THE TARGETING BENEFIT OF CONDITIONAL CASH TRANSFERS*

Katy Bergstrom† William Dodds‡

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 of the poorest households as access is tied to investments in children, which are normal goods. However, we argue that conditionalities based on children's school enrollment may actually improve the targeting of transfers to low consumption households. This is because sending a child to school can result in a discrete loss of child income, meaning that school enrollment may be negatively correlated with household consumption. The size of the targeting benefit is directly related to two elasticities already popular in the literature: the income effect of a UCT and the price effect of a CCT. We estimate these elasticities for a large CCT program in rural Mexico, Progresa, using variation in transfers to younger siblings to identify income effects. We find that the targeting benefit is almost as large as the cost of excluding some low-income households; this implies that if the only benefit of imposing conditions is improved targeting, 48% of the Progresa budget should go to a CCT over a UCT. *JEL* Codes: I28, I38, H21, H31, H52.

*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 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.

†Department of Economics, 579 Serra Mall, Stanford, CA 94305. Email: katyberg@stanford.edu. Tel: 650-440-9144.

‡Department of Economics, 579 Serra Mall, Stanford, CA 94305. Email: wdodds@stanford.edu.

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, these two aims can be at odds with one another as the poorest households may find the conditions too costly to comply with and thus be excluded from receiving aid (e.g., Baird et al, 2011; Freeland, 2007). Unconditional cash transfers (UCTs), cash transfers with "no strings attached", are therefore thought to be superior at alleviating current poverty.[1] Naturally, the debate over whether cash transfer programs should include conditions has been at the forefront of recent global policy discussions (Baird et al, 2013).

This paper argues that conditioning cash transfers on school attendance can improve targeting of transfers to low consumption households, making CCTs more effective at reducing current poverty than UCTs. The central idea is that sending a child to school rather than to work results in a discrete loss of child income, meaning that school enrollment may be negatively correlated with current household consumption. School enrollment can therefore act as a "tag" (an observable indicator) for low consumption households.[2] By conditioning transfers on schooling, governments may be able to target resources towards a group with lower consumption. We refer to this unexplored benefit of CCTs as the *targeting benefit*.

To illustrate the targeting benefit of CCTs, 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 discontinuously drop by the amount of the child's potential earnings; this discontinuous drop in household consumption is illustrated in Figure I below, where y_{child} denotes potential child income. Consequently, households that send their child to school may have lower consumption on average. Because a CCT allows the government to target transfers to the households sending to school, a CCT may be better at alleviating current poverty than a UCT.

1. Like CCTs, UCTs are cash transfers targeted to poor households; however, unlike CCTs, UCTs are not made conditional on human capital investments in children.

2. Given that we are conditioning on a decision (schooling) rather than an immutable characteristic, we refer to this benefit as a targeting benefit instead of a tagging benefit (see Akerlof, 1978).

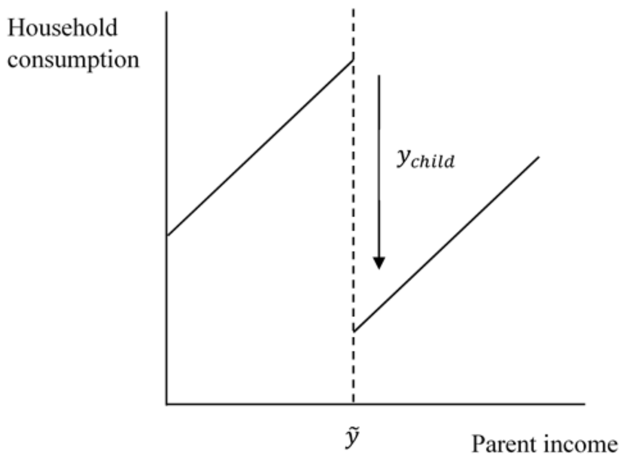

Figure I: Household consumption vs parent income

We first develop a theoretical framework to model the targeting benefit of CCTs. We consider a two-generation world in which households consist of a parent and 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 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 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 the set of (predetermined) eligible households and who has a fixed budget to redistribute during the first generation.[3] We assume parents make education decisions at the socially optimal level, thereby abstracting from conventional motives to condition transfers; consequently, the planner wants to transfer resources to the households that have the highest marginal utility of consumption in the first generation.[4] The planner chooses the share of the budget to allocate towards a constant UCT versus a constant CCT. By increasing the CCT, she increases the share of money received by the households sending their child to school (the enrolled households). We refer to the welfare gain experienced by the enrolled 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 households not sending their child to school (the unenrolled households). We refer to the welfare loss experienced by the unenrolled households as the exclusion cost. Our first result is that under some conditions it can be optimal for the planner to allocate some or all of her budget towards the CCT. This will be the case when the households with children enrolled in school have lower average consumption, so that they place

3. While in reality the planner also gets to choose the set of eligible households (e.g., through proxy means testing or geographic targeting), we abstract away from this decision so as to simplify the analysis. However, we investigate how changing the distribution of parental incomes in this eligible set affects the importance of the targeting benefit.

4. This is because the planner only has a budget to redistribute in the first generation, and because we assume (realistically) parents have no means of moving resources across generations. Hence, the planner's objective of maximizing total lifetime utility over both generations amounts to an objective of transferring towards those with the lowest consumption (highest marginal utility of consumption) today. Note, 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 et al (2003).

a greater value on receiving an extra dollar today relative to the unenrolled households, i.e., when the targeting benefit outweighs the exclusion cost.

One potential concern with our formulation is that while conditions can allow us to direct more resources towards households with low consumption today (and, hence, high marginal utility of consumption), conditions direct more money to enrolled households who have higher lifetime utility. This is in part because our baseline specification only considers a planner with a budget to redistribute in the current generation as this fully captures the insight behind the targeting benefit in a relatively parsimonious fashion. However, since children from households that invest in schooling today will have higher incomes in the future, current beneficiaries of CCTs will be less likely to meet eligibility criteria for future transfers, thereby freeing up resources for higher transfers to eligible beneficiaries in the next period. We show that in a model in which the planner has a budget to redistribute over multiple generations, low lifetime utility households will end up receiving more over the course of multiple generations.

The size of the targeting benefit relative to the exclusion cost depends on three factors: (1) the concavity of utility of consumption, (2) the distribution of parental incomes among eligible households, and (3) the size of potential child incomes (i.e., the cost of schooling). 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; hence, it can never be optimal to allocate any of the budget towards a CCT based on the targeting benefit alone. Second, we show that as parental income inequality increases among the eligible set of households, the targeting benefit falls while the exclusion costs 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 good). Third, the larger potential child incomes, the greater the "lumpiness" of the investment, thus the greater the targeting benefit relative to the exclusion cost (all else equal).

Is the empirical magnitude of the targeting benefit large enough to warrant policy relevance? To shed light on this question, we express the size of the targeting benefit relative to the exclusion cost in terms of empirically observable objects: the parental income distribution, 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 significant curvature in utility of consumption.

We then estimate income and price effects for Progresa, a large CCT program in rural Mexico. Started 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 transfers paid to mothers conditional on their school-age children attending school on a regular basis. These grants 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. The introduction of these transfers was randomized at the locality level: 63% of the localities received transfers immediately, while the other 37% started two years later. This variation allows us to identify price effects. 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 are 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 1.27.[6]

We use these estimates to evaluate the size of the targeting benefit relative to the exclusion cost under the observed Progresa transfers. We find that it is substantial - the targeting benefit is 91% as large as the exclusion cost under the observed transfers. We then calculate the share of the Progresa budget that should be made conditional on school attendance assuming that the only benefit of conditionalities is improved targeting (i.e., we calculate the share that equates the targeting benefit to the exclusion cost). We find that 48% of the budget should go towards a CCT, implying that the targeting benefit is a quantitatively important benefit of CCTs in this setting (in comparison, under the observed transfers, 64% of the budget goes towards a CCT).[7] 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 80% 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. Consequently, the households sending their teenage children to school have higher marginal utilities of consumption on average. Allocating some of the budget towards a CCT better targets transfers towards households who place a higher value on receiving an extra dollar today.

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 to estimating the size of the targeting

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 1.27.

7. I.e., of Progresa spending to households with a secondary school age child, 64% is spent on CCTs to these children, whereas 36% is spent on (effectively) unconditional transfers to these children's younger siblings.

benefit, Section V estimates the income and price effects for Progresa, Section VI determines the size of the targeting benefit in the context of Progresa and the optimal share of the Progresa budget to be allocated to a CCT based on the targeting benefit. Finally, Section VII concludes.

II Contribution to the Literature

Our paper contributes to the literature in four 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 (see, for example, Fiszbein and Schady, 2009; Baird et al, 2011; Baird et al, 2013; Benhassine et al, 2015).

Second, 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 consensus is that if the income effect is large relative 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, 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.

Third, 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; Case et al, 2005; 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.[8] 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

8. 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.

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

In this section we develop a model to highlight the targeting benefit of conditional cash transfers. To do so, we abstract away from conventional motives to impose conditions (e.g., parental under-investment motives and/or political economy motives). 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.C that this assumption is crucial to warrant any of the budget being allocated towards a CCT based on the targeting benefit.

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.C). 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 altruistic preferences of parents 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 this 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 additional heterogeneity is important. The parents' problem can be written as follows:

$$\max_{s\in\{0,1\}} \; u(c) + v(\mu, s)$$

$$\text{s.t. } c = y + y_c(1-s)$$

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). □

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:

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

Proof. See Appendix A.1. □

Assumption 1 ensures that parents who send their child to school earn on average more 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 strong, negative correlation between child ability and parental income.

III.A Social Planner Problem

We consider a utilitarian social planner with a fixed budget to distribute to eligible households in the first generation. Note, for simplicity, we assume eligibility has been predetermined (say via proxy-means testing and/or geographical targeting).[9] We assume the planner cannot condition transfers on parental income given the difficulty in verifying incomes in developing countries, nor can she observe child abilities. However, we assume she can observe schooling decisions.[10][11] 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 enrolled eligible households receive. By assuming a utilitarian planner with a fixed budget, we remove any direct motives for the planner to want to distort parents' enrollment decisions.[12] We can write the social planner's problem as follows:

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

where Y denotes the set of parental incomes for households that are eligible to receive the transfers, and M denotes the set of household specific child abilities. 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:

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

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

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

9. In reality, the set of eligible households is also a choice variable for the planner. We show later how changing the distribution of parental incomes of the eligible set of households affects the size of the targeting benefit (see Proposition 5).

10. I.e., we are operating in a second-best world. In a first-best world (i.e., a world in which the planner can observe parental incomes and child abilities), the planner would simply give money to those with the highest marginal utilities of consumption until either the budget runs out or marginal utilities are equated across all households (see Appendix A.4 for first best solution).

11. In Appendix A.5 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 $t_c(y)$. We show that it is always optimal to spend some of the budget on a CCT based on the targeting benefit alone, i.e., $t_c(y) > 0$ for some y.

12. First, we shut down under-enrollment motives to impose conditions by assuming the social planner problem coincides with the parent problem, e.g., we assume both planner and parents place the same value on the return to schooling. Second, we shut down any political economy motives to condition by assuming the planner has a fixed, exogenous budget to redistribute.

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

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

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:

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

The first term in Equation (3), what we denote as the targeting benefit, represents the net social welfare gain from increasing t_c for those sending to school. Specifically, $\left(1 + \frac{\partial t_u}{\partial t_c}\right)$ represents the net increase in transfers the enrolled 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 enrolled 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 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 $\frac{\partial t_u}{\partial t_c} < 0$.[13]

In this model, in order for the planner to allocate some of the budget towards a CCT, the targeting benefit must outweigh the exclusion cost at $t_c = 0, t_u = R$. As we will show in the next subsection, this will occur if the parents sending to school only earn slightly more than the parents not sending to school, and if the cost of schooling (i.e., forgone child income) is substantial relative to parental income, leading to the enrolled households having, on average, lower household consumption. Finally, while our model abstracts from conventional motives to impose conditions (e.g., we assume parents invest at the optimal level), we show in Appendix A.6 how our model can be adapted to incorporate the possibility that parents undervalue the return to schooling. In this framework, CCTs now offer an additional benefit over UCTs: they induce more parents to send their children to school. Thus, CCTs help correct for parental

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

under-investment more than UCTs. Our main framework abstracts from these conventional motives to highlight that the targeting benefit on its own is enough to warrant spending on a CCT. However, when quantifying the size of the targeting benefit relative to the exclusion cost in the context of Progresa, we also provide a back-of-the-envelope calculation to determine the size of the targeting benefit relative to the enrollment benefit CCTs offer when parents under-value the return to education (see Section VI.D).

III.B Results

Using the above framework we investigate the optimal allocation of the planner's budget towards a CCT. First, we focus on the case where utility is concave in consumption ($u_{cc} < 0$). We are able to show that, despite the planner having no direct incentive to distort schooling decisions, it may be optimal for her to allocate some or all of the budget towards a CCT. To understand why, consider a 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 towards the enrolled 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).[14] 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 it is beneficial to target transfers towards the enrolled households. This leads us to our main proposition.

14. 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.

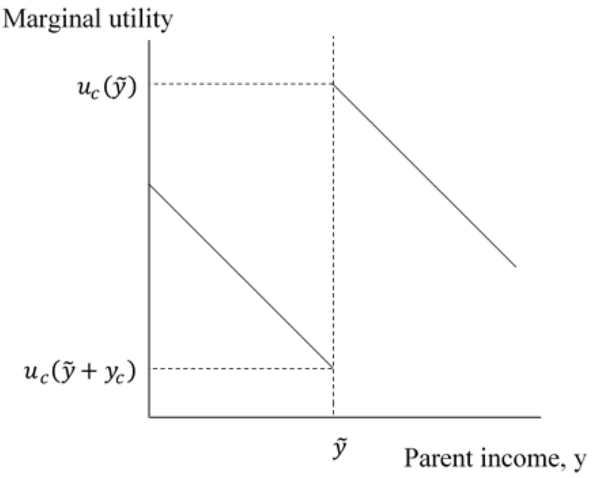

Figure II: Discontinuity in marginal utility of consumption (constant μ)

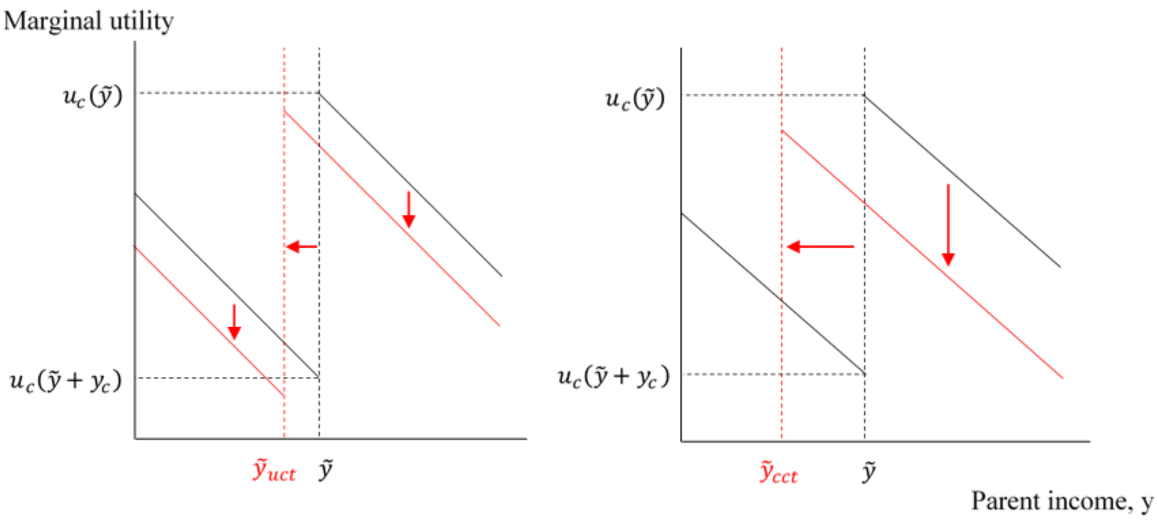

Figure III: Pure UCT (left) vs Pure CCT (right)

Proposition 3. *If utility is concave in consumption, a pure UCT may not be optimal.*

Proof. See Appendix A.7. □

We prove Proposition 3 by constructing an example where the enrolled households have higher marginal utilities of consumption on average relative to the non-enrolled households (under a pure UCT: $t_c = 0, t_u = R$), thus making it beneficial for the planner to target money towards the enrolled households (i.e., offer $t_c > 0$). We construct this example by imposing the following conditions: (1) curvature in the utility of consumption; (2) large child incomes relative to parental incomes; and (3) a parental income distribution with relatively little inequality so that the difference in average parental incomes between enrolled and non-enrolled households is relatively small.

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 3.1. *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. See Appendix A.8. □

Corollary 3.1 highlights that the targeting benefit can still be an important benefit of CCTs over UCTs under social welfare functions that place some weight on redistributing towards those households with lower lifetime utility (e.g., log utilitarian). Notably, Corollary 3.1 relies on $G(\cdot)$ being everywhere positive and continuous. This requirement rules out a Rawlsian social welfare function in which the planner only cares about redistributing to the lowest lifetime utility household (i.e., $G(u(c^*) + v(\mu, s^*)) = \min_{y,\mu} u(c^*) + v(\mu, s^*)$). If the planner is Rawlsian, it can never be optimal for her to allocate any of the budget towards a CCT, as the lowest lifetime utility household must be a household who does not send to school.[15] Corollary 3.1 shows that as long as we place some welfare value on redistributing towards those who value an extra dollar today the most, it can be optimal to allocate some of the budget towards a CCT.

We now 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 some parents (i.e., by imposing conditions, some parents choose to send to school who otherwise would not have if offered the same amount of money unconditionally). This distortion is costly in our framework given we abstract from any benefits associated with increased enrollment. Thus, 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.9. □

Next, we consider how the targeting benefit of CCTs (and thereby t_c^*) 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 CCTs are more desirable.

15. Note, when discussing CCTs under a Rawlsian objective, we are making two realtively 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 gaurantee that the lowest lifetime utility household is a household not sending their child to school.

Proposition 5. *Suppose $t_c^* > 0$ under $f^{(1)}(y, \mu)$. 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 enrolled are the same under both $f^{(1)}$ and $f^{(2)}$ at t_c^*.*[16] *Then under $f^{(2)}(y, \mu)$, the optimal local reform starting from t_c^* is to increase t_c.*

Proof. See Appendix A.10. □

The idea behind Proposition 5 is that as we shift the distribution of parental incomes for those sending to school to the left, the targeting benefit becomes more important. To see this graphically, Figure IV plots the density of parental incomes for those sending to school under $f^{(1)}$ and $f^{(2)}$ and the density of parental incomes for those not sending to school. Because $F^{(1)}(y|s^* = 1)$ FOSD $F^{(2)}(y|s^* = 1)$, the density of incomes under $f^{(1)}$ is situated to the right of that under $f^{(2)}$. Figure V translates these parent income densities into household income densities (where household income is equal to parental income plus constant child income if the child is not enrolled). It can be seen that household income under $f^{(2)}$ is lower than that under $f^{(1)}$. Thus, the benefit of transferring towards those households that send their child to school is higher under $f^{(2)}$ than $f^{(1)}$.

Proposition 5 is closely related to parental income inequality; because schooling is a normal good (i.e.,$\frac{\partial s^*}{\partial y} \geq 0$), if the enrolled households only have slightly higher parental incomes on average relative to the unenrolled households, this corresponds to a low degree of parental income inequality (loosely speaking). Conversely if the enrolled households have much higher parental incomes on average, this corresponds to a high 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).

Finally, we consider how changing the cost of schooling (i.e., changing the size of potential 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.

16. Actually, we need only $\frac{\partial t_u}{\partial t_c}$ is the same under $f^{(1)}, f^{(2)}$ at t_c^*. A sufficient condition for this is that total enrollment: $\int_Y \int_{\tilde{\mu}} f(\mu, y) d\mu dy$, average price effects: $\int_Y \frac{\partial \tilde{\mu}}{\partial t_c} f(\tilde{\mu}|y) f(y) dy$, and average income effects: $\int_Y \frac{\partial \tilde{\mu}}{\partial t_u} f(\tilde{\mu}|y) f(y) dy$ evaluated at t_c^* are the same under $f^{(1)}, f^{(2)}$.

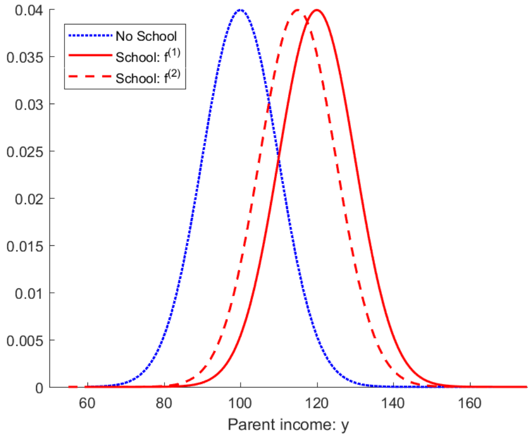

FIGURE IV: DISTRIBUTION OF PARENT INCOME (SPLIT BY ENROLLMENT)

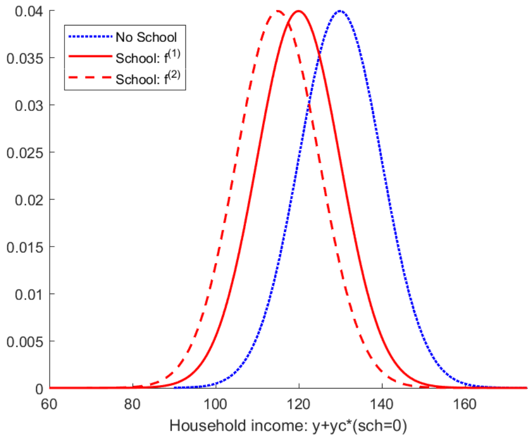

FIGURE V: DISTRIBUTION OF HOUSEHOLD INCOME (SPLIT BY ENROLLMENT)

III.C Extensions

We consider five extensions to our baseline model. Our first extension allows parents to freely borrow across generations. While we do not think this extension is realistic (i.e., markets to borrow against your child's future earnings do not exist), we believe this is a worthwhile extension to consider as it highlights that the lack of these borrowing markets is crucial for the targeting benefit to be relevant (i.e., for our main result, Proposition 3, to hold). In contrast, the remaining four extensions can be viewed as realistic extensions to our baseline model, and are considered to illustrate that our main result, Proposition 3, is not simply an artifact of simplifying assumptions we make in our baseline set-up.

First, we consider the scenario where parents can borrow freely across generations (for a complete discussion with the model set up, propositions, and proofs, see Appendix A.11). We show that in this model, a pure UCT is optimal (i.e., the targeting benefit can never justify spending on a CCT). This is because households are now able to perfectly smooth their consumption across the two generations. Thus, households will only send to school if doing so increases their total utility, i.e., schooling increases their consumption in both generations.

Hence, the households that send to school must have higher consumption today, implying it cannot be optimal to target transfers towards them.

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 therefore borrow against their own income (in the current 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.12 for model set-up and results).[17] 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.

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.14 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 the extension where parents make annual, binary schooling decisions over a set number of T years (see Appendix A.15 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. We are able to show that the planner's FOC is simply a sum of the annual targeting benefits and annual exclusion costs over the T years.

Finally, we consider the extension where the planner has money to redistribute over multiple generations. We do so to not only highlight that the targeting benefit is still relevant in the first generation, but also to illustrate that dynamic transfer schemes can mitigate concerns that we are on net transferring more to higher lifetime utility households. The key intuition is as follows: the households sending to school in the first generation still experience a loss in consumption today, thus making it potentially beneficial to target transfers towards them today. However, assuming returns to schooling are sufficiently high, children who went to school in the first generation do not meet eligibility criteria for transfers (either CCTs or UCTs) 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.13 for model set-up and results). In essence, the planner is alleviating the inefficiencies caused

17. 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.

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).

IV Sufficient Statistics for the Size of the Targeting Benefit

We've shown theoretically that the targeting benefit can be important, but is it quantitatively important in reality? In this section we develop a way to estimate the size of the targeting benefit relative to the exclusion cost of a CCT from empirically observable objects (sufficient statistics). This method will also allow us to determine the optimal CCT/UCT mix assuming the only benefit of imposing conditions is the targeting benefit. Hence, this method will be useful to determine whether the targeting benefit can be a quantitatively important benefit of CCTs relative to UCTs. To begin, we re-write the targeting benefit (TB), relative to the exclusion cost (EC) as follows:

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

where $S(y)=\int_{\tilde{\mu}(y,t_c,t_u)} f(\mu,y)d\mu$ denotes the proportion of households with parent income y sending their child to school (remembering $\tilde{\mu}(y,t_c,t_u)$ denotes the household with parental income y just indifferent between sending to school and not; see Equation (1)). Note for simplicity of notation we have suppressed that $S(y)$ is a function of the transfer schedule (t_c,t_u). 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 enrolled households value receiving an extra dollar relative to the unenrolled households. Thus, we first proceed with 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:

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

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

where $I(y)$ measures the increase in the share of parents (with income y) sending their child to school as we increase the unconditional transfer, and $P(y)$ measures the increase in the share of parents (with income y) sending their child to school when we increase the conditional transfer (again we have suppressed that both $I(y)$ and $P(y)$ are also functions of t_c, t_u). 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:

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

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

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

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

From Equation (7) we can see that the ratio of the income effect to the price effect for a given parental income level is proportional to the discontinuity in marginal utility of consumption that results from sending a child to school. Notably, for a given schedule (t_c, t_u), this relationship holds for all parental income levels y. Therefore, if we had sufficient data, we could estimate $I(y)$ and $P(y)$ for all parental income levels under the observed transfer schedule, thus allowing us to recover marginal utility of consumption at 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 $\frac{I(y)}{P(y)}$ 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 to pin down marginal utility of consumption as γ solves:

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

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 unconditional cash transfer. 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 remove 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, where averages are taken over the distribution of parental incomes. 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 $\bar{P} = \int_Y P(y) f(y) dy$, etc. This leads us to the following result: if one can observe the income and price effect schedules, $I(y)$ and $P(y)$, the shares enrolled, $S(y)$, and the density of parental incomes $f(y)$, one can evaluate the size of the targeting benefit relative to the exclusion cost under the observed transfer schedule.[18] It is worth mentioning that calculating TB/EC under the observed transfer schedule is equivalent to determining the optimal local reform assuming the only benefit of a CCT is the targeting benefit. For example, if we observe a ratio greater than 1, $\frac{TB}{EC} > 1$, this suggests it is optimal to increase the conditional cash transfer and decrease the unconditional cash transfer 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 benefit 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 not only need to observe how the income and price effects vary with parental income, but also how they vary with the transfer schedules themselves, i.e., we need to observe $I(y, t_c, t_u), P(y, t_c, t_u)$.

Lastly, we show that we can extend the above analysis to a model in which parents make annual schooling decisions over a number of years (see Appendix A.16). In this framework we relax the assumption that child income is constant, instead allowing it to vary with child age and gender. Consequently, in order to evaluate the size of the targeting benefit relative to the exclusion cost, one needs to observe the income effect schedule, the price effect schedule, and the enrollment schedule that are heterogeneous w.r.t. parental income, child age, and child gender, along with the joint density of parental income, child age, and child gender. When we move to our empirical exercise where we estimate the size of the targeting benefit for Progresa villages, we will implement this extended model thus allowing elasticities to vary with household and child characteristics.

18. One also needs to observe the cost of schooling y_c.

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 conditional cash transfer program in rural Mexico. We will focus in this section on estimating the quantities necessary to recover curvature of utility (the income and price effects) using evaluation data from Progresa. We assume, for simplicity, that the utility function takes the CRRA form ($u(c) = \frac{c^{1-\gamma}}{1-\gamma}$) 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 conditional cash transfer 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.[19] 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.

19. See Skoufias, Davis, and Behrman (1999) for more information on targeting of localities and households.

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.

TABLE I: Progresa Transfer Amounts (bimonthly, in pesos)

	1998, semester 1	1998, semester 2	1999, semester 1	1999, semester 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

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-Progresa 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 (3% 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 only 14 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)	168.6	116.0
Weekly income (mother, inc>0)	135.0	134.7
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)	131.3	83.0
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)	142.4	100.3
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 $1^{st} - 99^{th}$ percentiles). All incomes are inflation adjusted to be in 1998 values (inflation is calculated as the average percentage change in annual incomes by state).

Of course, 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).[20] 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 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

20. 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 day.

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 131 (142) 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.[21] Thus, combining these earning statistics with the fact that schooling directly competes with working, it is likely that parents forgo a discrete loss in consumption as a result of sending their child to school.

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 unconditional cash transfer and the conditional cash transfer, respectively. In order to recover these derivatives, we first need to discuss where our source of identifying variation in unconditional cash transfers comes from given Progresa only offered conditional cash transfers.[22]

To obtain variation in unconditional cash transfers, we will exploit the fact that enrollment of children below the age of 12 is nearly 100%. Figure VI plots the percentage of all children enrolled by age and gender in 1997.[23] 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 to the household.

21. 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.

22. 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.

23. 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.

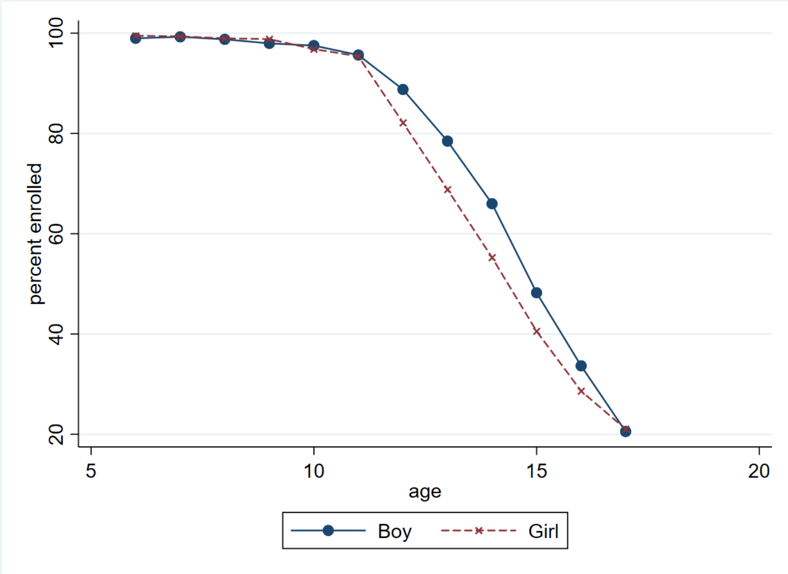

Figure VI: Enrollment by age and gender, 1997

Therefore, for children aged 12 and above, we observe both variation in conditional and unconditional cash transfers. Variation in conditional cash transfers 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 unconditional cash transfers 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 conditional and unconditional cash transfers 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, loosely speaking, 7 year-old children are enrolled in grade two and, thus, do not receive Progresa education grants, while 8 year-old children are enrolled in grade three and, thus, do receive Progresa 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: $I = \frac{1}{t_u}\left((f - e) - (h - g) - Pt_c\right)$. 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.[24] First we estimate a standard difference-in-difference regression to obtain the overall treatment effect of the Progresa program on enrollment:

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

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 and 1999 relative to eligible children aged 12-15 years in control localities.[25]

24. 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.

25. 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.

TABLE IV: Overall Treatment Effect

	(1)
treat×$\mathbb{1}$(year>1997)	0.0883***
	(0.015)
$\mathbb{E}$[Enroll|control,year > 97]	0.683
$\mathbb{E}$[Enroll|treat,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. treat × $\mathbb{1}$(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 $\mathbb{1}$(year > 1997).

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

$$Enroll_{it} = b_0 + b_1 t_{c,it} + b_2 t_{u,it} + \beta Z_{it} + \delta_t + \eta_i + e_{it} \tag{10}$$

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).[26] Specifically, $t_{u,it}$ is calculated as follows:

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

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.[27] 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).[28] Finally δ_t and η_i denote year and child fixed effects, respectively. Note, by including a child 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

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

27. 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).

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

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 15% level.[29] Notably, the presence of a positive income effect suggests utility is not linear in consumption. Specifically our constant price effect is equal to 0.0083 and our constant income effect is equal to 0.0040. 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 0.83 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 0.40 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 6.85 percentage points higher relative to control localities after 1997 (as $0.83 \times 6.86 + 0.40 \times 2.90 = 6.85$). 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, and/or perhaps due to the other benefits of Progresa such as the health and nutrition transfers that are not captured in t_c and t_u. 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 1.26 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 0.61 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 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.[30]

29. 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.

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

TABLE V: Constant Income and Price Effects

	(1) OLS	(2) IV	(3) IV	(4) IV
tc	0.00885***	0.00832***	0.00820***	0.00849***
	(0.002)	(0.002)	(0.002)	(0.002)
tu	0.00406$^+$	0.00399$^+$	0.00429$^+$	0.00395$^+$
	(0.003)	(0.003)	(0.003)	(0.003)
y^f	0.0000426	0.00379	0.00347	0.00379
	(0.000)	(0.005)	(0.005)	(0.005)
tc$^{(12+)}$				-0.000562
				(0.002)
$\mathbb{E}$[Enroll|control,year > 97]	0.694	0.693	0.693	0.693
$\mathbb{E}$[Enroll|treat,year > 97]	0.778	0.778	0.778	0.778
$\mathbb{E}$[tc|treat,year > 97]	6.854	6.864	6.864	6.864
$\mathbb{E}$[tu|treat,year > 97]	2.897	2.897	2.897	2.897
$\mathbb{E}$[famsize]	6.609	6.609	6.609	6.609
Observations	17365	17279	17279	17279
R-squared	0.145	0.120	0.127	0.120

Note: Standard errors are clustered at the locality level and are presented in parentheses. $^+$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 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 complete 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:

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

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 likely the case in this setting, see Table II), this will be a negative income shock to the household.[31]

V.E Heterogeneous Income and Price Effects

In the above regressions we assumed the income and price effects were constant. However, it is likely these effects are heterogeneous w.r.t. household characteristics, in particular parental income, the transfer levels themselves, and child characteristics that affect forgone child incomes and/or the returns to schooling (e.g., age and gender). Allowing for heterogeneity in these effects, we estimate the following regression:

$$\begin{aligned} Enroll_{it} =& b_0 + b_1 t_{c,it} + b_2 t_{u,it} + b_3 t_{c,it}^2 + b_4 t_{u,it}^2 + b_5 t_{c,it} t_{u,it} + b_6 t_{c,it} y_{it}^f + b_7 t_{u,it} y_{it}^f + b_8 t_{c,it} age_{it} \\ &+ b_9 t_{u,it} age_{it} + b_{10} t_{c,it} boy_i + b_{11} t_{u,it} boy_i + \beta Z_{it} + \delta_t + \eta_i + e_{it} \end{aligned} \tag{12}$$

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; age_{it} denotes child $i's$ age in years; boy_{it} denotes an indicator for whether child i is a boy; δ_t denotes year fixed effects; and η_i denotes child fixed effects. As in Equation (10), Z_{it} includes father's per-capita income, child age dummies, highest grade completed dummies, and child age interacted with grade. Finally, we instrument for father's income using the average locality wage (excluding the wage of the father in question).[32] In this specification, the price and income effects are given by:

$$P(X_{it}) = b_1 + 2b_3 t_{c,it} + b_5 t_{u,it} + b_6 y_{it}^f + b_8 age_{it} + b_{10} boy_i$$

$$I(X_{it}) = b_2 + 2b_4 t_{u,it} + b_5 t_{c,it} + b_7 y_{it}^f + b_9 age_{it} + b_{11} boy_i$$

where $X_{it} = \{t_{c,it}, t_{u,it}, y_{it}^f, age_{it}, boy_i\}$. Regression results of Equation (12) for both our OLS and IV estimates are presented in Table XIII in Appendix B.4. To test for heterogeneity in the income and price effects we conduct an F-test on all terms interacted with the conditional and unconditional cash transfers. In the OLS specification, we cannot reject the null of no heterogeneity, whereas in the IV specification we can. This discrepancy between IV and OLS appears to be driven by the fact that our IV estimates suggest price effects are significantly decreasing in father's income as indicated by the significant, negative coefficient on $t_c \times y^f$ (a

31. See Ferreira et al (2009) for a more in-depth analysis into how conditional transfers to siblings can have ambiguous effects on a child's enrollment.

32. 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.

result consistent with curvature in utility as $P'(y) \propto u_{cc}$). While our OLS specification also suggests price effects are decreasing in parental income, this relationship is insignificant and an order or magnitude lower, likely due to attenuation bias. There is little evidence to suggest there exists heterogeneity in income effects (in either our OLS or IV specifications), likely due to the much smaller variation in unconditional cash transfers relative to conditional cash transfers.

Moments of the income and price effect schedules are presented in Table VI below. We find significant average price effects of 0.0097 and 0.013 under our OLS and IV specifications, respectively, and average income effects of 0.0039 and 0.0035 under our OLS and IV specifications, respectively (see Appendix B.2 for calculation of moments). We then estimate the average value of the ratio of the income effect to the price effect using the following second-order Taylor series expansion:[33]

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

We find the average of this ratio is equal to 0.43 under our OLS specification and 0.25 under our IV specification. This lower ratio is primarily due to a higher average price effect under the IV specification. Finally, using the estimated coefficients from Regression (12), we calculate the implied increase in enrollment in treatment localities for the years 1998-1999 under the observed transfers. We calculate the implied increase in enrollment to be 9.2 percentage points under our OLS specification and 8.7 percentage points under our IV specification. These estimates are in line with the observed increase in enrollment of 8.8 percentage points (see column (1) of Table IV).

TABLE VI: Moments of the Income and Price Effect Schedules

	OLS	IV
$\mathbb{E}[P(X)]$	0.0097***	0.013***
	(0.0025)	(0.0034)
$\mathbb{E}[I(X)]$	0.0039	0.0035
	(0.0034)	(0.0030)
$\mathbb{E}\left[\frac{I(X)}{P(X)}\right]$	0.43	0.25
	(0.45)	(0.28)

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 100 bootstrap draws are presented in parentheses.

33. $\mathbb{E}\left[\frac{\hat{I}(X)}{\hat{P}(X)}\right]$ 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). However, it turns out that calculating $\mathbb{E}\left[\frac{\hat{I}(X)}{\hat{P}(X)}\right]$ gives us a very similar result to our second-order expansion, with $\mathbb{E}\left[\frac{\hat{I}(X)}{\hat{P}(X)}\right] = 0.28$ under our IV specification.

V.F Coefficient of Constant Relative Risk Aversion

Now that we've estimated income and price effects, we can use these estimates to calculate curvature of utility. We use our estimates 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):

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

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 are working 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 ratio, where we vary women's earnings from 0% to 100% of their husbands' earnings.

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 the 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 years 1998 and 1999, and using our estimates of $\mathbb{E}\left[\frac{I}{P}\right]$ from Table VI above, we solve Equation (14) for γ. Our results are presented in Table VII below. Our curvature estimate under our IV specification is 1.27. Interestingly this estimate is very similar to that estimated in Layard and Mayraz (2008) who find $\gamma = 1.26$, and is in-line with Chetty (2006) who finds a value of $\gamma \approx 1$, i.e., $u(c) = log(c)$. Our curvature estimate under our OLS specification is substantially higher with $\gamma = 2.50$. Given we suspect our OLS specification suffers from attenuation bias, we will proceed with our IV estimates for the remainder of the paper. We will see in the following section, however, that the size of the targeting benefit is increasing in γ. Thus, using γ from our IV specification will provide a conservative estimate for the size of the targeting benefit. Finally, we vary mothers' earnings from 0% to 100% as much as fathers' earnings, and find that γ ranges from $[0.87, 1.37]$ using our IV specification.[34]

34. In 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.

TABLE VII: Curvature of Utility

	OLS	IV
$\mathbb{E}\left[\frac{I}{P}\right]$	0.43	0.25
γ	2.50	1.27

Note: $\gamma's$ are 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, estimates of $\mathbb{E}\left[\frac{I}{P}\right]$ are taken from Table VI above.

VI The Size of the Targeting Benefit for Progresa

In this section, we estimate the importance of the targeting benefit for Progresa villages. To do so, we first estimate the size of the targeting benefit relative to the exclusion cost under the observed average Progresa grants. We then calculate the optimal CCT/UCT mix assuming the only benefit of CCTs is our targeting benefit. Third, we show sensitivity of our results to three key parameters: the curvature of utility, parental income inequality, and the size of child incomes. Finally, to conclude the section, we provide a back-of-the-envelope estimate for the size of the targeting benefit relative to the enrollment benefit CCTs offer when parents under-value the return to schooling.

VI.A Size of the Targeting Benefit under Progresa Grants

We now estimate the size of the targeting benefit relative to the exclusion cost under the observed average Progresa transfer schedule of $\bar{t}_c = 6.86$ 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:

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

where N denotes our sample size; $S_{it}(\bar{t}_c, \bar{t}_u)$ denotes predicted enrollment for child i in year t under the average Progresa grants (where we predict $S_{it}(\bar{t}_c, \bar{t}_u)$ using our IV specification of Equation (12)); $\bar{S} = \frac{1}{N}\sum_{it} S_{it}(\bar{t}_c, \bar{t}_u)$ denotes average enrollment under average grants and is equal to 0.77; γ is taken from our IV specification in Table VII above, $\gamma = 1.27$; and, finally, average income and price effects are taken from our IV specification in Table VI above, $\bar{I} = 0.0035$, $\bar{P} = 0.013$.

Our estimate for $\frac{TB}{EC}$ is presented in column (1) of Table VIII below. We find a ratio equal to 0.91, 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 (if the only benefit of conditions is improved targeting), 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: as we vary mothers' earnings from 0% to 100% as much as fathers' earnings, we find that $\frac{TB}{EC}$ varies between $[0.89, 0.92]$.[35]

TABLE VIII: Size of the Targeting Benefit

	(1) Observed	(2) Optimal
t_c	6.86	5.27
t_u	2.90	4.23
Share of budget to CCT	64%	48%
TB/EC	0.91	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 Owing Solely to the Targeting Benefit

Next we calculate the optimal CCT assuming the only benefit of conditions is improved targeting. 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: $R = \bar{S}\bar{t_c} + \bar{t}_u = 0.77 \times 6.86 + 2.90 \approx 8$, where R represents the per-child budget (in pesos, per-capita, per-week).[36] 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:

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

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

where $S_{it}(t_c, t_u)$ denotes predicted enrollment of child i under the transfer schedule (t_c, t_u); $\bar{S}(t_c, t_u) = \frac{1}{N}\sum_{it} S_{it}(t_c, t_u)$ denotes the average share of children enrolled under the transfer schedule (t_c, t_u); $\bar{P}(t_c, t_u) = \frac{1}{N}\sum_{it} P_{it}(t_c, t_u)$ denotes the average price effect under the schedule (t_c, t_u); and $\bar{I}(t_c, t_u) = \frac{1}{N}\sum_{it} I_{it}(t_c, t_u)$ denotes the average income effect under the schedule (t_c, t_u). Note, to calculate $S_{it}(t_c, t_u)$, we use the following relationship:[37]

35. 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.

36. For example, if an eligible child aged 12-15 is from a family of size 6, we would estimate that Progresa has $6 \times 8 = 48$ pesos to offer this child unconditionally each week.

37. We recalculate the optimal transfers using our average price and income effect schedules to update S_{it} and find similar results (we do so because our heterogeneous income and price effect schedules are not estimated precisely). Specifically, we update the shares enrolled as follows: $S_{it}(t_c, t_u) = S_{it}(\bar{t}_c, \bar{t}_u) + (t_c - \bar{t}_c)\bar{P} + (t_u - \bar{t}_u)\bar{I}$. We find that 56% of the Progresa budget should be allocated towards a CCT.

$$S_{it}(t_c, t_u) = S_{it}(\bar{t}_c, \bar{t}_u) + \int_{\bar{t}_c}^{t_c} P_{it}(v, \bar{t}_u)dv + \int_{\bar{t}_u}^{t_u} I_{it}(t_c, v)dv$$

Results for the optimal transfers are presented in column (2) of Table VIII above. Our findings suggest that 48% of the Progresa budget should be allocated towards a CCT (compared to the observed proportion of 64%), with the optimal transfers equal to $t_c^* = 5.27, t_u^* = 4.23$ pesos, per-capita, per-week. It is important to re-emphasize that these results are driven purely by the targeting benefit (i.e., there are no other benefits of imposing conditions in this exercise). Including additional benefits of conditioning would simply increase t_c^* (similarly, including additional costs would decrease t_c^*). Lastly, our findings are robust to the magnitude of mothers' earnings: as we vary mothers' earnings from 0% to 100% as much as fathers' earnings, we find that the optimal share of the budget to a CCT varies between [46%, 49%].

Discussion 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. As shown in Table II, mean child income for boys is around 78% as large as mean income of fathers (similarly, our predicted measure of child income gives a mean child income for boys around 75% as large). Third, there is little difference in the density of per-capita parental incomes for those sending to school and those not sending to school. Highlighting this, Figure VII plots the density of per-capita parental income split by those sending their 14 year-old child to school and those not sending their 14 year-old child to school in our sample (under zero transfers).[38][39] 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 VIII plots the density of per-capita household income for those sending their 14 year-old child to school and those not sending their 14 year-old child to school, 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 over 3 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 the unconditional cash transfer. Reducing the unconditional cash transfer is particularly costly for the low parental income households

38. Note, figures look very similar when we look at households with a 12, 13, or 15 year-old child.

39. 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 $\frac{S_{it}(0,0)}{\bar{S}(0,0)}$, where $S_{it}(0,0)$ 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 $\frac{1-S_{it}(0,0)}{1-\bar{S}(0,0)}$.

who do not send their child to school.

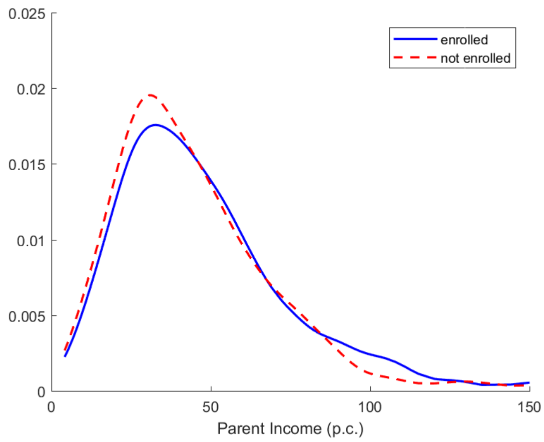

Figure VII: Distribution of Parent Income

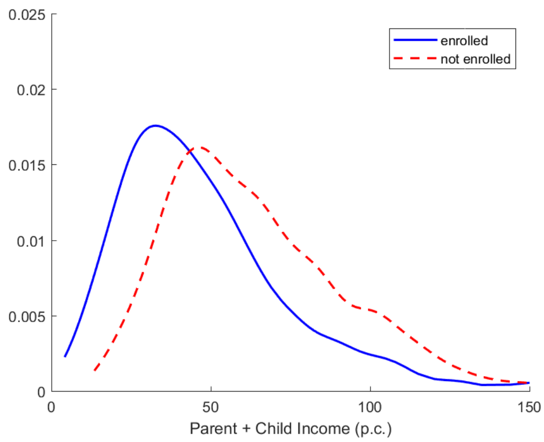

Figure VIII: Distribution of Parent + Child Income

VI.C 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.

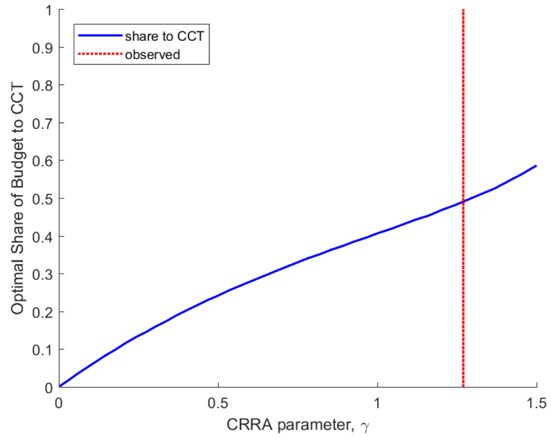

Figure IX: Varying Curvature of Utility

Varying Parental Income Inequality. Next we vary the mean difference in (per-capita) parental incomes between those parents sending to school and those parents not sending to school, $\mathbb{E}[y|s=1]-\mathbb{E}[y|s=0]$. To do so, we vary 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. Increasing this derivative will increase $\mathbb{E}[y|s=1]-\mathbb{E}[y|s=0]$ as $\mathbb{E}[y|s=1]=\frac{1}{N}\sum_{it=1}^{N}\frac{S_{it}y_{it}}{\bar{S}}$ and $\mathbb{E}[y|s=0]=\frac{1}{N}\sum_{it=1}^{N}\frac{(1-S_{it})y_{it}}{1-\bar{S}}$. Figure X plots the optimal share of the budget to a CCT vs. the mean difference in (per-capita) parental incomes for those sending to school and those not sending to school (under zero transfers). As expected by Proposition 5, the optimal CCT is decreasing in this difference.

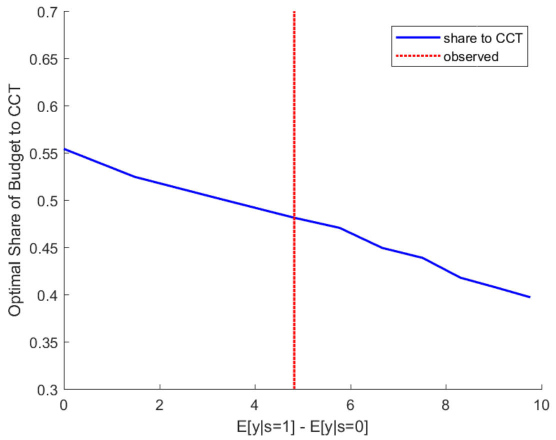

Figure X: Varying Parental Income Inequality

Varying Child Income. Finally, we 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 XI 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 XI is constructed holding enrollment decisions fixed).

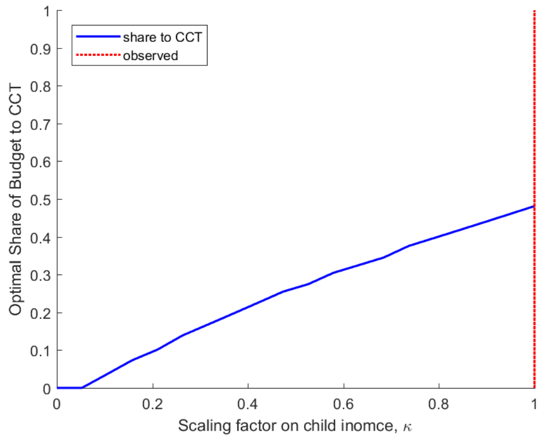

FIGURE XI: VARYING CHILD INCOME

VI.D 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. However, a common argument for offering CCTs over UCTs 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 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.6 for the planner's problem and derivation of the first order condition below):

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

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:

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

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).[40] Finally, we assume that annual utility of consumption is CRRA with parameter $\gamma = 1.27$. 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 discreteized version of Equation (15) under the average Progresa grants (for our sample of treated, eligible children in years 1998 and 1999):

40. 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.

$$\frac{TB}{EB} = \frac{\left(1 - \frac{\bar{S}+t_c\bar{P}}{1+\bar{t}_c\bar{I}}\right)\sum_{it=1}^{N}\left(y_{it} + \bar{t}_u + \bar{t}_c\right)^{-\gamma} S_{it}\left(\bar{t}_c, \bar{t}_u\right)}{N\left(\bar{P} - \bar{I}\left(\frac{\bar{S}+t_c\bar{P}}{1+\bar{t}_c\bar{I}}\right)\right)17.5\frac{(1.08\bar{y})^{1-\gamma}-(1.02\bar{y})^{1-\gamma}}{1-\gamma}} = 0.30$$

We find that under the average Progresa grants, the targeting benefit is 30% 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 $\frac{TB}{EB} = 0.87$, whereas if the actual return to an additional year of schooling is 12%, then $\frac{TB}{EB} = 0.18$. Thus, while the enrollment benefit dominates the targeting benefit, the targeting benefit is still comparable to the enrollment benefit, and thus should be taken into account when trading off the costs and benefits of CCTs vs. UCTs.

VII Conclusion

This paper argues that cash transfers made conditional on school attendance can better target transfers towards low consumption households relative to unconditional cash transfers. This is because sending a child to school can result in a discrete loss of child income, so that schooling may be negatively correlated with household consumption. By conditioning transfers on schooling, governments may be able to target transfers towards a group with lower consumption. We refer to this unexplored benefit of CCTs as the targeting benefit.

We formalize this intuition by developing a theoretical framework to model the targeting benefit associated with imposing conditions on schooling. We show that the targeting benefit alone can justify the planner allocating some or all of her budget to a CCT over a UCT. 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 (the cost associated with excluding households who do not comply with the conditions) in terms of empirically observable quantities. We show that the two relevant empirical quantities are the income effect schedule of a UCT and the price effect schedule of a CCT. These schedules 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 these schedules 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 48% of the Progresa budget should be allocated to a CCT over a UCT based on the targeting benefit alone. This implies that the targeting benefit can be a quantitatively important benefit of CCTs. Three key empirical factors are driving this finding: (1) forgone child incomes are large in this setting; (2) income differences between parents sending to school and parents not sending to school are small; and (3) we find substantial curvature in utility (our income and price effect estimates imply a coefficient of constant relative risk aversion of 1.27). 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, its 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 determining the optimal extent of conditioning. Second, it is important to have an idea on both the cost of schooling (e.g., potential child incomes) as well as the distribution of parental incomes because these factors influence the magnitude of the targeting benefit. And finally, although we abstract from under-investment motives in this paper so as to focus solely on the targeting benefit, it is important to know the extent to which private enrollment levels are below socially optimal levels. For example, if parents undervalue the returns to schooling and price effects are large relative to income effects, this alone may be enough to justify imposing conditions.

Combining all these factors together, Tables IX and X 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 IX). 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, on should offer a CCT (see Table X).

TABLE IX: Large Income Effects: $I \approx P$

	No under-enrollment	High under-enrollment
$cons_{sch} << cons_{no\ sch}$	CCT	CCT
$cons_{sch} >> cons_{no\ sch}$	UCT	UCT

TABLE X: Small Income Effects: $I << P$

	No under-enrollment	High under-enrollment
$cons_{sch} << cons_{no\ sch}$	UCT	CCT
$cons_{sch} >> cons_{no\ sch}$	UCT	CCT

Looking forward, we believe much of our analysis of 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): 12061240. 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): 488500.

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. Shaprio: "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, 229254.

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, 386414.

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 (2009): "Own and Sibling Effects of Conditional Cash Transfer Programs: Theory and Evidence from Cambodia," *World Bank Policy Research Working Paper*

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), 145161.

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 Progresa/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 Progresa 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 Proposition 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 $\frac{\partial(1-F(\tilde{\mu}(y)|y))}{\partial y} = -f(\tilde{\mu}(y)|y)\frac{\partial\tilde{\mu}(y)}{\partial y} + \frac{\partial(1-F(\tilde{\mu}|y))}{\partial y} \geq 0$.[41] 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 $t_c^* < y_c$.

A.4 *First Best*

We now assume the planner can observe both child ability μ, and parental income y. Thus, the planner can now offer transfers $t(y,\mu)$ (we still assume the planner can only distribute money, i.e., $t(y,\mu) \geq 0\;\forall\; y,\mu$). The planner solves the following:

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

$$
\begin{aligned}
\max_{t(y,\mu)} & \int_Y \int_M u\left(c^*(y,\mu)\right) + v(\mu, s^*(y,\mu))\ dF(\mu,y) \\
\text{s.t.} & \int_Y \int_M t(y,\mu) dF(\mu,y) \leq R \\
\& \ & t(y,\mu) \geq 0 \ \forall \ y, \mu
\end{aligned}
$$

We can write the Planner's Lagrangian as follows

$$
\int_Y \int_M u\left(c^*(y,\mu)\right) + v(\mu, s^*(y,\mu))\ dF(\mu,y) + \lambda \left(R - \int_Y \int_M t(y,\mu) dF(\mu,y) \right)
$$
$$
+ \int_Y \int_M \delta(\mu,y) \left(t(y,\mu) - 0\right) d\mu dy
$$

where λ denotes the Lagrange multiplier on the budget constraint, and $\delta(\mu,y)$ denotes the Lagrange multiplier on the non-negativity constraint. We now consider perturbing $t(y,\mu)$ by $d\tau$ on the intervals $[y, y+\epsilon]$, $[\mu, \mu+\epsilon]$. 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 schedule $t(y,\mu)$ must satisfy the following condition:

$$
\epsilon^2 d\tau \left(u_c(c^*(y,\mu)) - \lambda + \delta(\mu,y)\right) = 0
$$

Thus, $t(y,\mu)$ either solves $u_c(y - ks^*(y,\mu) + t(y,\mu)) = \lambda$ or $t(y,\mu) = 0$. Therefore, marginal utility of consumption is equated for everyone who gets a non-zero transfer. It's easy to show that the people who get transfers are those with the highest marginal utilities of consumption. In other words, the first best solution is to keep giving money to the highest marginal utility households until we run out of money, or until marginal utilities are equated across all households.

A.5 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:

$$\max_{t_u(y),t_c(y)} \int_Y \int_{\tilde{\mu}} u\left(y + t_u(y) + t_c(y)\right) + v(\mu, 1)\ dF(\mu, y)$$

$$\int_Y \int^{\tilde{\mu}} u\left(y + y_c + t_u(y)\right) + v(\mu, 0)\ dF(\mu, y)$$

$$\text{s.t.} \int_Y t_u(y)dF(y) + \int_Y \int_{\tilde{\mu}} t_c(y)dF(\mu, y) \leq R$$

$$\&\ t_c(y) \geq 0\ \forall\ y$$

$$\&\ t_u(y) \geq 0\ \forall\ y$$

We can write the Lagrangian as follows

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

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:

$$\epsilon d\tau \left(u_c(y + t_u(y) + t_c(y))S(y) + u_c(y + y_c + t_u(y))(1 - S(y)) - \lambda\left(1 + t_c(y)I(y)\right) + \delta_2(y)\right) = 0 \tag{16}$$

$$\epsilon d\tau \left(u_c(y + t_u(y) + t_c(y))S(y) - \lambda\left(S(y) + t_c(y)P(y)\right) + \delta_1(y)\right) = 0 \tag{17}$$

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

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

And from Equation (17) we get

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

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.6 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:

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

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

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

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:

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

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:

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

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).[42] 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.

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

A.7 Proof of Proposition 3

Proof. 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. □

A.8 Proof of Corollary 3.1

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:

$$\frac{G'(6 - \delta)}{G'(6 + \delta)} < \frac{(6 - \delta + 10)^{1/2}}{(6 + \delta + 3)^{1/2}} < (16/9)^{1/2} \tag{21}$$

we will have:

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

Inequality (21) 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 $t_c = 0$:

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

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.9 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

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

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.10 Proof of Proposition 5

Proof. First, let us denote $\mathbb{P}(s^* = 1|y) = \int_{\tilde{\mu}(y)} f(\mu|y)d\mu = S(y)$ 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:

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

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

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

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 $\text{TB}(t_c^*,f^{(2)})\geq\text{TB}(t_c^*,f^{(1)})$. Morever, 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 $\text{EC}(t_c^*,f^{(2)})\geq\text{EC}(t_c^*,f^{(1)})$, 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.11 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 assumption: $\frac{\partial^2 y_2}{\partial\mu\partial s}>\frac{-v''}{v'}\frac{\partial y_2}{\partial\mu}\frac{\partial y_2}{\partial s}$. In words, schooling and ability are sufficiently strong complements.[43] This assumption implies schooling is weakly increasing in ability μ. We can now derive the following proposition.

43. 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.

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

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.[44] Thus, there exists a $\tilde{\mu}(y)$ s.t. households with income y and $\mu < \tilde{\mu}(y)$ do not send to school and vice versa.

We now differentiate the Planner's problem with respect to t_u, using the budget constraint to write $t_c(t_u)$, and get the following FOC

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

where $c^*(1;y,\mu)$ ($c^*(0;y,\mu)$) is the optimal consumption level of a household with parental income y and child ability μ if they send to school (do not send to school). 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. E.g. $b^*(1;y,\mu)$ solves: $u_c(y + t_c + t_u + b) = v_c(y_2(\mu,1) - rb)$. Next,

$$\frac{\partial t_c}{\partial t_u} = \frac{-1}{\int_Y \int_{\tilde{\mu}} f(\mu,y)d\mu dy}\left(1 - t_c\left(\int_Y \frac{\partial\tilde{\mu}}{\partial t_u} f(\tilde{\mu}|y)f(y)dy\right)\right)$$

If $t_c = 0, t_u = R$ the FOC for t_u becomes:

$$-\frac{1}{\int_Y \int_{\tilde{\mu}} f(\mu,y)d\mu dy}\int_Y \int_{\tilde{\mu}(y)} u_c\left(c^*(1;y,\mu)\right) f(\mu,y)d\mu dy +$$
$$\frac{1}{\int_Y \int^{\tilde{\mu}} f(\mu,y)d\mu dy}\int_Y \int^{\tilde{\mu}(y)} u_c\left(c^*(0;y,\mu)\right) f(\mu,y)d\mu dy$$

But this is just the difference in the expected marginal utility between those who do not send to school and those who do. But this is necessarily positive as a household will only send to school if doing so increases their consumption today $c^*(1;y,\mu) > c^*(0;y,\mu)\ \forall\ y,\mu$.[45] Thus, marginal utility is higher for those who do not enroll. Thus, $t_c = 0$ is optimal. □

By allowing households to borrow across generations, they are able to smooth their consumption across the two periods (the household first order condition gives $u_c(c) = v_c(c_2)$). Thus, households will only send to school if doing so increases their total utility, i.e., schooling increases their consumption in both generations. Therefore, it can never be optimal to transfer

44. $\frac{\partial W}{\partial\mu} = \beta v'(c_2(s))\frac{\partial y_2(\mu,s)}{\partial\mu}$ which is increasing in s by our assumption. Hence $\frac{\partial W}{\partial\mu}|_{s^*=1} > \frac{\partial W}{\partial\mu}|_{s^*=0}$; $\frac{\partial^2 W}{\partial b^*\partial\mu} = 0$; and $\frac{\partial W}{\partial b^*} = 0\ \forall\ s$ by our household's FOC for b, where W denotes the maximand of the Planner's problem.

45. Household's FOC for b gives $u_c(c^*(s)) = v_c(c_2^*(s))$ where c is consumption in the first generation, and c_2 is consumption in the second generation. A household would only ever send to school if doing so increases their total utility: $u(c^*(1)) + v(c_2^*(1)) \geq u(c^*(0)) + v(c_2^*(0))$. By the FOC for b we know if c increases, c_2 must increase. Thus, if a household sends to school it must be true that $c^*(1) \geq c^*(0)$.

more to those sending to school, as that would result in transferring to households with higher levels of consumption.

A.12 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

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

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.13 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:

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

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:

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

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:

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

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

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

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:

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

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:

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

Notably, 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 $\frac{\partial t_{c1}}{\partial t_{u1}} \geq -1$; see Proposition **??** above).

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.14 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:

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

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 $y = nl$:

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

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:*

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

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:

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

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. □

A.15 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., $y_t = y$ $\forall$ $t \in 1, 2, ..., T$, and that their 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 $V(t+1,\mu) > V(t,\mu)$ $\forall$ $t \in 0, ..., T-1$, $V_2(t,\mu) > 0$, 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 $s_{t+n}^* = 0$ $\forall$ $n \in 1, ..., T-t$. 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 $n \in 1, ..., T-t$. 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 $s_t = s_t^* \ \forall \ i = 1, .., m-1, m+2, ...T$, $s_m = 1$, $s_{m+1} = 0$ 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. □

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

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

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 $i \in 1, ..., m - 1$ and $\mu \leq \tilde{\mu}_m \implies \mu < \tilde{\mu}_i$ for $i \in m + 1, ..., T$.[46] Thus, the set of households sending to school in year m will have $\mu \geq \tilde{\mu}_m$.

46. 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 $u(y + t_u + y_{cm+1}) - u(y + t_u + t_c) < V(m + 1, \mu) - V(m, \mu)$. 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.

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 $b_0, ..., b_0 + B$, and will offer each child in these cohorts transfers for their full T years of schooling. Thus, the planner will offer transfers from years $b_0, ..., T + b_0 + B$. 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 $m \in \{b, ..., T+b\}$ in which they stop sending their child to school:

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

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 $m + 1$:

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

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 follows[47]

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

where $F(\mu, y, t, b)$ denotes the joint CDF of child abilities, parental incomes, cohorts, and

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 $i \in 1, ..., m$. One can easily repeat a similar exercise to show $\mu \leq \tilde{\mu}_{m+1} \implies \mu < \tilde{\mu}_i$ for $i \in m + 2, ..., T$.

47. 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.

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:

$$\sum_{b=b_0}^{b_0+B}\sum_{t=b}^{T+b}\Bigg(\left(1+\frac{\partial t_u}{\partial t_c}\right)\int_Y u_c(y+t_u+t_c)S(y,t,b)dF(y,t,b)+ \\ \frac{\partial t_u}{\partial t_c}\int_Y u_c(y+t_u+y_c(t-b))(1-S(y,t,b))dF(y,t,b)\Bigg)=0 \tag{22}$$

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 (22) is simply the average targeting benefit across all cohorts and years plus the average exclusion cost across all cohorts and years.

A.16 Sufficient Statistics for Annual Schooling Decisions Model

As with our baseline model, we can show in our annual schooling decisions model (discussed above) that the ratio of the income to price effect still allows us to determine the curvature of utility. First, the income effect and price effect for parents with income y, a child in cohort b, in year t is given by:

$$I(y,b,t)=\frac{\partial S(y,b,t)}{\partial t_c}=\frac{u_c(y+t_u+t_c)-u_c(y+t_u+y_c(t-b))}{V_2(t-b+1,\tilde{\mu})-V_2(t-b,\tilde{\mu})}dF(\tilde{\mu}|y,t,b)$$

$$P(y,b,t)=\frac{\partial S(y,b,t)}{\partial t_c}=\frac{u_c(y+t_u+t_c)}{V_2(t-b+1,\tilde{\mu})-V_2(t-b,\tilde{\mu})}dF(\tilde{\mu}|y,t,b)$$

Taking the ratio we get

$$\frac{I(y,b,t)}{P(y,b,t)}=1-\frac{u_c(y+t_u+y_c(t-b))}{u_c(y+t_u+t_c)}$$

Thus, if we observe $I(y,b,t)$, $P(y,b,t)$, $S(y,b,t)$, and the cdf $F(y,t,b)$, we can determine the optimal t_c^* (via solving Equation (22)). If we make assumption that the density of parental incomes and child abilities is constant across cohorts and years, and that cohorts are the same size, we can rewrite our FOC as follows:

$$\frac{1}{TB}\sum_{b=b_0}^{b_0+B}\sum_{t=b}^{T+b}\Bigg(\left(1+\frac{\partial t_u}{\partial t_c}\right)\int_Y u_c(y+t_u+t_c)S(y,t-b)dF(y)+ \frac{\partial t_u}{\partial t_c}\int_Y u_c(y+t_u+y_c(t-b))(1-S(y,t-b))dF(y)\Bigg)=0 \tag{23}$$

where $\frac{1}{TB}\sum_b\sum_t\int_Y\int_M dF(y,\mu) = 1$, where $F(y,\mu)$ denotes the density of incomes and abilities for a given cohort in a given year; and, where $S(y,t-b)=\int_{\tilde{\mu}(y,t-b)} dF(\mu|y)$ denotes the share of households with parental income y sending their child of age $t-b+6$ to school (assuming school starts at age 6). Thus, now the sufficient statistics are: $I(y,t-b)$, $P(y,t-b)$, $S(y,t-b)$, and $F(y)$. In words, we need to observe the income effects, price effects, and enrollment shares for each parental income level at each child age, along with the density of parental incomes. Note, if we also think child income is not only a function of age, but also a function of gender, i.e., $y_c(t-b, boy)$, our sufficient statistics would now be functions of parental income, child age, and child gender, e.g., $I(y,t-b,boy)$ etc.

B Appendix

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

	(1)
mean hourly wage in locality	1.688***
	(0.350)
$\mathbb{E}[y^f]$	26.71
Observations	17691
R-squared	0.00958

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^f_{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. 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: $t_{c,it}$, $t_{u,it}$, indicators for age of child i, indicators for highest grade completed of child i, age interacted with highest grade completed, year fixed effects, and child fixed effects.

B.2 Calculating Moments of Income and Price Effects

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

$$\mathbb{E}\left[P(X)\right]=\frac{1}{N}\sum_{it=1}^{N}\hat{b}_1+2\hat{b}_3t_{c,it}+\hat{b}_5t_{u,it}+\hat{b}_6y^f_{it}+\hat{b}_8age_{it}+\hat{b}_{10}boy_i$$

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:

$$\text{Var}\left(P(X)\right) = 4\tilde{b}_3^2\text{Var}(t_{c,it}) + \tilde{b}_5^2\text{Var}(t_{u,it}) + ... + 4\widetilde{b_3 b_5}\text{Cov}(t_{c,it}, t_{u,it}) + ... + 2\widetilde{b_8 b_{10}}\text{Cov}(age_{it}, boy_i)$$

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):

$$\begin{aligned}\log(hw_{ilt}) =& \alpha_1 + \alpha_2 age_{it} + \alpha_3 hgc_{it} + \alpha_4 boy_i + \alpha_5 log(town\_hw_{lt}) + \alpha_6 eligible_i + \alpha_7 control_l + \\ & \alpha_8 \mathbb{1}(year = 98)_t + \alpha_9 \mathbb{1}(year = 99)_t + \alpha_{10}\left(control_l \times \mathbb{1}(year = 98)_t\right) + \\ & \alpha_{11}\left(control_l \times \mathbb{1}(year = 99)_t\right) + e_{it}\end{aligned} \tag{24}$$

where hw_{ilt} 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 XI. 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). Not surprisingly, we also find that child wages increased in localities that received the Progresa grants (this is shown by the negative and significant coefficients on $year98 \times control$ and $year99 \times control$). This is consistent with a reduction in child labor supply in these localities after 1997. Prior to the grants, we find no evidence to suggest children in control localities received different wages to those in treatment localities (shown by the insignificant coefficient on *control*), and we find no evidence to suggest that those children from eligible households received different wages to those from non-eligible households.

TABLE XI: Mincer Regression

	(1)
age	0.0445***
	(0.006)
hgc	0.00373
	(0.004)
boy	0.0826***
	(0.020)
log(town wage)	0.791***
	(0.039)
eligible	-0.0175
	(0.021)
control	0.0193
	(0.026)
$\mathbb{1}(year = 98)$	0.0215
	(0.022)
$\mathbb{1}(year = 99)$	0.0163
	(0.023)
$\mathbb{1}(year = 98) \times$ control	-0.0628*
	(0.036)
$\mathbb{1}(year = 99) \times$ control	-0.0617*
	(0.035)
Observations	4454
R-squared	0.244

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.

Using Equation (24), 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.[48] Average potential incomes for 12-15 year-old boys and girls are presented in Table XII below.

TABLE XII: Summary Statistics: Potential Child Incomes

Children aged 12-15	Mean	Std. Dev.
Predicted income (boy)	126.9	31.6
Predicted income (girl)	116.9	28.8

Note: Average weekly predicted incomes in pesos.

48. 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.

B.4 Heterogeneous Income and Price Effects: Equation (12)

TABLE XIII: HETEROGENEOUS INCOME AND PRICE EFFECTS

	(1)	(2)
tc	0.0416**	0.0440*
	(2.03)	(1.96)
tu	0.00868	0.0159
	(0.25)	(0.42)
tc^2	-0.000158	0.000328
	(-0.47)	(0.70)
tu^2	-0.00144*	-0.00134
	(-1.70)	(-1.43)
tc× tu	-0.000317	-0.000177
	(-0.45)	(-0.23)
tc× y^f	-0.0000503	-0.000370**
	(-1.30)	(-2.18)
tu× y^f	0.0000920	-0.0000676
	(1.06)	(-0.17)
tc× age	-0.00202	-0.00184
	(-1.34)	(-1.13)
tu× age	0.000283	-0.000115
	(0.12)	(-0.05)
tc× boy	-0.00113	-0.00110
	(-0.45)	(-0.43)
tu× boy	-0.000516	-0.0000871
	(-0.11)	(-0.02)
y^f	0.000162	0.00346
	(0.49)	(0.66)
Observations	17365	17279
R-squared	0.146	0.131
Pvalue on interactions	0.374	0.0820

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). Column (1): OLS; column (2): instrument father's per-capita income and all interaction terms with father's income with average locality wage (excluding the wage of the father in question). Regressors not shown: highest grade completed dummies, age dummies, grade interacted with age, year fixed effects, and child fixed effects.