^{(2)}(y|s^* = 1)$, $1 + \frac{\partial t_u}{\partial t_c}(t_c^*, f^{(2)}) > 0$, and u_c is decreasing, we get that the $TB(t_c^*, F^{(2)}) \geq TB(t_c^*, F^{(1)})$. Moreover, because $F^{(2)}(y|s^* = 0)$ (weakly) $FOSD F^{(1)}(y|s^* = 0)$, $\frac{\partial t_u}{\partial t_c}(t_c^*, F^{(2)}) < 0$, and u_c is decreasing we also get that $-EC(t_c^*, F^{(2)}) \geq -EC(t_c^*, F^{(1)})$, with one of these two inequalities being strict. Hence, $FOC(t_c^*, F^{(2)}) > 0$, so that increasing t_c from t_c^* increases welfare; hence the optimal local reform is to increase t_c . □

A.9 Borrowing across generations

We adjust our simple model to allow for borrowing across generations. Parents solve the following problem

$$\begin{aligned}
& \max_{s \in \{0,1\}, b} u(c) + \beta v(c_2) \\
& \text{s.t. } c = y + b + (1-s)y_c + t_u + t_c s \\
& \text{and } c_2 = y_2(\mu, s) - rb
\end{aligned}$$

where b denotes the amount parents choose to borrow in the first generation, $r = 1/\beta$ denotes the across generation interest rate, v denotes the utility of consumption in the second generation, $y_2(\mu, s)$ denotes the income of a grown-up child with ability μ that received schooling s in the first generation, and β denotes the across generation discount rate.

We make the following functional form assumptions: $u_{cc} < 0$, $v_{cc} < 0$, $\frac{\partial y_2}{\partial \mu} > 0$, $\frac{\partial y_2}{\partial s} > 0$ and $\frac{\partial^2 y_2}{\partial \mu \partial s} > \frac{-v''}{v'} \frac{\partial y_2}{\partial \mu} \frac{\partial y_2}{\partial s}$. The last assumption implies schooling and ability are sufficiently strong complements.⁴⁰ We can now derive the following proposition.

Proposition 6. *If parents can borrow across generations, a pure UCT is optimal (locally).*

Proof. First, by our assumption $\frac{\partial^2 y_2}{\partial \mu \partial s} > \frac{-v''}{v'} \frac{\partial y_2}{\partial \mu} \frac{\partial y_2}{\partial s}$ we can show $\frac{\partial s^*}{\partial \mu} \geq 0$ by Topkis.⁴¹ This implies for a given parental income level y , there exists a cut-off ability $\tilde{\mu}(y)$ such that a household with parental income y and $\mu \geq \tilde{\mu}(y)$ will send to school (and vice versa). We now differentiate the Planner's problem with respect to t_c to get the following FOC:

$$\left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y \int_{\tilde{\mu}(y)}^{\infty} u_c(c^*(1, y, \mu)) dF(\mu, y) + \frac{\partial t_u}{\partial t_c} \int_Y \int_0^{\tilde{\mu}(y)} u_c(c^*(0, y, \mu)) dF(\mu, y)$$

where $c^*(s, y, \mu)$ is the optimal consumption level of household y, μ if they choose schooling level s . Specifically, $c^*(1; y, \mu) = y + t_c + t_u + b^*(1; y, \mu)$ and $c^*(0, y, \mu) = y + y_c + t_u + b^*(0; y, \mu)$, where $b^*(s; y, \mu)$ denotes the optimal level of borrowing for household y, μ if they choose schooling level s , specifically $b^*(s, y, \mu)$ solves

$$u_c(y + st_c + (1-s)y_c + t_u + b) = v_c(y_2(\mu, s) - rb) \tag{21}$$

Next, we evaluate our FOC at $t_c = 0, t_u = R$ noting that $\frac{\partial t_c}{\partial t_u} \Big|_{t_c=0} = - \int_Y \int_{\tilde{\mu}(y)} dF(y, \mu) =$

40. Increasing ability has two effects on schooling: first, a substitution effect which results in an increase in schooling, and second, an income effect which results in a decrease in schooling. This assumption implies the substitution effect dominates.

41. Denote $W(\mu, b, s)$ the household maximand. We will show that W is supermodular in μ, b, s . First $\frac{\partial^2 W}{\partial \mu \partial s} = \beta v_c(c_2) \frac{\partial^2 y_2}{\partial \mu \partial s} + \beta v_{cc}(c_2) \frac{\partial y_2}{\partial \mu} \frac{\partial y_2}{\partial s} > 0$ by assumption. Second, $\frac{\partial^2 W}{\partial \mu \partial b} = -v_{cc}(c_2) \frac{\partial y_2}{\partial \mu} > 0$ as $v_{cc} < 0$ and $\frac{\partial y_2}{\partial \mu} > 0$. Third, $\frac{\partial^2 W}{\partial b \partial s} = u_{cc}(c)(t_c - y_c) - v_{cc}(c_2) \frac{\partial y_2}{\partial s} > 0$ for $t_c < y_c$.

$-\bar{S}$, where \bar{S} denotes average enrollment. Our FOC becomes:

$$(1 - \bar{S}) \int_Y \int_{\tilde{\mu}(y)} u_c(c^*(1; y, \mu)) dF(\mu, y) dy - \bar{S} \int_Y \int^{\tilde{\mu}(y)} u_c(c^*(0; y, \mu)) dF(\mu, y)$$

Dividing all terms in our FOC by $\bar{S}(1 - \bar{S})$ we get:

$$\frac{1}{\bar{S}} \int_Y \int_{\tilde{\mu}(y)} u_c(c^*(1; y, \mu)) dF(\mu, y) - \frac{1}{1 - \bar{S}} \int_Y \int^{\tilde{\mu}(y)} u_c(c^*(0; y, \mu)) dF(\mu, y) \quad (22)$$

But this is just the difference in the expected marginal utility between those who send to school and those who do not. This difference will be negative meaning $t_c = 0, t_u = R$ is locally optimal. Why is this term negative? First, we can show that if a household chooses to send their child to school (i.e., $s^* = 1$), it must be the case that $c^*(1, y, \mu) > c^*(0, y, \mu)$. Why? A household will send to school iff:

$$u(c^*(1, y, \mu)) + v(c_2^*(1, y, \mu)) \geq u(c^*(0, y, \mu)) + v(c_2^*(0, y, \mu)) \quad (23)$$

where $c_2^*(s, y, \mu) = y_2(\mu, s) - rb^*(s, y, \mu)$. Because we know $b^*(s, y, \mu)$ satisfies $u_c(c^*(s, y, \mu)) = v_c(c_2^*(s, y, \mu))$, it must be the case that if a household chooses to send to school then $c^*(1, y, \mu) \geq c^*(0, y, \mu)$. Suppose not, i.e., $c^*(1, y, \mu) < c^*(0, y, \mu)$ and $s^* = 1$. By Equation (21), it must be the case that $c_2^*(1, y, \mu) < c_2^*(0, y, \mu)$ (as $u_{cc} < 0, v_{cc} < 0$). But then Equation (23) does not hold, meaning it could not have been optimal to send to school. By similar arguments, if a household chooses to not send to school, it must be the case that $c^*(1, y, \mu) < c^*(0, y, \mu)$, and if a household is indifferent between schooling or not, $c^*(1, y, \mu) = c^*(0, y, \mu)$. By this last statement, we therefore know $c^*(1, y, \tilde{\mu}(y)) = c^*(0, y, \tilde{\mu}(y)) \forall y$.

Second, we can show that $c^*(1, y, \mu)$ and $c^*(0, y, \mu)$ are increasing in μ . Why? $\frac{\partial c^*(s, y, \mu)}{\partial \mu} = \frac{\partial b^*(s, y, \mu)}{\partial \mu} = \frac{v_{cc}(c_2^*(s, y, \mu)) \frac{\partial y_2(\mu, s)}{\partial \mu}}{u_{cc}(c^*(s, y, \mu)) + rv_{cc}(c_2^*(s, y, \mu))} > 0$ as $v_{cc} < 0, u_{cc} < 0, \frac{\partial y_2(\mu, s)}{\partial \mu} > 0$. Thus, combining the fact that $c^*(1, y, \tilde{\mu}(y)) = c^*(0, y, \tilde{\mu}(y)) \forall y$, $\frac{\partial c^*(1, y, \mu)}{\partial \mu} > 0$, and $\frac{\partial c^*(0, y, \mu)}{\partial \mu} > 0$ it must be the case Equation (22) is negative, meaning that $t_c = 0, t_u = R$ is locally optimal. \square

A.10 Borrowing within a generation

A seemingly more realistic possibility is that parents are unable to borrow against their child's future income, yet are able to borrow against their own future income, in

say a year's time. For example, perhaps a parent is thinking about sending their child to a year of secondary school, but they are able to borrow against their income next year (in which they won't have to bear costs of education). To consider how borrowing within a generation affects the targeting benefit of CCTs, we need to adjust our model to allow for multiple periods within a generation. We consider the following simple extension:

$$\begin{aligned} & \max_{s \in \{0,1\}, b} u(c_1) + \delta u(c_2) + v(\mu, s) \\ \text{s.t. } & c_1 = y + b + (1 - s)y_c + t_u + t_c s \\ & \text{and } c_2 = y - rb \end{aligned}$$

where δ denotes the within generation discount rate, c_1 denotes consumption in the first period of the first generation when parents must decide whether to send to school or not, c_2 denotes consumption in the second period of the first generation when parents do not have to make a schooling decision, and $v(\mu, s)$ denotes household utility in the second generation. b denotes the amount households borrow and $r = 1/\delta$ denotes the interest rate. Without loss of generality, we could have chosen to model the schooling decision in the second period of the first generation and households choosing how much to save during the first period.

It's simple to show that with borrowing, households smooth consumption across the first two periods of the first generation: $c_1 = c_2$. As such, we can rewrite the above problem as a maximization problem over s :

$$\begin{aligned} & \max_{s \in \{0,1\}} u(c_1)(1 + \delta) + v(\mu, s) \\ \text{s.t. } & c_1 = y + \frac{1}{1 + \frac{1}{r}}(t_u + t_c s + (1 - s)y_c) \end{aligned}$$

Let's redefine the following: $t'_u = t_u \frac{1}{1+\delta}$; $t'_c = t_c \frac{1}{1+\delta}$; $y'_c = y_c \frac{1}{1+\delta}$; $R' = R \frac{1}{1+\delta}$; $v'(\mu, s) = v(\mu, s) \frac{1}{1+\delta}$ noting that $\frac{1}{1+\delta} = \frac{1}{1+\frac{1}{r}}$ as $r = 1/\delta$. The planner's problem is now

$$\begin{aligned} W &= \max_{t'_u, t'_c} \int_Y \int_M u(c_1^*(y, \mu)) + v'(\mu, s^*) f(\mu, y) d\mu dy \\ \text{s.t. } & t'_u \int_Y f(y) dy + t'_c \int_Y \int_M s^* f(\mu, y) d\mu dy \leq R' \\ & c_1^* = y + (1 - s^*)y'_c + t'_u + t'_c s^* \end{aligned}$$

This is functionally equivalent to the original problem without any borrowing. As

such, the Proof of Proposition 3 still holds in this world, so that it can be optimal to have a CCT.

A.11 Considering labor supply

In the above model we assume parents are endowed with income y . We now relax this assumption and include labor supply decisions. We are able to show that it can still be beneficial to allocate some of the budget towards a CCT. We keep the social planner's problem the same as in Section III, but now we consider a modified household problem with labor supply decisions:

$$\begin{aligned} \max_{s \in \{0,1\}, l} \quad & u(c, l) + v(\mu, s) \\ \text{s.t.} \quad & c = nl + (1 - s)y_c + t_u + t_c s \end{aligned}$$

where l denotes the labor supply choice of parents in the first generation, n denotes the heterogeneous productivity of parents, c denotes consumption in the first generation, and $v(\mu, s)$ denotes utility in the second generation. For simplicity we keep child income constant at y_c . We assume $F(\mu|n)$ FOSD $F(\mu|n')$ for $n > n'$, i.e., parent productivity and child ability are positively correlated. Finally we assume $u_c > 0, u_{cc} < 0, u_l < 0, u_{ll} \leq 0, u_{cl} = 0$.

We first show that schooling is weakly increasing in parent productivity n , holding child ability μ constant:

Lemma 7. *Schooling, $s^*(n, \mu)$, is weakly increasing in parent productivity n , so that there exists a cutoff \tilde{n} such that those parents with $n < \tilde{n}$ do not send to school and those with $n \geq \tilde{n}$ send to school.*

Proof. First, we rewrite the problem slightly using a change of variables $y = nl$:

$$\begin{aligned} \max_{s \in \{0,1\}, y} \quad & u(c, y/n) + v(\mu, s) \\ \text{s.t.} \quad & c = y + (1 - s)y_c + t_u + t_c s \end{aligned}$$

We show that the problem has increasing differences in (s, y, n) . We simply check all three cross partial derivatives of $f(s, y, n) = u(c, y/n) + v(\mu, s)$ are weakly positive, remembering that $u_{cl} = 0$. First, it's easy to see that $\frac{\partial^2 f}{\partial n \partial s} = 0$.

$$\frac{\partial^2 f}{\partial y \partial s} = u_{cc}(c, y/n)(-y_c + t_c) \geq 0$$

$$\frac{\partial^2 f}{\partial y \partial n} = -u_l(c, y/n) \frac{1}{n^2} - u_{ll}(c, y/n) \frac{y}{n^3} \geq 0$$

Thus, we have increasing differences in (s, y, n) , so by Topkis' Theorem, we know that s is increasing in n . \square

Next, we show that there still exists a jump in marginal utility of consumption around the parent who is indifferent between sending to school and not (holding child ability, μ , constant)

Lemma 8. *There still exists a discontinuity in household income $z(n) \equiv y(n) + (1 - s(n))y_c$, where $y(n) = nl(n)$. In particular, around \tilde{n} , we have that:*

$$\lim_{n \rightarrow \tilde{n}^-} z(n) > \lim_{n \rightarrow \tilde{n}^+} z(n)$$

Proof. Suppose not, so that $z^-(n) \leq z^+(n)$, with $z^-(n) = y^-(n) + y_c$ and $z^+(n) = y^+(n)$. Thus, $y^-(n) < y^+(n)$ as $y_c > 0$. This implies $u_c(z^-(n), l^-(n)) \geq u_c(z^+(n), l^+(n))$ as $u_{cc} < 0, u_{cl} = 0$. By the first order condition we know $u_c(z(n), l(n))n = -u_l(z(n), l(n))$. Thus, $-u_l(z^-(n), l^-(n)) \geq -u_l(z^+(n), l^+(n))$. Since $u_{ll} \leq 0, u_{cl} = 0$, this implies $l^-(n) \geq l^+(n)$. This implies $y^-(n)/n \geq y^+(n)/n$, which contradicts $y^-(n) < y^+(n)$. \square

Lemmas 7 and 8 imply that, for a given child ability type μ , marginal utility of consumption is decreasing in parent productivity n until \tilde{n} where there exists a discrete jump up in marginal utility of amount $u_c(\tilde{n}l^*(1)) - u_c(\tilde{n}l^*(0) + y_c)$ (where $l^*(1)$ denotes optimal labor supply if household \tilde{n} sends to school, and $l^*(0)$ denotes optimal labor supply if household \tilde{n} does not send to school). Thus, Figure II is still relevant when we consider labor supply. Essentially, households do not adjust their labor supply fully to offset the discrete cost of schooling, thus, creating a jump in marginal utility of consumption between those just indifferent between sending to school and not. Finally, we show the following proposition

Proposition 9. *If utility is concave in consumption and parents can freely adjust their labor supply, a pure UCT is not necessarily optimal.*

Proof. For simplicity, let's assume away μ heterogeneity. The Planner's FOC can be

written as follows:

$$\int_{\tilde{n}} u_c(nl(n, 1) + t_u + t_c - k, l(n, 1)) \left(1 + \frac{\partial t_u}{\partial t_c}\right) f(n) dn + \int_{\tilde{n}} u_c(nl(n, 0) + t_u, l(n, 0)) \frac{\partial t_u}{\partial t_c} f(n) dn$$

Suppose that $u(c, l) = 10c^{1/2} - l^2/2$ and $\mu = 10$. Further suppose $v(10, 1) = 10$ and $v(10, 0) = 0$. Further, suppose $R = 3$, $y_c = 7$. Given this, the indifferent type is $\tilde{n} \approx 2.39$ for $t_u = 3, t_c = 0$. Finally, suppose $f(n) \sim Unif[\tilde{n} - 0.5, \tilde{n} + 0.5]$. Consider the FOC of the Planner's problem with respect to t_c starting from $t_u = 3, t_c = 0$ noting that $\frac{\partial t_u}{\partial t_c} = -0.5$:

$$\int_{\tilde{n}}^{\tilde{n}+0.5} \frac{5}{2} (nl(n, 1) + 3)^{-1/2} dn - \int_{\tilde{n}-0.5}^{\tilde{n}} \frac{5}{2} (nl(n, 0) + 3 + 7)^{-1/2} dn = 0.23 > 0$$

Hence, we increase social welfare by increasing t_c , which means that a pure UCT cannot be optimal in this setting. \square

A.12 Transfers in multiple generations

We now consider a world where in each generation households consist of a parent and a child, and parents must decide whether to send their child to school or not. If parents send to school, children earn more in the following generation when they are parents; however, sending a child to school results in a discrete loss of household consumption as parents forgo child income y_c (assumed to be constant across generations). We assume all parents with incomes less than some threshold \bar{y} are eligible to receive conditional and unconditional cash transfers. Households solve the following problem:

$$\max_{s_1 \in \{0,1\}, s_2 \in \{0,1\}} u\left(y_1 + t_{u1} + t_{c1}s_1 + (1 - s_1)y_c\right) + \beta u\left(y_2(s_1, \mu) + (t_{u2} + t_{c2}s_2)\mathbb{1}(y_2(s_1, \mu) \leq \bar{y}) + (1 - s_2)y_c\right) + \beta^2 V(s_2, \mu)$$

where s_1 denotes whether a parent sends their child to school in generation 1; s_2 denotes whether a parent sends their child to school in generation 2; y_1 denotes parent income in generation 1 (where $y_1 < \bar{y}$, i.e., we restrict to parents who are below the eligibility threshold in generation 1); $y_2(s_1, \mu)$ denotes parent income in generation 2 (increasing in both schooling and ability); t_{ci}, t_{ui} denote the conditional and unconditional

transfers offered in generation i ; and, $V(s_2, \mu)$ denotes utility in generation 3 which depends on whether the child in generation 2 went to school and this child's ability (note, for simplicity, ability is assumed constant across generations within the same household). Further, for simplicity, we assume that if a child went to school in generation 1, their income as an adult surpasses the eligibility threshold, i.e., $y_2(1, \mu) > \bar{y} \forall \mu$, and if a child did not go to school in generation 1, their income as an adult does not surpass the eligibility threshold, i.e., $y_2(0, \mu) \leq \bar{y} \forall \mu$.

Denote $\tilde{\mu}_2^{(1)}$ as the parent who is just indifferent between sending to school or not in generation 2 who went to school in generation 1:

$$u\left(y_2(1, \tilde{\mu}_2^{(1)}) + y_c\right) + \beta V(0, \tilde{\mu}_2^{(1)}) = u\left(y_2(1, \tilde{\mu}_2^{(1)})\right) + \beta V(1, \tilde{\mu}_2^{(1)})$$

Denote $\tilde{\mu}_2^{(0)}$ as the parent who is just indifferent between sending to school or not in generation 2 who did not go to school in generation 1:

$$u\left(y_2(0, \tilde{\mu}_2^{(0)}) + y_c + t_{u2}\right) + \beta V(0, \tilde{\mu}_2^{(0)}) = u\left(y_2(0, \tilde{\mu}_2^{(0)}) + t_{u2} + t_{c2}\right) + \beta V(1, \tilde{\mu}_2^{(0)})$$

Finally, denote $\tilde{\mu}_1$ as the indifferent household in generation 1:

$$\begin{aligned} u\left(y_1 + t_{u1} + t_{c1}\right) + \beta u\left(y_2(1, \tilde{\mu}_1) + (1 - s_2^{*(1)})y_c\right) + \beta^2 V(s_2^{*(1)}, \tilde{\mu}) = \\ u\left(y_1 + t_{u1} + y_c\right) + \beta u\left(y_2(0, \tilde{\mu}_1) + t_{u2} + s_2^{*(0)}t_{c2} + (1 - s_2^{*(0)})y_c\right) + \beta^2 V(s_2^{*(0)}, \tilde{\mu}_1) \end{aligned}$$

where $s_2^{*(1)}$ ($s_2^{*(0)}$) denotes optimal schooling decision in generation 2 if the parent went to school (did not go to school) in generation 1.

The planner has a budget R in each generation and can offer a conditional and unconditional cash transfer to households in each generation. The planner's objective is to maximize total lifetime utility:

$$\begin{aligned}
& \max_{t_{c1}, t_{u1}, t_{c2}, t_{u2}} \int_{Y_1} \int_{\tilde{\mu}_2^{(1)}}^{\infty} u(y_1 + t_{u1} + t_{c1}) + \beta u(y_2(1, \mu)) + \beta^2 V(1, \mu) dF(y_1, \mu) + \\
& \int_{Y_1} \int_{\tilde{\mu}_1}^{\tilde{\mu}_2^{(1)}} u(y_1 + t_{u1} + t_{c1}) + \beta u(y_2(1, \mu) + y_c) + \beta^2 V(0, \mu) dF(y_1, \mu) + \\
& \int_{Y_1} \int_{\tilde{\mu}_2^{(0)}}^{\tilde{\mu}_1} u(y_1 + y_c) + \beta u(y_2(0, \mu) + t_{u2} + t_{c2}) + \beta^2 V(1, \mu) dF(y_1, \mu) + \\
& \int_{Y_1} \int_0^{\tilde{\mu}_2^{(0)}} u(y_1 + y_c) + \beta u(y_2(0, \mu) + t_{u2} + y_c) + \beta^2 V(0, \mu) dF(y_1, \mu) \\
& \text{s.t. } t_{u1} + t_{c1} \int_{Y_1} \int_{\tilde{\mu}_1}^{\infty} dF(y_1, \mu) \leq R \\
& \text{s.t. } t_{u2} \int_{Y_1} \int_0^{\tilde{\mu}_1} dF(Y_1, \mu) + t_{c2} \int_{Y_1} \int_{\tilde{\mu}_2^{(0)}}^{\tilde{\mu}_1} dF(y_1, \mu) \leq R
\end{aligned}$$

where for simplicity we assume $\tilde{\mu}_2^{(0)} < \tilde{\mu}_1 < \tilde{\mu}_2^{(1)}$. The planner's first order condition w.r.t. t_{c1} is given by:

$$\begin{aligned}
& \underbrace{\left(1 + \frac{\partial t_{u1}}{\partial t_{c1}}\right) \int_{Y_1} \int_{\tilde{\mu}_1}^{\infty} u_c(y_1 + t_{u1} + t_{c1}) dF(y_1, \mu)}_{\text{Targeting Benefit}} + \underbrace{\frac{\partial t_{u1}}{\partial t_{c1}} \int_{Y_1} \int_{\tilde{\mu}_2^{(0)}}^{\tilde{\mu}_1} u_c(y_1 + t_{u1} + y_c) dF(y_1, \mu)}_{\text{-Exclusion Cost}} \\
& - \lambda_2 \underbrace{\left((t_{u2} + t_{c2}) \int_{Y_1} \left(\frac{\partial \tilde{\mu}_1}{\partial t_{c1}} + \frac{\partial \tilde{\mu}_1}{\partial t_{u1}} \frac{\partial t_{u1}}{\partial t_{c1}} \right) f(\tilde{\mu}_1) dF(y_1) \right)}_{\text{Budgetary Benefit}} = 0
\end{aligned}$$

Note, the targeting benefit and exclusion cost are identical to Equation (3); however, there is now another positive term in the planner's first order condition which we term the budgetary benefit (where λ_2 denotes the lagrange multiplier on the planner's budget constraint in generation 2). This term captures the fact that increasing t_{c1} reduces the number of eligible beneficiaries for transfers tomorrow, thus increasing the size of the transfers offered to households tomorrow. This term is positive (assuming $\frac{\partial t_{c1}}{\partial t_{u1}} \geq -1$).

Lastly, it is easy to show that the average total transfers received by households who send their child to school in generation 1 is less than the average total transfers received by households who do not send their child to school in generation 1 as $t_{u1} + t_{c1} < t_{u1} + R/(1 - S_1)$ (where S_1 denotes the share enrolled in the first generation). Thus, once we consider welfare programs offered in future generations, it is easy to mitigate concerns that we are on net transferring more to higher lifetime utility households.

A.13 Proving Proposition 3 for all Continuous Social Welfare Functions

In proving Proposition 3, we assumed the planner was a utilitarian. Because the planner only has money to give out in the first generation and because parents cannot move resources across generations, the planner's utilitarian objective translates into an objective to redistribute towards those with the highest marginal utility of consumption in the first generation. However, now suppose the planner also cares about directing money towards those with the lowest lifetime utility (i.e., the lowest utility over the two generations combined). We now proceed to show that Proposition 3 is not a consequence of the utilitarian social welfare function; rather, we can show that a pure UCT is not necessarily optimal under a large class of social welfare functions.

Corollary 9.1. *If utility is concave in consumption, a pure UCT may not be optimal under any continuous social welfare function $\int_Y \int_M G(u(c^*) + v(\mu, s^*)) f(\mu, y) d\mu dy$ with $G(\cdot) > 0$.*

Proof. As in Proposition 3, suppose that $u(c) = 10c^{1/2}$, assume all individuals have $\mu = 10$, $v(10, 1) = 10$ and $v(10, 0) = 0$, and $R = 3$, $y_c = 7$. Now, however, $f(y) \sim Unif[6 - \epsilon, 6 + \epsilon]$. It's still true that parents with income above 6 send their children to school when $t_u = R$, $t_c = 0$. Consider the FOC of the Planner's problem with respect to t_c (when we substitute the budget constraint to write $t_u(t_c)$) when $t_u = R$, $t_c = 0$, noting that $\frac{\partial t_u}{\partial t_c} = -1/2$. Using the shorthand $G(y) \equiv G(u(c^*(y)) + v(10, s^*(y)))$, we can write the planner's FOC as:

$$\begin{aligned} & \int_6^{6+\epsilon} G'(y) u_c(y + t_u) \left(1 + \frac{\partial t_u}{\partial t_c}\right) f(y) dy + \int_{6-\epsilon}^6 G'(y) u_c(y + y_c + t_u) \frac{\partial t_u}{\partial t_c} f(y) dy \\ &= \int_6^{6+\epsilon} G'(y) (y + 3)^{-1/2} \frac{5}{2} \frac{1}{2\epsilon} dy - \int_{6-\epsilon}^6 G'(y) (y + 7 + 3)^{-1/2} \frac{5}{2} \frac{1}{2\epsilon} dy \end{aligned}$$

Suppose $\epsilon < 1$. Then $\forall \delta \in [0, \epsilon]$, as long as:

$$\frac{G'(6 - \delta)}{G'(6 + \delta)} < \frac{(6 - \delta + 10)^{1/2}}{(6 + \delta + 3)^{1/2}} < (16/9)^{1/2} \quad (24)$$

we will have:

$$G'(6 + \delta)(6 + \delta + 3)^{-1/2} - G'(6 - \delta)(6 - \delta + 10)^{-1/2} > 0$$

Inequality (24) will be true for sufficiently small δ by continuity of G' (as $\frac{G'(6-\delta)}{G'(6+\delta)} \approx 1$, while $\frac{(6-\delta+10)^{1/2}}{(6+\delta+3)^{1/2}} \approx (16/9)^{1/2}$). This means that for sufficiently small ϵ , the planner's FOC

will be positive at $t_c = 0$:

$$\begin{aligned} & \int_6^{6+\epsilon} G'(y)(y+3)^{-1/2} \frac{5}{2} \frac{1}{2\epsilon} dy - \int_{6-\epsilon}^6 G'(y)(y+10)^{-1/2} \frac{5}{2} \frac{1}{2\epsilon} dy \\ &= \int_0^\epsilon \left(G'(6+\delta)(6+\delta+3)^{-1/2} \frac{5}{2} \frac{1}{2\epsilon} - G'(6-\delta)(6-\delta+10)^{-1/2} \frac{5}{2} \frac{1}{2\epsilon} \right) d\delta > 0 \end{aligned}$$

Hence, this FOC is positive for sufficiently small ϵ , so we can increase social welfare by increasing t_c , which means that a pure UCT is again suboptimal. □

A.14 Annual Schooling Decisions Model

In this subsection we consider the extension where parents make multiple annual schooling decisions over T years of their child's life. We denote the cost of schooling in each year t as y_{ct} . For simplicity we assume parental income is constant across the years, i.e., $y_t = y \forall t \in 1, 2, \dots, T$, and that there is no discounting over the T years.

Parent Problem. We begin with the parent problem:

$$\max_{\{s_t \in \{0,1\}\}_{t=1}^T} \sum_{t=1}^T u(y + y_{ct}(1 - s_t) + t_c s_t + t_u) + V\left(\sum_{t=1}^T s_t, \mu\right)$$

where $V(t+1, \mu) > V(t, \mu) \forall t \in 0, \dots, T-1$, $V_2(t, \mu) > 0$, and $V_2(t+1, \mu) > V_2(t, \mu)$, i.e., child ability and total years at school are complements. To make the planner's problem tractable, we have imposed that both the conditional and unconditional transfers are constant across the T years. Finally, we assume child income is increasing in child age, i.e. $y_{ct+1} > y_{ct}$. With this assumption, we can then make the following claim:

Claim 10. *If $s_t^* = 0$ then $s_{t+n}^* = 0 \forall n \in 1, \dots, T-t$. In words, if you stop sending to school for a year, you stop sending to school for all remaining years.*

Proof. Suppose not, i.e. $s_t^* = 0$, $s_{t+n}^* = 1$ for some $n \in 1, \dots, T-t$. This implies that for some year m , we have the following: $s_m^* = 0$, $s_{m+1}^* = 1$. Total utility from $s_m^* = 0$, $s_{m+1}^* = 1$ is given by:

$$\sum_{t=1}^{m-1} u(y + y_{ct}(1 - s_t^*) + t_u + t_c s_t^*) + u(y + y_{cm} + t_u) + u(y + t_u + t_c) + \sum_{t=m+2}^T u(y + y_{ct}(1 - s_t^*) + t_c s_t^* + t_u) + V\left(\sum_{t=m+2}^T s_t^* + \sum_{t=1}^{m+1} s_t^*, \mu\right)$$

We will now show $s_t = s_t^* \forall i = 1, \dots, m-1, m+2, \dots, T$, $s_m = 1$, $s_{m+1} = 0$ generates higher parental utility. Total utility from $s_m^* = 1$, $s_{m+1}^* = 0$ is given by:

$$\sum_{t=1}^{m-1} u(y + y_{ct}(1 - s_t^*) + t_u + t_c s_t^*) + u(y + t_u + t_c) + u(y + t_u + y_{cm+1}) + \sum_{t=m+2}^T u(y + y_{ct}(1 - s_t^*) + t_c s_t^* + t_u) + V\left(\sum_{t=m+2}^T s_t^* + \sum_{t=1}^{m+1} s_t^*, \mu\right)$$

By assumption $y_{cm+1} > y_{cm}$, hence, utility is higher under $s_m^* = 1$, $s_{m+1}^* = 0$. Thus, $s_m^* = 0$, $s_{m+1}^* = 1$ could not have been optimal. \square

Thus, we can rewrite the parent problem as an optimal stopping problem as follows:

$$\max_{m \in \{1, \dots, T+1\}} \sum_{t=1}^{m-1} u(y + t_u + t_c) + \sum_{t=m}^T u(y + t_u + y_{ct}) + V(m-1, \mu)$$

where parents now choose the year m to stop sending their child to school. Finally, let $\tilde{\mu}_m$ denote the household who is indifferent between stopping in year m or $m+1$ (note, $\tilde{\mu}_m$ will be a function of parent income, the transfer schedule, and child income in year m):

$$u(y + t_u + t_c) + V(m, \tilde{\mu}_m) = u(y + t_u + y_{cm}) + V(m-1, \tilde{\mu}_m)$$

Because we assume ability and total years of schooling are complements, we know that all households with $\mu > \tilde{\mu}_m$ will strictly prefer stopping in year $m+1$ to stopping in year m , and vice versa. Moreover, because we assume V has decreasing differences in total years of schooling, we know that if $\mu \geq \tilde{\mu}_m \implies \mu > \tilde{\mu}_i$ for $i \in 1, \dots, m-1$ and $\mu \leq \tilde{\mu}_m \implies \mu < \tilde{\mu}_i$ for $i \in m+1, \dots, T$.⁴² Thus, the set of households sending to school

42. If $\mu > \tilde{\mu}_{m+1} \implies u(y + t_u + t_c) + V(m+1, \mu) > u(y + t_u + y_{cm+1}) + V(m, \mu)$. Rearranging gives

in year m will have $\mu \geq \tilde{\mu}_m$. We now describe the planner's problem.

Planner's Problem. We assume the planner has a budget R to give the next B cohorts where we define a cohort to be the year in which a child can start school, i.e., the planner will offer transfers to the set of children who can start school in years $b_0, \dots, b_0 + B$, and will offer each child in these cohorts transfers for their full T years of schooling. Thus, the planner will offer transfers from years $b_0, \dots, T + b_0 + B$. The planner's objective is to maximize total parental utility of all parents who will have a child in the next B cohorts.

Before we write the planner's problem, it is first useful to re-write the parents' optimal stopping problem explicitly in terms of calendar years and child cohort, i.e., parents of a child who starts school in year b get to pick a year $m \in \{b, \dots, T + b\}$ in which they stop sending their child to school:

$$\max_{m \in \{b, \dots, T+b\}} \sum_{t=b}^{m-1} u(y + t_u + t_c) + \sum_{t=m}^{T+b} u(y + t_u + y_c(t-b)) + V(m-b, \mu)$$

where child income is a function of child age which can be determined by $t-b$, e.g., if children start school at age 6 and they can start school in year b (i.e., they are in cohort b), their age in year t is given by $t-b+6$. Now let $\tilde{\mu}(y, m-b)$ denote the household (earning parental income y) who is indifferent between stopping in year m or in year $m+1$:

$$u(y + t_u + t_c) + V(m-b+1, \tilde{\mu}) = u(y + t_u + y_c(m-b)) + V(m-b, \tilde{\mu})$$

By the same logic as before, households with $\mu \geq \tilde{\mu}(y, m-b)$ will send to school in year m and vice versa. Again, for ease of notation we have omitted that $\tilde{\mu}$ is also a function of the transfer schedule (t_u, t_c) . We write the planner's problem as follows⁴³

$u(y + t_u + y_{cm+1}) - u(y + t_u + t_c) < V(m+1, \mu) - V(m, \mu)$. But $u(y + t_u + y_{cm}) - u(y + t_u + t_c) < u(y + t_u + y_{cm+1}) - u(y + t_u + t_c) < V(m+1, \mu) - V(m, \mu) < V(m, \mu) - V(m-1, \mu)$ where the first inequality comes from our assumption that $y_{cm} < y_{cm+1}$ and the last inequality comes from our decreasing differences assumption. Hence $u(y + t_u + y_{cm}) - u(y + t_u + t_c) < V(m, \mu) - V(m-1, \mu) \implies \mu > \tilde{\mu}_m$. But if $\mu > \tilde{\mu}_m$, by the same logic, $\mu > \tilde{\mu}_{m-1}$ etc. Hence, if $\mu \geq \tilde{\mu}_{m+1} \implies \mu > \tilde{\mu}_i$ for $i \in 1, \dots, m$. One can easily repeat a similar exercise to show $\mu \leq \tilde{\mu}_{m+1} \implies \mu < \tilde{\mu}_i$ for $i \in m+2, \dots, T$.

43. Note, we have assumed utility of parents is constant when their child is not of school-going age and have therefore are only concerned with maximizing utility of parents when they are making schooling decisions.

$$\begin{aligned}
& \max_{t_c, t_u} \sum_{b=b_0}^{b_0+B} \sum_{t=b}^{T+b} \left(\int_Y \int_{\tilde{\mu}(y, t-b)} u(y + t_u + t_c) + V(t - b + 1, \mu) - V(t - b, \mu) dF(\mu, y, t, b) + \right. \\
& \quad \left. \int_Y \int_{\tilde{\mu}(y, t-b)} u(y + t_u + y_c(t - b)) dF(\mu, y, t, b) \right) \\
& \text{s.t. } t_u + t_c \sum_{b=b_0}^{b_0+B} \sum_{t=b}^{T+b} \int_Y \int_{\tilde{\mu}(y, t-b)} dF(\mu, y, t, b) \leq R
\end{aligned}$$

where $F(\mu, y, t, b)$ denotes the joint CDF of child abilities, parental incomes, cohorts, and calendar years, i.e., $\sum_{b=b_0}^{b_0+B} \sum_{t=b}^{T+b} \int_Y \int_M dF(y, \mu, t, b) = 1$. Taking the FOC w.r.t. t_c we get:

$$\begin{aligned}
& \sum_{b=b_0}^{b_0+B} \sum_{t=b}^{T+b} \left(\left(1 + \frac{\partial t_u}{\partial t_c} \right) \int_Y u_c(y + t_u + t_c) S(y, t, b) dF(y, t, b) + \right. \\
& \quad \left. \frac{\partial t_u}{\partial t_c} \int_Y u_c(y + t_u + y_c(t - b)) (1 - S(y, t, b)) dF(y, t, b) \right) = 0
\end{aligned} \tag{25}$$

where $S(y, t, b) = \int_{\tilde{\mu}(y, t-b)} dF(\mu|y, t, b)$ denotes the share of households with parental income y and a child in cohort b sending their child to school in year t (note this share is also a function of the transfer schedule as $\tilde{\mu}$ is a function of the transfer schedule), and where $\frac{\partial t_c}{\partial t_u}$ is given by:

$$\frac{\partial t_u}{\partial t_c} = - \frac{t_c \bar{S} + t_c \sum_b \sum_t \int_Y \frac{\partial S(y, b, t)}{\partial t_c} dF(y, b, t)}{1 + t_c \sum_b \sum_t \int_Y \frac{\partial S(y, b, t)}{\partial t_u} dF(y, b, t)}$$

where \bar{S} denotes average enrollment over all cohorts, all years, and all parental income levels: $\sum_b \sum_t \int_Y S(y, t, b) dF(y, t, b)$. Equation (25) is simply the average targeting benefit across all cohorts and years plus the average exclusion cost across all cohorts and years.

B APPENDIX

B.1 First Stage Regressions: IV Estimation of Equation (10)

	(1)
hourly wage (town)	2.021*** (0.355)
$\mathbb{E}[y^f]$	27.19
Observations	17847
R-squared	0.0126

Note: Standard errors are clustered at the locality level and are presented in parentheses. Dependent: child i 's father's weekly income (in pesos, per-capita), y_{it}^f . Sample: unbalanced panel of children aged 12-15 years in 1997-1999 from eligible (poor) households with two parents and one parent reports to be the head of the household. Mean hourly wage in locality is calculated as the average wage in a locality in year t excluding the wage of child i 's father. Regressors not shown: child age dummies, highest grade complete dummies, age interacted with highest grade completed, year dummies, and a child fixed effect.

B.2 Calculating Moments of Income and Price Effects

To calculate the average price effect (for example), we do the following:

$$\mathbb{E}[P(X)] = \frac{1}{N} \sum_{it=1}^N \hat{b}_1 + \hat{b}_3 y_{it} + 2\hat{b}_5 t_{c,it} + \hat{b}_7 t_{u,it}$$

where \hat{b}_j denotes the j^{th} estimated coefficient from Regression (12), and where N denotes the number of eligible children living in treatment localities in years 1998 and 1999 (i.e., we average over the set of eligible, treated children for years 1998 and 1999).

What about second-order moments of the income and price effect schedules (used in Equation (13))? Well, to calculate the variance of the price effect, we do the following:

$$\begin{aligned} \text{Var}(P(X)) &= \tilde{b}_3^2 \text{Var}(y_{it}) + 4\tilde{b}_5^2 \text{Var}(t_{c,it}) + \tilde{b}_7^2 \text{Var}(t_{u,it}) + 4\widetilde{b_3 b_5} \text{Cov}(y_{it}, t_{c,it}) + 4\widetilde{b_5 b_7} \text{Cov}(t_{c,it}, t_{u,it}) \\ &\quad + 2\widetilde{b_3 b_7} \text{Cov}(y_{it}, t_{u,it}) \end{aligned}$$

where

$$\tilde{b}_j = \mathbb{E}[\hat{b}_j^2] - \text{Var}(\hat{b}_j)$$

$$\widetilde{b_j b_m} = \mathbb{E}[\hat{b}_j \hat{b}_m] - \text{Cov}(\hat{b}_j, \hat{b}_m)$$

etc. Note, we do this adjustment because $\mathbb{E}[\hat{b}_j^2] \neq b_j^2$ (i.e., \hat{b}_j^2 is a biased estimate for b_j^2). To obtain the various moments of the \hat{b} 's, we bootstrap.

B.3 Predicting Child Incomes

Following a similar specification to Attanasio et al (2011), we estimate the following Mincer regression for all individuals aged 12-16 years over the three survey years (1997, 1998, 1999):

$$\begin{aligned} \log(hw_{it}) = & \alpha_1 + \alpha_2 age_{it} + \alpha_3 hgc_{it} + \alpha_4 boy_i + \alpha_5 \log(town_hw_{lt}) + \alpha_6 eligible_i + \alpha_7 control_l + \\ & \alpha_8 \mathbb{1}(year = 98)_t + \alpha_9 \mathbb{1}(year = 99)_t + \alpha_{10} (control_l \times \mathbb{1}(year = 98)_t) + \\ & \alpha_{11} (control_l \times \mathbb{1}(year = 99)_t) + e_{it} \end{aligned} \tag{26}$$

where hw_{it} denotes the (inflation-adjusted) hourly wage of child i in locality l in year t ; age_{it} and hgc_{it} denote age and highest grade completed of child i in year t , respectively; boy_i takes value 1 if child i is a boy; $\log(town_hw_{lt})$ denotes the log of the median wage in locality l at time t ; $eligible_i$ denotes whether child i is eligible for Progresa grants (i.e., they are from a poor household); and $control_l$ denotes whether child i lives in a control locality. Results are presented in Table XII. Our results are similar to those of Attanasio et al (2011) (see their Equation (9)). Like Attanasio et al (2011), we find a strong and significant child age effect, a strong and significant locality-level wage effect, and a small and insignificant education effect (likely reflecting the limited types of jobs available in these villages).

TABLE XII: MINCER REGRESSION

	(1)
age	0.0440*** (0.007)
hgc	0.00109 (0.004)
boy	0.102*** (0.025)
log(town wage)	0.830*** (0.047)
Observations	4586
R-squared	0.213

Note: Standard errors are clustered at the locality level and are presented in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: log hourly wage. Estimated on all individuals aged 12-16 years reporting positive hourly wages for years 1997, 1998, and 1999. Regressors not shown: year dummies, dummy for living in a control area, dummy for being an eligible household, control interacted with year dummies.

Using Equation (26), we estimate potential child earnings, $y_{c,it}$, for all children in our

sample. To do so, we assume all children would work 40 hours per-week if they worked.⁴⁴ Average potential incomes for 12-15 year-old boys and girls in 1997 are presented in Table XIII below.

TABLE XIII: SUMMARY STATISTICS: POTENTIAL CHILD INCOMES

Children aged 12-15	Mean	Std. Dev.
Predicted income (boy)	125.5	33.2
Predicted income (girl)	113.5	29.5

Note: Average weekly predicted incomes in pesos.

B.4 Heterogeneous Income and Price Effects: More Covariates

We now allow for price and income effects to be heterogeneous w.r.t. child age and gender. Specifically we estimate:

$$\begin{aligned}
 Enroll_{it} = & b_0 + b_1 t_{c,it} + b_2 t_{u,it} + b_3 t_{c,it} y_{it}^f + b_4 t_{u,it} y_{it}^f + b_5 t_{c,it}^2 + b_6 t_{u,it}^2 + b_7 t_{c,it} t_{u,it} \\
 & + b_8 t_{c,it} age_{it} + b_9 t_{u,it} age_{it} + b_{10} t_{c,it} boy_i + b_{11} t_{u,it} boy_i + \beta Z_{it} + \delta_t + \eta_i + e_{it}
 \end{aligned}$$

Regression results are presented in Table XIV. Moments of the income and price effect schedules are presented in Table XV.

44. Note, for those children actually working, their mean hours worked is 40 hours per-week. However, their median hours worked is 48 hours per-week. We choose to multiply predicted wages by 40 instead of 48 so as to underestimate potential child earnings, and thus, underestimate the size of the targeting benefit.

TABLE XIV: HETEROGENEOUS INCOME AND PRICE EFFECTS: MORE COVARIATES

	(1)
tc	0.0418* (0.022)
tu	0.0259 (0.036)
tc ²	0.0000586 (0.000)
tu ²	-0.00121 (0.001)
tc × tu	-0.0000300 (0.001)
tc × y^f	-0.000269* (0.000)
tu × y^f	-0.000288 (0.000)
tc × age	-0.00170 (0.002)
tu × age	-0.000505 (0.002)
tc × boy	-0.000902 (0.003)
tu × boy	-0.000411 (0.005)
y^f	0.00402 (0.004)
Observations	17432
R-squared	0.124

Note: Standard errors are clustered at the locality level and are presented in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $Enroll_{it}$. Sample: unbalanced panel of children aged 12-15 years in 1997-1999 from eligible (poor) households with two parents and one parent reports to be the head of the household. y^f denotes father's weekly per-capita income; tc (tu) denotes the weekly, offered per-capita CCT (UCT). We instrument for father's per-capita income with average locality wage (excluding the wage of the father in question), and we instrument for $t_c \times y^f$ and $t_u \times y^f$ with the CCT and UCT interacted with the average locality wage (excluding the wage of the father in question). Regressors not shown: child age dummies, highest grade complete dummies, age interacted with highest grade completed, year dummies, and a child fixed effect.

TABLE XV: MOMENTS OF THE INCOME AND PRICE EFFECT SCHEDULES: MORE COVARIATES

$\mathbb{E}[P(X)]$	0.0117*** (0.0033)
$\mathbb{E}[I(X)]$	0.0041 (0.0035)
$\mathbb{E}\left[\frac{I(X)}{P(X)}\right]$	0.22 (0.55)

Note: Moments are calculated for the treated, eligible sample of children aged 12-15 years in 1998 and 1999 (see Appendix B.2 for explicit calculation). Bootstrapped standard errors from 500 bootstrap draws are presented in parentheses.

B.5 Sensitivity

Varying the Curvature of Utility. We now consider how varying some of our key parameters affects the optimal CCT. First we consider varying the curvature of utility of consumption, γ . We do so by scaling the income effect function, $I(X_{it})$, which then scales the average income effect, which then scales $\mathbb{E}\left[\frac{I}{P}\right]$ (see Equation (13)), which then scales our estimate of curvature, γ (see Equation (14)). A higher scaling factor (i.e., larger income effects) implies a higher γ . We then recalculate the optimal share of the budget allocated towards a CCT under the different values of γ . Figure IX illustrates how the optimal share of the budget to a CCT varies with γ . As to be expected by Proposition 4, when $\gamma = 0$, a pure UCT is optimal. However, as curvature increases, so does t_c^* . This is because the enrolled households have lower consumption on average; thus, as curvature increases, the extent to which the enrolled households value an extra dollar relative to the unenrolled households increases.

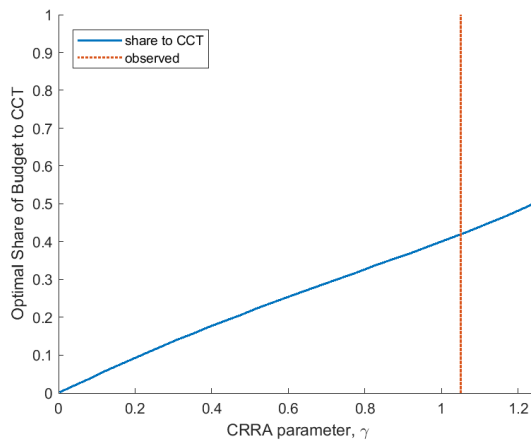


FIGURE IX: VARYING CURVATURE OF UTILITY

Varying Child Income. We now vary our estimate of potential child income, y_c , by scaling our predicted measure by $\kappa \in [0, 1]$. Holding enrollment decisions constant, the optimal CCT will be increasing in the cost of schooling, i.e., child income. This relationship is highlighted in Figure X below. As child income falls (i.e., κ falls), the drop in consumption experienced by households who send to school falls; thus, the size of the targeting benefit relative to the exclusion cost falls, decreasing the optimal CCT. When the cost of schooling is zero (i.e., $\kappa = 0$), optimal spending on the CCT is also 0 as parents incur no loss in consumption when sending their children to school (of course, in our model, if schooling were free, enrollment would be 100% so it wouldn't matter if you offered a CCT vs. a UCT. However, Figure X is constructed holding enrollment decisions fixed).

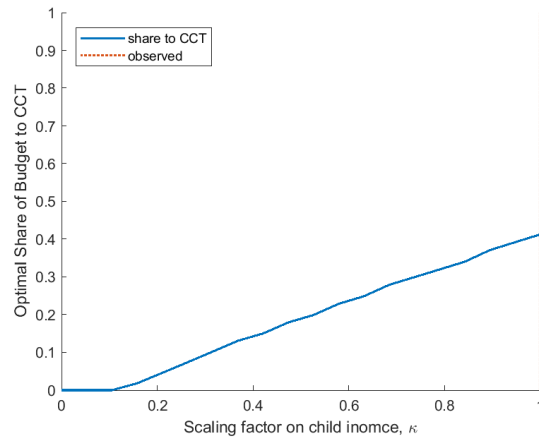


FIGURE X: VARYING CHILD INCOME