

The Targeting Benefit of Conditional Cash Transfers *

Katy Bergstrom[†]

William Dodds[‡]

[CLICK HERE FOR LATEST VERSION](#)

November 25, 2018

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 the poorest households, as access is tied to the consumption of normal goods. However, we argue that conditionalities based on children's schooling may actually improve the targeting of transfers to low consumption households. Sending a child to school can result in a discrete loss of child income, so that schooling is negatively correlated with household consumption. Thus, schooling decisions may act as a useful "tag" for cash transfers. 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, 55% of the Progresa budget should go to a CCT over a UCT.

Keywords: conditional cash transfer, unconditional cash transfer, targeting benefit, exclusion cost, income effect, price effect.

*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.

[†]Department of Economics, Stanford University. Email: katyberg@stanford.edu.

[‡]Department of Economics, Stanford University. Email: wdodds@stanford.edu.

1 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 1 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.

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

²Given 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)).

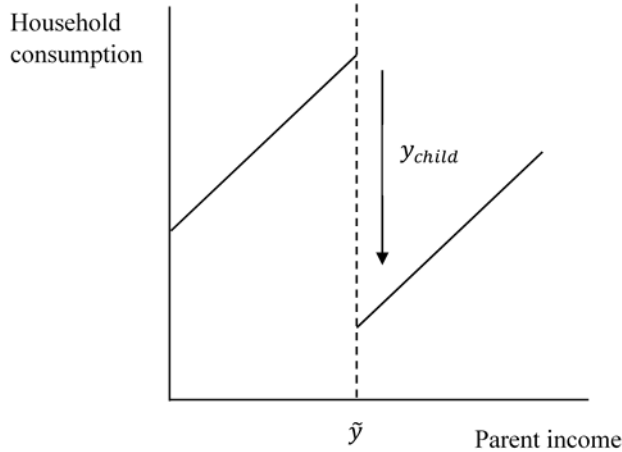


Figure 1: Household consumption versus parent income

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

³While 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. We show how changing the distribution of parental incomes in this set affects the importance of the targeting benefit.

⁴Because the planner only has a budget to redistribute in the first generation, maximizing total lifetime utility over both generations amounts to transferring towards those with the lowest consumption (highest marginal utility of consumption) today. 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).

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, freeing up resources for higher transfers to eligible beneficiaries in the next period. We show that in a dynamic transfer scheme, under some assumptions around the returns to schooling, 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 important 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, the targeting benefit can never exceed the exclusion cost; hence it can never be optimal to allocate any of a budget towards a CCT based on the targeting benefit alone. With linear utility, marginal utilities are constant across all households, thus eliminating any desire for the planner to target transfers towards specific households. Second, we show that as parental income inequality increases among the eligible set of households, the targeting benefit falls while the exclusion 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 show that the size of the targeting benefit relative to the exclusion cost can be written 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) there is substantial curvature in the return to schooling. However, if (2) and (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 with the dual objectives of alleviating current poverty and increasing children’s human capital to reduce the transmission of poverty to the next generation. 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: 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 such that 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 and price 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 1.37.

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 89% 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 55% 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).⁶ 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 2 discusses our contribution to the literature, Section 3 sets-up our theoretical framework and derives our main theoretical results, Section 4 develops our sufficient statistics approach to estimating the size of the targeting ben-

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

⁶I.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.

efit, Section 5 estimates the income and price effects for Progresa, Section 6 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 7 concludes.

2 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), and 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), and 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), and 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 schooling 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 (Schultz 2004; de Janvry et al (2006), Filmer and Schady (2011), Case et al (2005), Edmonds (2006), Edmonds and Schady (2012), and many others). Importantly, our novel identification strategy allows us to estimate these elasticities in the same setting using a reduced-form specification.⁷ While Todd and Wolpin (2006) and de Brauw and Hoddinott

⁷Other 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.

(2011) estimate both behavioral responses in the setting of Progresa, the former uses a model-based simulation exercise to do so, while the latter can only pin down income effects relative to price effects. Finally, we add to the literature that investigates alternative reforms of the Progresa program. Todd and Wolpin (2006) and Attanasio et al (2011) both examine the effects of counter-factual policies on school enrollment. Our paper differs in that we investigate the optimal allocation of the budget towards a CCT vs. UCT so as to best alleviate current poverty. In doing so we highlight the role conditions can play in improving the targeting of cash transfers.

3 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. 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 3.3 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 3.3). 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 (with this share usually rising with parent income), suggesting additional heterogeneity is important. The parents' problem can be written as follows

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

where $u(c)$ denotes household utility over consumption c in the first generation (where $u_c > 0, u_{cc} \leq 0$), y denotes heterogeneous parental income, s denotes whether parents decide to

send to school or not, and y_c denotes child income (assumed to be 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 which 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 this section, we will refer to μ as child ability.

We make the following functional form assumptions on $v(\mu, s)$: first, v is a positive function of child ability and schooling, namely $v_1(\mu, s) > 0$ 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 3.1. *Schooling, s^* , is weakly increasing in ability type μ , $\frac{\partial s^*(\mu, y)}{\partial \mu} \geq 0$.*

Proof. A household will send to school iff $u(y) + v(\mu, 1) \geq u(y + y_c) + v(\mu, 0)$. Since we assume $v_1(\mu, 1) > v_1(\mu, 0)$, $s^*(\mu, y)$ is weakly increasing in μ (by Topkis). \square

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

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

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

Proposition 3.2. *If Assumption 1 holds, $\mathbb{E}[y|s^* = 0] \leq \mathbb{E}[y|s^* = 1]$.*

Proof. See Appendix A.1. \square

Proposition 3.2 highlights an undesirable feature of conditional cash transfers: higher income parents are more likely to meet the conditions and therefore receive the transfer. This is consistent with the view that CCTs are conditioned on the consumption of normal goods (Rodríguez-Castelán, 2017). We now describe the social planner problem.

3.1 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).⁸ Within the set of eligible households we assume the planner cannot condition transfers on parental income or child ability, but can condition on schooling decisions.^{9,10} Thus, she 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 to distort enrollment decisions.¹¹ We can write the social planner's problem as follows

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

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

⁸In reality, the set of eligible households is also a choice variable for the planner. We show later on how changing the distribution of parental incomes of the eligible set of households affects the size of the targeting benefit (see Proposition 3.5).

⁹I.e., we are operating in a second-best world. In a first-best world, 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).

¹⁰In Appendix A.5 we consider the extension where the planner can condition transfers on parental income and schooling decisions within the set of eligible households, but cannot observe ability μ . Thus, the planner can now offer UCTs and CCTs that vary with parental income: $t_u(y)$ and $t_c(y)$. In this setting, it is always optimal to offer a CCT, i.e., $t_c(y) > 0$ for some y . The reason we assume the planner cannot condition on parental income is for realism given the difficulty in verifying incomes in developing countries.

¹¹First, we shut down any under-enrollment motives to impose conditions by assuming the social planner problem coincides with the parent problem of maximizing total household utility. Second, we shut down any political economy motives to condition by assuming the planner has a fixed, exogenous budget to redistribute.

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

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

$$u(y + t_u + t_c) + v(\tilde{\mu}, 1) = u(y + y_c + t_u) + v(\tilde{\mu}, 0) \tag{2}$$

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

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

where $\frac{\partial \tilde{\mu}}{\partial t_c}$ is the total derivative w.r.t. t_c , i.e., $\frac{\partial \tilde{\mu}(y, t_c, t_u(t_c))}{\partial t_c} = \frac{\partial \tilde{\mu}}{\partial t_c} + \frac{\partial \tilde{\mu}}{\partial t_u} \frac{\partial t_u}{\partial t_c}$.

The first term in Equation 1, 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 1, 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 $\frac{\partial t_u}{\partial t_c}$ 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 $\frac{\partial t_u}{\partial t_c} < 0$.¹² It is worth mentioning that the utility change experienced by the households who are now induced to invest is second order (by the envelope theorem) given these households were near indifferent prior to the change in t_c .

Thus, 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

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

in the next subsection, this can occur if the parents sending to school do not earn substantially more than the parents not sending to school, and if the cost of schooling (e.g., child income) is substantial relative to parental income, leading to the enrolled households having, on average, lower household consumption. Finally, it is worth noting that if the planner were to value increased enrollment (e.g., parents undervalue the return on schooling), this would work alongside the targeting benefit in the sense that now the planner would offer a CCT if the combination of the targeting benefit plus the additional gain in enrollment from offering a CCT over a UCT outweighed the exclusion cost (see Appendix A.6 for a model with parents undervaluing the return to schooling). Alternatively, if CCTs induced additional administrative costs over UCTs, then this too could easily be included in the above model (e.g., if CCTs induced an additional fixed cost, C , the planner's budget constraint could be expressed as $t_u + t_c \int_Y \int_{\bar{\mu}} dF(y, \mu) = R - C\mathbb{1}(t_c > 0)$).

3.2 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, we proceed with the following example. 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. \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 is illustrated in Figure 2 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 and therefore lower their marginal utility of consumption by a greater amount. However, this comes at the cost of missing out on transferring to those households with very low parental incomes who find it too costly to send their child to school (Figure 3 illustrates the effect of a pure UCT and a pure CCT on marginal utility of consumption, respectively).¹³ Depending on the size of jump up in marginal utility (which in turn will depend on the cost of schooling, y_c , and the concavity of utility) and the distribution of parental incomes, it may be the case that it is beneficial to target transfers towards the enrolled households. This leads us to our main proposition.

¹³Moreover, 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 3). This is viewed as costly in our set-up as the planner has no incentive to distort enrollment decisions.

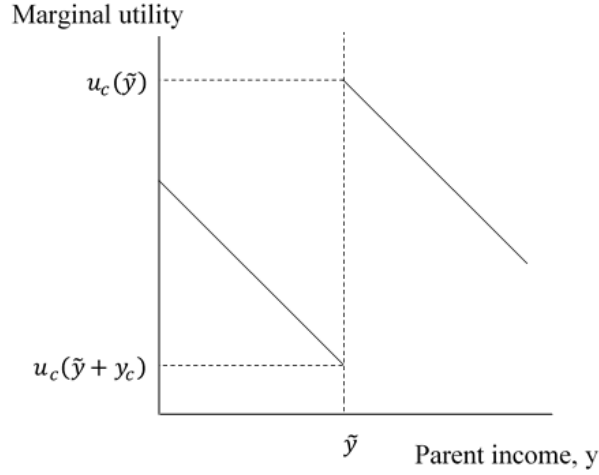


Figure 2: Discontinuity in marginal utility of consumption (constant μ)

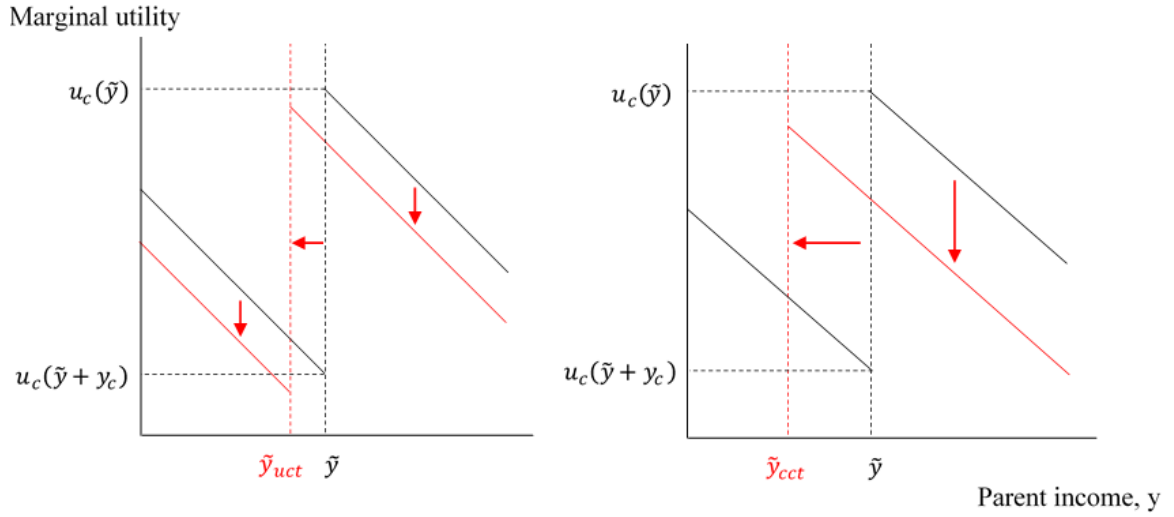


Figure 3: Pure UCT (left) vs Pure CCT (right)

Proposition 3.3. *If utility is concave in consumption, a pure UCT is not necessarily optimal.*

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. □

Moreover, we can extend the logic of Proposition 3.3 to show that a pure UCT is not necessarily optimal within a large class of social welfare functions.

Corollary 3.3.1. *If utility is concave in consumption, a pure UCT is not necessarily 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.7 □

Corollary 3.3.1 highlights that the targeting benefit can still be an important benefit of CCTs over UCTs under social welfare functions that place a greater weight on redistributing towards those with lower lifetime utility (e.g., log utilitarian). Notably, Corollary 3.3.1 relies on $G(\cdot)$ being everywhere positive and continuous. These requirements rule 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 for any given parental income level, the set of households sending to school must have higher lifetime utility than the set of households not sending to school. Corollary 3.3.1 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 3.4. *If utility is linear in consumption, a pure UCT is optimal.*

Proof. See Appendix A.8. □

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 useful.

Proposition 3.5. *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^* .¹⁴ Then under $f_2(y, \mu)$, the optimal local reform starting from t_c^* is to increase t_c .*

Proof. See Appendix A.9. □

The idea behind Proposition 3.5 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 4 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 5 translates these parent income densities into household income densities. 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 3.5 is closely related to parental income inequality; because schooling is a normal good (see Proposition 3.2), 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 3.5, 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

¹⁴ 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 .

poorer (lower parental income) households. This will work in favor of reducing the targeting benefit, making a comparative static in y_c unfeasible.

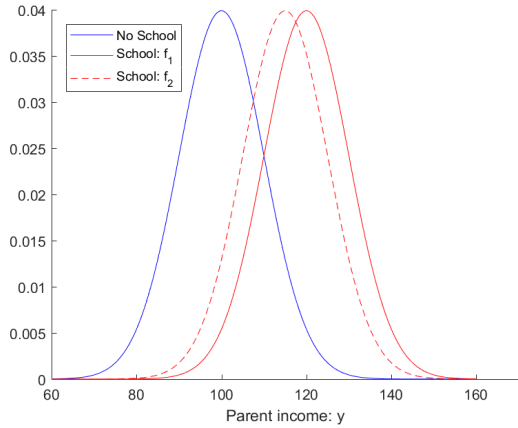


Figure 4: Dist. of parent income

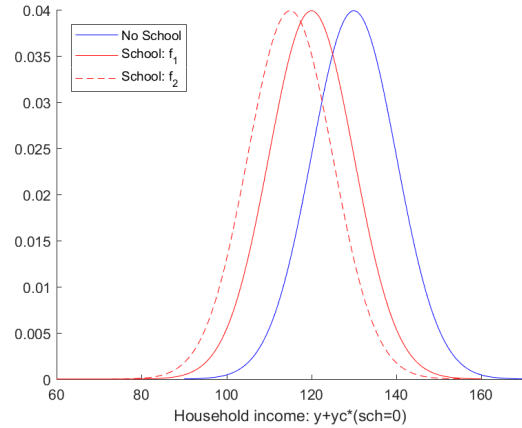


Figure 5: Dist. of household income

3.3 Extensions

We consider five extensions of our baseline model. 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.10). While in practice these markets do not exist, such an exercise is useful to illustrate that our main result, Proposition 3.3, relies crucially on the lack of these markets. We show that when households are able to borrow across generations, a pure UCT is optimal. This is because households are now able to 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, in this world, the households that send to school must have higher consumption today, implying it cannot be optimal to target transfers towards them.

Second, we consider a more realistic extension where we introduce 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 (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; see Appendix A.11 for model set-up and results). 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 consider the extension where the planner has a fixed budget in both the first

and second generation to redistribute (see Appendix A.12). As before, adults in the second generation (i.e., the children in the first generation) earn an income that depends on their ability and whether they went to school as a child. Now, however, these adults have their own child and choose whether or not to send their child to school in the second generation. The planner gets to choose the optimal mix of a CCT and UCT in both the first and second generation to maximize total lifetime utility of households (where this mix can differ across the two generations). Moreover, we assume that the adults in the second generation who went to school when they were children earn high enough incomes so that they are ineligible to receive transfers in the second generation (note, parents in the first generation take this into account when choosing whether to send to school or not). We show that the first order condition for the CCT in the first generation is the same as in Equation 1; however, there is one additional term capturing how increasing the CCT today reduces the set of eligible households tomorrow. Like the targeting benefit, this term works in favor of increasing the CCT in the first generation. Importantly, we are able to show that the total amount of money received across both generations by households that do not send to school in the first generation is greater than the total money received by the households that do send to school in the first generation. Thus, taking into account transfer schemes in later generations can help mitigate the concern that we are on net transferring more to higher lifetime utility households.

Fourth, we incorporate 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.13 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.

Finally, we consider the extension where parents make annual, binary schooling decisions over a set number of T years (see Appendix A.14). 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.

4 Sufficient Statistics for the Size of the Targeting Benefit

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). Moreover, this method will also allow us to determine the optimal CCT based solely on the targeting benefit (i.e., the optimal CCT assuming the only benefit of imposing conditions is improved targeting). While in theory we can construct examples where the targeting benefit is substantial, empirically

we do not know if it is. 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:

$$\frac{TB}{EC} = - \frac{\left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y u_c(y + t_u + t_c) S(y) f(y) dy}{\frac{\partial t_u}{\partial t_c} \int_Y u_c(y + y_c + t_u) (1 - S(y)) f(y) dy} \quad (3)$$

where $S(y) = \int_{\tilde{\mu}(y)} f(\mu, y) d\mu$ denotes the proportion of households with parent income y sending their child to school, remembering $\tilde{\mu}(y)$ denotes the household with parental income y just indifferent between sending to school and not (see Equation 2). Note for simplicity of notation we have suppressed that $S(y)$ is a function of the transfer schedule (as $\tilde{\mu}(y)$ is a function of the transfer schedule). It can be seen from Equation 3, that in order to determine the relative size of the targeting benefit, one needs to know the curvature of utility. This is intuitive, as this ratio is essentially capturing 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 2) with respect to t_u and t_c , we obtain explicit formulas for the income and price effects:

$$I(y) = \frac{u_c(y + t_u + t_c) - u_c(y + y_c + t_u)}{v_1(\tilde{\mu}, 1) - v_1(\tilde{\mu}, 0)} f(\tilde{\mu}(y)|y) \quad (4)$$

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

Taking the ratio of Equations 4 and 5 we get:

$$\frac{I(y)}{P(y)} = 1 - \frac{u_c(y + y_c + t_u)}{u_c(y + t_u + t_c)} \quad (6)$$

From Equation 6 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. The greater this ratio, the greater the discontinuity in marginal utility of consumption, i.e., the greater the curvature in utility of consumption. Notably, for a given schedule (t_c, t_u) , this relationship holds for all parental income levels y . Therefore, if we can observe the income and price effect schedules under the given transfer schedule, we can pin down the curvature of utility of consumption. As an example, consider the case where utility of consumption is CRRA with coefficient of relative risk aversion γ . To determine γ , one only needs to observe the expected value of the ratio of income to price effects as γ solves:

$$\mathbb{E}_Y \left[\frac{I(y)}{P(y)} \right] = 1 - \mathbb{E}_Y \left[\frac{(y + y_c + t_u)^{-\gamma}}{(y + t_u + t_c)^{-\gamma}} \right] \quad (7)$$

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, i.e., 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 there is substantial curvature in the return to schooling (this can be seen by the terms $f(\tilde{\mu}|y)$ and $v_1(\tilde{\mu}, 1) - v_1(\tilde{\mu}, 0)$ in Equation 4, respectively). Fortunately, however, if there is a large mass of indifferent households, or if there is substantial curvature in the return to schooling, 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 5 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 “curvature of 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, the only unknown quantity in Equation 3 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 $\frac{\partial t_u}{\partial t_c}$:

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

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.¹⁵ 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, $TB/EC > 1$, 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.15). 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 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.

5 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

¹⁵One also needs to observe the cost of schooling y_c and the size of the budget R .

income and price effects, using evaluation data from Progresa. More specifically, we focus on estimating the expected value of the ratio of these two effects so as to pin down the coefficient of constant relative risk aversion. First, we will briefly discuss the Progresa program. Second, we will describe key facts about child incomes, child labor supply, and suggestive 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.

5.1 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 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 and vaccinate their children, as well as 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 1 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 then targeted via proxy-means testing, and split into two groups: poor and non-poor. Poor households were eligible for Progresa transfers while non-poor households were ineligible.¹⁶ The program was initially implemented as follows: 506 of the targeted localities were randomly chosen across 7 states, and of these, 320 were randomly chosen to be offered the transfers immediately, starting May 1998. We refer to these 320 localities as the treated localities. The other 186 localities, the control localities, started the program 2 years later. Thus, eligible households in the treated localities started receiving transfers in May 1998, while eligible households in control localities did not start receiving transfers until May 2000.

Prior to the start of the program, a comprehensive survey was carried out in September

¹⁶See Skoufias, Davis, and Behrman (1999) for more information on targeting of localities and households.

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 covers approximately 24,000 households over 1997, 1998, and 1999.

Table 1: Progresa Transfer Amounts (bimonthly, in pesos)

	1998 1st semester	1998 2nd semester	1999 1st semester	1999 2nd semester
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

5.2 Key facts on child labor supply, child income, and the discreteness of schooling

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 2 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 job. Conversely, of those children who report they attend school, only 7.5% (3%) of boys (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.¹⁷ Thus, it appears that going to school places a

¹⁷We assume if a child reports to be working 30 or more hours a week, that they are not in school. This is a minor assumption as very few children report being in school working 30 or more hours a week.

major constraint both on labor market participation and on hours worked in the labor market for children.

Table 2: 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
Estimates using SP (2000): all children aged 12		
Hours worked per-day (not enrolled)	7.56	
Hours in school per-day (enrolled)	7.15	

Note: Sample: all households in 1997 with two parents where one parent reports to be the head, and with at least one child aged 12-15 years. Has job takes value 1 if individual reported to work in a paid or unpaid job last week, or reported to have a job but did not work last week. Weekly income is summarized for those reporting positive incomes (we remove those reporting positive incomes outside of the 1st – 99th percentiles). All incomes are inflation adjusted to be in 1998 values (inflation is calculated as the average percentage change in annual incomes by state). Estimates using SP (2000) come from graphs 5 and 6 in Soukias and Parker (2000).

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 (2000), domestic activities and some 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 to which children carry out domestic work and other unpaid activities. Using the estimates from Parker and Skoufias (2000), we approximate that 12 year-old children not enrolled in school spend around 7.56 hours per-day working in all activities, while 12 year-old children in school spend around 7.15 hours per-day in school (these results are similar when splitting by gender).¹⁸ Thus, we conclude that children not in school

¹⁸Parker and Skoufias (2000) find that children aged 12 spend around 360 minutes a day in school, and around

spend a substantial proportion of their time working, and that school directly competes with work.

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.¹⁹ Thus, combing these earning statistics with the reasonable assumption 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. Finally, in Appendix A.16 we provide further evidence that schooling results in a discrete loss of consumption. We do so by investigating the effect that sending a teenage child to school has on weekly, per-capita household consumption. We find a significant, negative relationship between schooling and consumption (controlling for household characteristics).

5.3 Identification strategy

Our goal is to estimate income and price effects w.r.t. 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.²⁰

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 6 plots the percentage of all children enrolled by age and gender in 1997.²¹ Just over 98% of children are enrolled below the age of 12. This is consistent with findings of previous studies, for example, Attanasio et al (2011) find limited to zero effects of the Progresa grants on enrollment for children under the age of 12 (see Table 2 of their paper), while Todd and Wolpin (2006) state “Because attendance, in the absence of any subsidy, is almost universal through the elementary school ages, subsidizing attendance at the lower grade levels, as under the existing program, is essentially an income transfer”. Thus, it seems reasonable to assume that for children younger than 12 years of age,

170 minutes a day working in all activities. Taking enrollment at age 12 to be 85% , this implies children in school spend $365/0.85 = 7.15$ hours in school. Assuming children in school also spend 2 hours a day working, we estimate that children aged 12 not in school work $(170 - 120 \times 0.85) / 0.15 = 7.56$ hours. See graphs 5 and 6 of SP (2000).

¹⁹It 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.

²⁰De 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 (they find of the 4383 households that received at least one education transfer between March and August of 1999 for children’s school attendance, 464 of them did not receive the relevant (E1) form). 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 allowing them to only identify income effects relative to price effects.

²¹Note, 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.

the conditionalities of the transfer are not binding. We therefore view transfers to children under this age as unconditional transfers to the household.

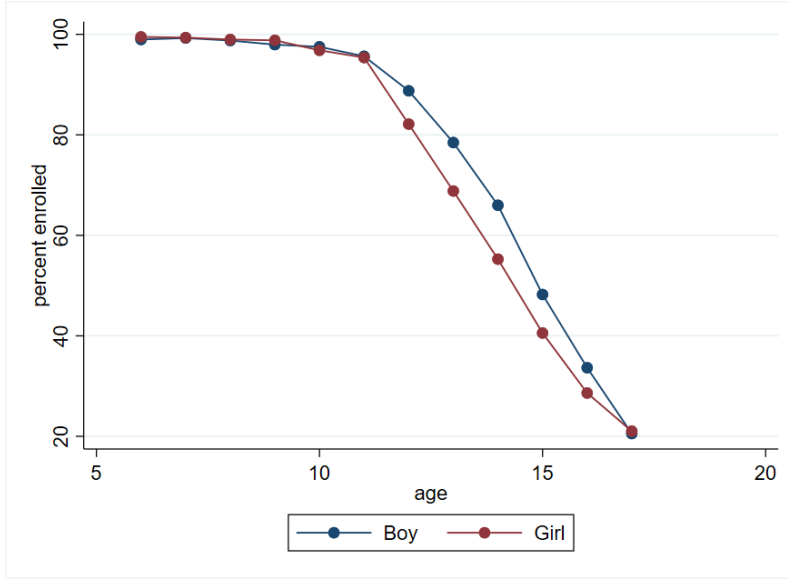


Figure 6: Enrollment by age and gender, 1997

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 1). 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 3 illustrates our identification strategy for constant income and price effects. Restricting to eligible (poor) households, children aged 12 and above are split into four groups: those with no siblings younger than 12 in treatment localities, T_1 ; those with no siblings younger than 12 in control localities, C_1 ; those with siblings younger than 12 in treatment localities, T_2 ; and those with siblings younger than 12 in control localities, C_2 . Further, denote \bar{t}_{c1} and \bar{t}_{c2} the average, per-capita CCT offered to children in groups T_1 and T_2 , respectively, and denote \bar{t}_{u2} the average, per-capita transfers received by siblings under the age of 12 in group T_2 . The price effect is then given by $P = \frac{1}{\bar{t}_{c1}} ((b - a) - (d - c))$, while the income effect is given by $I = \frac{1}{\bar{t}_{u2}} ((f - e) - (h - g) - P\bar{t}_{c2})$. Of course, this is simply an illustration as we have not controlled adequately for the effects of grade, gender, and sibling composition on enrollment.

Table 3: Illustration of Identification Strategy: Constant Income and Price Effects

	Share Enrolled (before grants start)	Share Enrolled (after grants start)
No siblings < 12 years-old		
Treatment: T_1	a	b
Control: C_1	c	d
Has siblings < 12 years-old		
Treatment: T_2	e	f
Control: C_2	g	h

5.4 Constant price and income effects

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

$$Enroll_{it} = a_0 + a_1 treat_i + a_2 after97 + a_3 (treat_i \times after97) + v_{it} \quad (8)$$

where $Enroll_{it}$ takes value 1 if child i in year t is reported to be currently attending school; $treat_i$ takes value 1 if child i lives in a treatment locality; and $after97$ takes value 1 if it is after 1997. Results are presented in Column (2) of Table 4. We find that eligible children aged 12-15 years in treatment localities are 8.8 percentage points more likely to be enrolled in 1998 and 1999 relative to eligible children aged 12-15 years in control localities.²³

We proceed to estimate constant income and price effects w.r.t. annual enrollment of children aged 12-15 years. Specifically we estimate the following regression:

$$Enroll_{it} = b_0 + b_1 t_{c,it} + b_2 t_{u,it} + \beta Z_{it} + year_t + \eta_i + e_{it} \quad (9)$$

²²This drops our sample from around 29500 12-15 year-olds to 21500 12-15 year-olds over the three years.

²³When 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 - see Column (1) of Table 4.

where Z_{it} denotes a vector of child i characteristics in year t including highest grade completed dummies, age dummies, and age interacted with grade. δ_t denotes year fixed effects, and η_i denotes child fixed effects. By including a child fixed effect we control for child gender, area characteristics, and other fixed child and family characteristics. In Appendix A.17 we show that our coefficient estimates on t_c and t_u are robust in size and significance when including additional time-varying sibling controls including the number of siblings in different age and grade brackets in year t . $t_{c,it}$ denotes the per-capita conditional cash transfer 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 :

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

where j denotes the j^{th} sibling of child i , $t_{c,jt}$ denotes the offered conditional transfer to sibling j at time t , and N_{it} denotes child i 's family size in year (measured as both parents plus all children aged 18 and below).²⁴ ²⁵

Under the assumption that $\mathbb{E}[e_{it}|X_{it}] = 0$, price and income effects are given by $P = \frac{\partial \mathbb{E}[Enroll_{it}|X_{it}]}{\partial t_{c,it}} = b_1$ and $I = \frac{\partial \mathbb{E}[Enroll_{it}|X_{it}]}{\partial t_{u,it}} = b_2$. Results of Equation 9 are presented in Column (3) of Table 4. We find significant constant income and price effects. Notably, the presence of significant income effects suggests utility is not linear in consumption. Specifically our constant price effect is equal to 0.0084 and our constant income effect is equal to 0.0036. 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 0.84 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 0.36 percentage points. Given the average t_c offered to children in treatment localities is 6.92 pesos (per-capita, per-week), and the average t_u offered to children is 2.85 pesos (per-capita, per-week), enrollment in treatment localities should be 6.9 percentage points higher relative to control localities after 1997. This estimate is lower than that in Column (2), 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 6.6 members, a 1 USD increase in per-week, per-household conditional transfers leads to a 1.3 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 0.55 percentage point increase in enrollment of children aged 12-15.

²⁴Note, all results are robust to using a constant family size measure, where we measure family size as both parents plus all children under the age of 18 in 1999.

²⁵Note, to adjust transfers for inflation, we use the 1998 semester 1 transfer schedule to determine the CCT offered to each child in both 1998 and 1999.

Table 4: Constant Income and Price Effects

	(1)	(2)	(3)	(4)
$\text{treat} \times \mathbb{1}(\text{year}=1998)$	0.0923*** (0.016)			
$\text{treat} \times \mathbb{1}(\text{year}=1999)$	0.0845*** (0.016)			
$\text{treat} \times \mathbb{1}(\text{year}>1997)$		0.0883*** (0.014)		
tc			0.00843*** (0.001)	0.00847*** (0.001)
tu			0.00360* (0.002)	0.00358* (0.002)
tc (sibs \in [12,15])				-0.000149 (0.001)
F test pvalue	0.608			
$\mathbb{E}[\text{Enroll} \text{control}, \text{year} > 97]$	0.683	0.683	0.683	0.683
$\mathbb{E}[\text{Enroll} \text{treat}, \text{year} > 97]$	0.767	0.767	0.767	0.767
$\mathbb{E}[\text{tc} \text{treat}, \text{year} > 97]$			6.921	6.921
$\mathbb{E}[\text{tu} \text{treat}, \text{year} > 97]$			2.852	2.852
$\mathbb{E}[\text{famsize}]$			6.609	6.609
Observations	21570	21570	21570	21570
R-squared	0.0142	0.0142	0.155	0.155

Note: Robust standard errors are presented in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: Enroll_{it} . Sample for columns (1)-(4): 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. $\text{treat} \times \mathbb{1}(\text{year} > 1997)$ takes value one if child i lives in a treated locality and it is after 1997. tc denotes the weekly, per-capita transfer offered to child i . tu denotes the weekly, per-capita transfer to all siblings under the age of 12. tc (sibs [12-15]) denotes weekly, per-capita transfer offered to all siblings aged 12-15 years. Column (1) includes a treatment indicator and year dummies. Column (2) includes a treatment indicator and an indicator for after 1997. Columns (3) and (4) include child's highest grade completed dummies, age dummies, grade interacted with age, year dummies, and a child fixed effect.

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:

$$\text{Enroll}_{it} = b_0 + b_1 t_{c,it} + b_2 t_{u,it} + b_3 t_{c,it}^{(12+)} + \beta Z_{it} + \delta_t + \eta_i + e_{it} \quad (10)$$

where

$$t_{c,it}^{(12+)} = \frac{1}{N_{it}} \sum_j t_{c,jt} \mathbb{1}(\text{age}_{jt} \in [12, 15])$$

Results of regression 10 are presented in Column (4) of Table 4. 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 2), this will be a negative income shock to the household.²⁶

5.5 Estimating 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:

$$Enroll_{it} = b_0 + \vec{b}_1 t_{c,it} \times X_{it} + \vec{b}_2 t_{u,it} \times X_{it} + \vec{b}_3 Z_{it} + \delta_t + \eta_i + e_{it} \quad (11)$$

where $X_{it} = [1, t_{c,it}, t_{u,it}, y_{it}^f, age_{it}, boy_{it}]$, 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 (almost all mothers in our sample do not report earning an income); age_{it} denotes child i 's age in years; and, boy_{it} denotes an indicator for the gender of child i . Finally, Z_{it} denotes the same matrix of variables as in Equation 10 along with the additional covariates in X_{it} not already captured in Z_{it} , and δ_t, η_i denote year and child fixed effects.

One issue with estimating Equation 11 is that father's income is likely measured with error. Highlighting this, the correlation between reported incomes across 1997-1998 and 1998-1999 is only 0.42. Thus, estimating Equation 11 via OLS will likely result in attenuation bias. Moreover, there is a potential endogeneity issue with including father's income as parents may adjust their labor supply in response to schooling decisions (although Parker and Skoufias (2000) find no evidence that parents adjust their labor supply in response to receiving grants, suggesting endogeneity may not be an issue). Therefore, we instrument for father i 's income with the mean hourly wage in father i 's locality, 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 regressions are presented in Appendix A.19.

²⁶See 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.

Regression results of Equation 11 for both our OLS and IV estimates are presented in Table 14 in Appendix A.20. 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 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 both our OLS and IV specifications), likely due to the much smaller variation in unconditional cash transfers relative to conditional cash transfers. Finally, our IV estimate for the coefficient on father's income is an order of magnitude higher than that found in the OLS specification (0.0035 vs. 0.0002), again, consistent with attenuation bias in the OLS specification.

Moments of the income and price effect schedules are presented in Table 5 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. 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:

$$\mathbb{E} \left[\frac{I(X)}{P(X)} \right] \approx \mathbb{E}[I(X)] \left(\frac{1}{\mathbb{E}[P(X)]} + \frac{1}{\mathbb{E}[P(X)]^3} \text{Var}(P(X)) \right) - \frac{1}{\mathbb{E}[P(X)]^2} \text{Cov}(I(X), P(X)) \quad (12)$$

We find the average of this ratio is equal to 0.43 under our OLS specification and 0.25 under our IV specification, implying our curvature estimate will be lower under our IV specification. This lower ratio is primarily due to a higher average price effect under the IV specification. Note, when we estimate second-order moments of the income and price effect schedules to plug into Equation 12, we adjust these moments for the bias that results from using point estimates.²⁷ Finally, we estimate that enrollment in treatment localities is 9.2 percentage points higher in 1998-1999 relative to control localities under our OLS specification, and 8.7 percentage points higher under our IV specification. This is in line with the observed increase in enrollment of 8.8 percentage points (see Column 2 of Table 4).

²⁷For example, denote $P(X) = \beta X$, giving $\text{Var}(P(X)) = \beta^2 \text{Var}(X)$. However, we estimate $\hat{P}(X) = \hat{\beta} X$. Since $\mathbb{E}[\hat{\beta}^2] \neq \beta^2$, we proceed with the following adjustment to obtain an unbiased estimate of $\text{Var}(P(X))$: $\text{Var}(P(X)) = (\mathbb{E}[\hat{\beta}^2] - \text{Var}(\hat{\beta})) \text{Var}(X)$, where we bootstrap to obtain moments of $\hat{\beta}$'s. Note, for robustness, we also calculate (1) $\mathbb{E}_X[\hat{I}(X)/\hat{P}(X)]$ and (2) the second order Taylor expansion of Equation 12 using biased but consistent estimates of second order moments of the income and price effects (e.g., we use $\text{Var}(P(X)) = \hat{\beta}^2 \text{Var}(X)$). Estimates of $\mathbb{E}[I(X)/P(X)]$ using these latter two methods are 0.28 and 0.22 using our IV specification, respectively.

Table 5: Moments of the Income and Price Effect Schedules

	OLS	IV
$\mathbb{E}[P(X)]$	0.0097*** (0.0025)	0.013*** (0.0034)
$\mathbb{E}[I(X)]$	0.0039 (0.0034)	0.0035 (0.0030)
$\mathbb{E}\left[\frac{I(X)}{P(X)}\right]$	0.43 (0.45)	0.25 (0.27)

Note: Moments are taken for the treated, eligible sample of children aged 12-15 years in 1998 and 1999. Bootstrapped standard errors from 100 bootstrap draws for $\mathbb{E}[P(X)]$, $\mathbb{E}[I(X)]$, and $\mathbb{E}[I(X)]/\mathbb{E}[P(X)]$ are presented in parentheses.

5.6 Estimating the coefficient of constant relative risk aversion

Using our estimates of $\mathbb{E}\left[\frac{I}{P}\right]$ from both our constant and heterogeneous income and price effect regressions, we determine the coefficient of relative risk aversion, γ , via the following relationship:

$$\mathbb{E}\left[\frac{I}{P}\right] = 1 - \mathbb{E}\left[\frac{(y + t_u + y_c)^{-\gamma}}{(y + t_u + t_c)^{-\gamma}}\right] \quad (13)$$

Before solving for γ , we need to address two issues. First, we need a measure of 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 for the home, 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 2). 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. It is worth noting that such an assumption will likely work against the targeting benefit as it essentially increases the variance of parental income from $\text{Var}(y^f)$ to $1.8^2\text{Var}(y^f)$ where y^f denotes father's income.

Second, we need a measure of potential child income. 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, y_c is simply a function of child age,

child gender, and area characteristics (see Appendix A.18 for details on how we estimate y_c).

Our results are presented in Table 6 below. Our curvature estimate under our IV specification is 1.37, in line with (although slightly higher than) the curvature estimates found in Chetty (2006). Our curvature estimates for both the heterogeneous OLS specification and the constant elasticity specification are both 2.63. These specifications give the same value of γ as they both give the same ratio of $\mathbb{E}[I/P]$ (this is perhaps not surprisingly given we reject the null hypothesis of heterogeneity in the heterogeneous OLS specification). Finally, it is worth mentioning that it is inconsistent with theory to have $\gamma > 0$ and have both constant income *and* price effects. For this reason, and for reasons of endogeneity and measurement error in father’s income, we will proceed using our IV estimates. Finally, we vary mothers’ earnings from 0% to 100% as much as fathers’ earnings, and find that γ ranges from $[0.91, 1.49]$ using our IV specification.

Table 6: Curvature of Utility

	Constant	Heterogeneous: OLS	Heterogeneous: IV
$\mathbb{E}[\frac{I}{P}]$	0.43	0.43	0.25
γ	2.63	2.63	1.37

Note: γ 's are estimated using Equation 13 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.

6 The Size of the Targeting Benefit for Progresa

6.1 Size of the targeting benefit under the observed Progresa grants

We now estimate the size of the targeting benefit relative to the exclusion cost under the average, observed Progresa transfer schedule of $\bar{t}_c = 6.9$ and $\bar{t}_u = 2.9$ pesos, per-capita, per-week. To do so, we evaluate the following discretized version of Equation 3 for our sample of treated, eligible children in 1998 and 1999:

$$\frac{TB}{EC} = \frac{\left(1 - \frac{\bar{S} + t_c \bar{P}}{1 + t_c \bar{I}}\right) \sum_i^N (y_i + \bar{t}_u + \bar{t}_c)^{-\gamma} S(X_i)}{\left(\frac{\bar{S} + \bar{t}_c \bar{P}}{1 + t_c \bar{I}}\right) \sum_i^N (y_i + \bar{t}_u + y_{ci})^{-\gamma} (1 - S(X_i))}$$

where N denotes our sample size, and $S(X_i)$ denotes predicted enrollment for children with characteristics X_i under the average Progresa grants, where X_i includes parental income, child age, and child gender (we predict $S(X_i)$ using our IV specification of Equation 11). The value of γ is taken from our IV specification in Table 6 above: $\gamma = 1.37$; the average income and price effects are taken from our IV specification in Table 5 above: $\bar{I} = 0.0035$, $\bar{P} = 0.013$; and, observed average enrollment under the Progresa grants is equal to $\bar{S} = 0.77$. Our estimate for TB/EC is presented in Column (1) of Table 7 below. We find a ratio equal to 0.89, 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 TB/EC varies between [0.87, 0.90].

Table 7: Size of the Targeting Benefit

	(1) Observed	(2) Optimal
t_c	6.9	5.9
t_u	2.9	3.7
Share of budget to CCT	64%	55%
TB/EC	0.89	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.

6.2 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: $R = \bar{S}\bar{t}_c + \bar{t}_u = 0.77 \times 6.9 + 2.9 \approx 8$, where R represents the per-child budget (in pesos, per-capita, per-week).²⁸ Then, using data on our treated, eligible sample of children aged 12-15 in 1998 and 1999, we determine the transfer schedule that satisfies $TB/EC = 1$. Specifically we solve the following for t_c and t_u :

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

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

where $S(X_i, t_c, t_u)$ denotes the share of children with characteristics X_i enrolled under the transfer schedule t_c, t_u , and $\bar{S}(t_c, t_u)$ is the average (total) share of children enrolled under the transfer schedule t_c, t_u . We assume that for all households, enrollment varies with t_c and t_u at rate \bar{P} and \bar{I} , respectively, i.e., $S(X_i, t_c, t_u) = S(X_i, \bar{t}_c, \bar{t}_u) + (t_c - \bar{t}_c)\bar{P} + (t_u - \bar{t}_u)\bar{I}$.²⁹

²⁸For 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.

²⁹We assume this because we are unable to precisely estimate heterogeneity in our income and price effect

Results for the optimal transfers are presented in Column (2) of Table 7 above. Our findings suggest that 55% of the Progresa budget should be allocated towards a CCT (compared to the observed proportion of 64%), with the optimal transfers equal to $t_c^* = 5.9, t_u^* = 3.7$ 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). Adding additional benefits of conditioning would simply increase t_c^* (similarly, adding 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 [52%, 55%].

6.3 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 quickly in consumption. Second, child incomes are high relative to parent incomes. As shown in Table 2, 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 7 plots the density of per-capita parental income for those sending their 14 year-old child to school and those not sending their 14 year-old child to school in our sample (note, figures look very similar when we look at households with a 12, 13, or 15 year-old child).³⁰ 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 8 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, the opportunity cost of raising t_c by a dollar is very high in terms of reducing the unconditional cash transfer. Reducing the unconditional cash

schedules. However, we recalculate the optimal transfers using our imprecise price and income effect schedules and find very similar results. Specifically, we update the shares enrolled as follows: $S(X_i, t_c, t_u) = S(X_i, \bar{t}_c, \bar{t}_u) + \int_{\bar{t}_c}^{t_c} P(X_i, v, \bar{t}_u) dv + \int_{\bar{t}_u}^{t_u} I(X_i, t_c, v) dv$. We find that 47% of the Progresa budget should be allocated towards a CCT.

³⁰The density of parental incomes for those sending to school is given by $f(y|enroll = 1) = f(y)S(y)/\bar{S}$, where $S(y)$ is obtained from the relationship between enrollment and parental income in our IV specification of Equation 11.

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

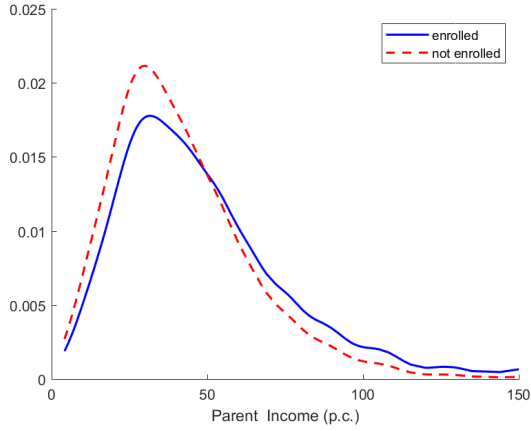


Figure 7: Distribution of Parent Income

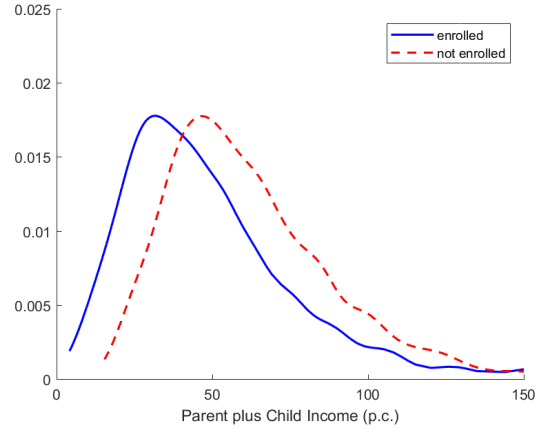


Figure 8: Distribution of Parent + Child Income

6.4 Counterfactuals

6.4.1 Varying the curvature of utility

To conclude this section, we 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 observed average income effect, $\mathbb{E}[I]$, which then scales $\mathbb{E}[I/P]$ (see Equation 12), which then scales our estimate of curvature, γ (see Equation 13). We then recalculate the optimal share of the budget allocated towards a CCT under our new values for γ . Figure 9 illustrates how the optimal share of the budget to a CCT varies with γ . As to be expected by Proposition 3.4, 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.

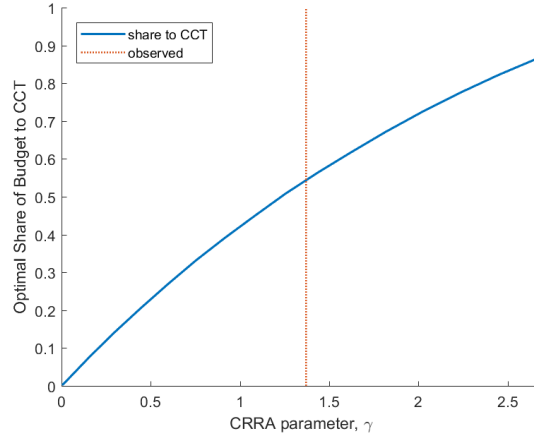


Figure 9: Varying Curvature of Utility

6.4.2 Varying parental income inequality

Next we vary the mean difference in 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 $\frac{\partial S(X)}{\partial y}$ keeping \bar{S} constant. Increasing this derivative will increase $\mathbb{E}[y|s = 1] - \mathbb{E}[y|s = 0]$ as $\mathbb{E}[y|s = 1] = \frac{1}{N} \sum_{i=1}^N \frac{S(X_i)y_i}{\bar{S}}$ and $\mathbb{E}[y|s = 0] = \frac{1}{N} \sum_{i=1}^N \frac{(1-S(X_i))y_i}{1-\bar{S}}$. As expected by Proposition 3.5, the optimal CCT is decreasing in this difference (see Figure 10 below).

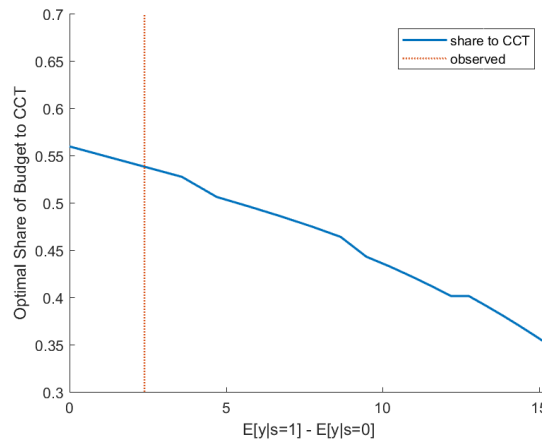


Figure 10: Varying Parental Income Inequality

6.4.3 Varying child income

Finally, we vary our estimate of predicted child income, y_c , by scaling our predicted measure by $\kappa \in [0.25, 1]$. 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 11 below. As we decrease child income, we decrease the discrete loss in household income for those households sending to school. Hence, we decrease the size of the targeting benefit relative to the exclusion cost, thus, decreasing the optimal CCT.

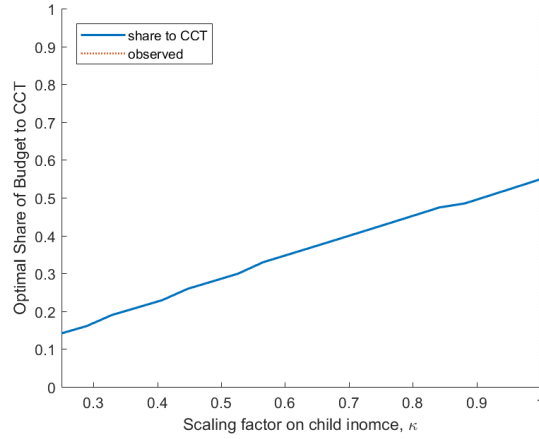


Figure 11: Varying Child Income

7 Conclusion

This paper argues that cash transfers made conditional on school attendance can better target low consumption households relative to an unconditional cash transfer. 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 w.r.t. 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 these children are enrolled. Using these elasticities we estimate that 55% 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 1.37). Thus, by allocating some of the budget towards a CCT, the planner can better target transfers towards households who place a higher value on receiving an extra dollar today.

Moving forward, we believe our results have several implications for the design of cash transfer programs. First, it's 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, it is important to know the extent to which private enrollment levels are below the social levels. For example, if there are large externalities associated with increased school enrollment and price effects are large relative to income effects, this alone may be enough to justify imposing conditions.

Combining all these factors together, Tables 8 and 9 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 8). 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 there are externalities associated with increased schooling). If there is no under-enrollment, one should offer a UCT so as to not distort parental enrollment decisions, whereas, if there is substantial under-enrollment, one should offer a CCT (see Table 9).

Table 8: Large Income Effects: $I \approx P$

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

Table 9: Small Income Effects: $I \ll P$

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

Looking forward, we believe much of our analysis 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): 1206-1240. <http://dx.doi.org/10.1257/aer.102.4.1206>
- AKERLOF, G. (1970): “The Market for Lemons: Quality Uncertainty and the Market Mechanism,” *Quarterly Journal of Economics* vol. 84 (3): 488-500.
- ANGELUCCI, M. AND G. DE GIORGI (2009): “Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles’ Consumption,” *American Economic Review* vol. 99, 486-508.
- AKRESH, R., D. DE WALQUE, AND H. KAZIANGA (2013): “Cash Transfers and Child Schooling: Evidence from a Randomized Evaluation of the Role of Conditionality,” *World Bank Policy Research Working Paper* 6340.
- ATTANASIO O., 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 Brazil's Bolsa Escola Program” *World Bank Economic Review* vol. 17, 229-254.
- COADY, D., M. GROSH, AND J. HODDINOTT (2004): “Targeting of Transfers in Developing Countries: Review of Lessons and Experience,” *The International Bank for Reconstruction and Development, The World Bank*.
- COHEN, A., AND L. EINAV (2007): “Estimating Risk Preferences from Deductible Choice,” *American Economic Review*. vol. 97(3), 745-788.
- CHETTY, R. (2006): “A New Method of Estimating Risk Aversion,” *American Economic Review* vol. 96(5), 1821-1834.
- DE BRAUW, A. AND J. HODDINOTT (2011): “Must conditional cash transfer programs be conditioned to be effective? The impact of conditioning transfers on school enrollment in Mexico,” *Journal of Development Economics* vol. 96, 356-370.
- DE JANVRY, A. F. FINAN, E. SADOULET, AND R. VAKIS (2006): “Serve as Safety Nets in Keeping Children at School and from Working When Exposed to Shocks?,” *Journal*

- of Development Economics* vol. 79, 349-373.
- EDMONDS, E. (2006): “Child Labor and Schooling Responses to Anticipated Income in South Africa,” *Journal of Development Economics* vol. 81, 386-414.
- EDMONDS, E., AND SCHADY, N. (2012): “Poverty Alleviation and Child Labor,” *American Economic Journal: Economic Policy* vol. 4(4), 100-124.
- FERREIRA, F., D. FILMER, AND N. SCHADY (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>
- KAPLOW (2005): “The Value of a Statistical Life and the Coefficient of Relative Risk Aversion,” *Journal of Risk and Uncertainty* vol. 31(1), 23-34.
- KOCHERLAKOTA, N. (2005): “Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation,” *Econometrica* vol. 73(5), 1587-1621.
- MEHRA, R. AND PRESCOTT E. (1985): “The Equity Premium: A Puzzle,” *Journal of Monetary Economics* vol. 15(2), 145-161.

- LAYARD, R., S. NICKELL, AND G. MAYRAZ (2008): “The Marginal Utility of Income,” *Journal of Public Economics* vol. 92, 1846-1857.
- PARKER, S. AND E. SKOUFIAS (2000): “The Impact of PROGRESA on Work, Leisure and Time Allocation,” *International Food Policy Research Institute*
- PARKER, S. AND P. TODD (2017): “Conditional Cash Transfers: The Case of Progres/Oportunidades,” *Journal of Economic Literature* vol. 55(3), 866-915.
- RAVALLION, M. (2009): “How Relevant is Targeting to the Success of an Antipoverty Program? A Case Study for China,” *World Bank Research Observer* vol. 24(2): 205-231.
- RODRÍGUEZ-CASTELÁN, C. (2017): “Conditionality as Targeting? Participation and Distributional Effects of Conditional Cash Transfers,” *Poverty and Equity Global Practice Group*
- SCHADY, N., AND M. ARAUJO (2008): “Cash Transfers, Conditions, and School Enrollment in Ecuador,” *Economica* vol. 8, 4370.
- SCHULTZ, T. (2004): “School subsidies for the poor: evaluating the Mexican Progres poverty program,” *Journal of Development Economics* vol. 74(1), 199-250.
- SKOUFIAS, E., B. DAVIS, AND J. BEHRMAN (1999): “An Evaluation of the Selection of Beneficiary Households in the Education, Health, and Nutrition Program (PROGRESA) of Mexico,” *International Food Policy Research Institute*
- THORISSON, H. (1995): “Coupling Methods in Probability Theory,” *Scandinavian Journal of Statistics* vol. 22(2), 159-182.
- TODD, P. AND K. WOLPIN (2006): “Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility,” *American Economic Review* vol. 96(5), 1384-1417.

A Appendix

A.1 Proof of Proposition 3.2

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

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

This covariance term will be non-negative if $\frac{\partial(1-F(\tilde{\mu}(y)|y))}{\partial y} = -f(\tilde{\mu}(y)|y) \frac{\partial \tilde{\mu}(y)}{\partial y} + \frac{\partial(1-F(\tilde{\mu}(y)|y))}{\partial y} \geq 0$.³¹ 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 $\bar{P} = \int_Y -\frac{\partial \tilde{\mu}(y)}{\partial t_c} f(\tilde{\mu}(y)|y) f(y) dy = \int_Y \frac{u_c(y + y_c + t_u)}{v_1(\tilde{\mu}, 1) - v_1(\tilde{\mu}, 0)} f(\tilde{\mu}(y)|y) f(y) dy > 0$, $\bar{I} = \int_Y \frac{-\partial \tilde{\mu}(y)}{\partial t_u} f(\tilde{\mu}(y)|y) f(y) dy = \int_Y \frac{u_c(y + t_c + t_u) - u_c(y + y_c + t_u)}{v_1(\tilde{\mu}, 1) - v_1(\tilde{\mu}, 0)} f(\tilde{\mu}(y)|y) f(y) dy = 0$ 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 $t_c^* < y_c$.

³¹See Thorisson (1995) for a proof that the covariance of two increasing functions of a random variable is positive.

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:

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

We can write the Planner's Lagrangian as follows

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

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:

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

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:

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

We can write the Lagrangian as follows

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

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

$$\epsilon d\tau (u_c(y + t_u(y) + t_c(y))S(y) + u_c(y + y_c + t_u(y))(1 - S(y)) - \lambda(1 + t_c(y)I(y)) + \delta_2(y)) = 0 \quad (14)$$

$$\epsilon d\tau (u_c(y + t_u(y) + t_c(y))S(y) - \lambda(S(y) + t_c(y)P(y)) + \delta_1(y)) = 0 \quad (15)$$

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 14 we get

$$u_c(y + t_u(y) + t_c(y))S(y) + u_c(y + y_c + t_u(y))(1 - S(y)) = \lambda \quad (16)$$

And from Equation 15 we get

$$u_c(y + t_u(y) + t_c(y))S(y) = \lambda S(y) - \delta_1(y) < \lambda S(y) \quad (17)$$

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 16 we get $u_c(y + t_u(y) + t_c(y)) > \lambda$, a contradiction with Equation 17. 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. We assume the parent's problem remains the same, however, we now assume a more specific functional form on utility in the second generation: $v(\mu, s) = \beta u(y_2(\mu, s))$ where β denotes the common discount rate across generations, and $y_2(\mu, s)$ denotes the way in which parents believe schooling and ability translate into earnings when the child is an adult. We assume $\frac{\partial y_2}{\partial \mu} > 0$, $\frac{\partial y_2}{\partial s} > 0$, $\frac{\partial^2 y_2}{\partial \mu \partial s} > 0$. Further, we assume parents mis-perceive/under-value the return to schooling so that the correct function can be expressed as follows $y_2(\mu, \theta s)$ where $\theta > 1$. The planner's problem is as follows:

$$\begin{aligned} W = \max_{t_u, t_c} & \int_Y \int_{\tilde{\mu}} u(y + t_u + t_c) + \beta u(y_2(\mu, \theta)) f(\mu, y) d\mu dy + \\ & \int_Y \int_{\tilde{\mu}} u(y + y_c + t_u) + \beta u(y_2(\mu, 0)) f(\mu, y) d\mu dy \\ \text{s.t. } & t_u + t_c \int_Y \int_{\tilde{\mu}} f(\mu, y) d\mu dy \leq R \end{aligned}$$

where $\tilde{\mu}$ denotes the indifferent household defined implicitly as follows:

$$u(y + t_u + t_c) + \beta u(y_2(\tilde{\mu}, 1)) = u(y + y_c + t_u) + \beta u(y_2(\tilde{\mu}, 0))$$

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

$$\begin{aligned} & \underbrace{\left(1 + \frac{\partial t_u}{\partial t_c}\right) \int_Y \int_{\tilde{\mu}}^{\infty} u_c(y + t_u + t_c) f(\mu, y) d\mu dy}_{\text{Targeting Benefit (TB)} > 0} + \underbrace{\frac{\partial t_u}{\partial t_c} \int_Y \int^{\tilde{\mu}} u_c(y + y_c + t_u) f(\mu, y) d\mu dy}_{\text{Exclusion Cost (EC)} < 0} \\ & + \underbrace{\int_Y \left(\frac{\partial \tilde{\mu}}{\partial t_u} \frac{\partial t_u}{\partial t_c} + \frac{\partial \tilde{\mu}}{\partial t_c} \right) \beta (u(y_2(\tilde{\mu}, 1)) - u(y_2(\tilde{\mu}, \theta))) f(\tilde{\mu}|y) f(y) dy}_{\text{Enrollment Benefit (EB)} > 0} = 0 \end{aligned} \quad (18)$$

where we have a new term relative to Equation 1 which we denote the ‘‘Enrollment Benefit’’. This term captures the fact the raising t_c leads to an increase in the number of parents sending their child to school. This term is positive as $u(y_2(\mu, 1)) - u(y_2(\mu, \theta)) < 0$ and $\frac{\partial \tilde{\mu}}{\partial t_u} \frac{\partial t_u}{\partial t_c} + \frac{\partial \tilde{\mu}}{\partial t_c} \leq 0$ shown below. Thus, the enrollment benefit works in favor of increasing t_c . We now derive the following proposition:

Proposition A.1. *As long as $\frac{\partial t_u}{\partial t_c} \geq -1$ (i.e., the targeting benefit is positive), the optimal CCT is increasing in the extent to which parents under-value the return to schooling, θ .*

Proof. By Topkis' Theorem, it's sufficient to show that $\frac{\partial^2 W}{\partial t_c^* \partial \theta} > 0$. A sufficient condition for this to hold is $\frac{\partial \tilde{\mu}}{\partial t_u} \frac{\partial t_u}{\partial t_c} + \frac{\partial \tilde{\mu}}{\partial t_c} < 0 \forall y$ (remembering $\tilde{\mu}$ is a function of y). Suppose not, i.e., $\frac{\partial \tilde{\mu}}{\partial t_u} \frac{\partial t_u}{\partial t_c} + \frac{\partial \tilde{\mu}}{\partial t_c} \geq 0$ for some y . We know $\frac{\partial \tilde{\mu}}{\partial t_c} < \frac{\partial \tilde{\mu}}{\partial t_u} \leq 0 \forall y$. Thus, if $\frac{\partial \tilde{\mu}}{\partial t_u} \frac{\partial t_u}{\partial t_c} + \frac{\partial \tilde{\mu}}{\partial t_c} \geq 0$ for some y , $\implies \frac{\partial t_u}{\partial t_c} < -1$, a contradiction. \square

A.7 Proof of Corollary 3.3.1

Proof. As in Proposition 3.3, 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, $f(y) \sim \text{Unif}[6 - \epsilon, 6 + \epsilon]$. It's still true that parents with income above 6 send their children to school when $t_u = R$, $t_c = 0$. Consider the FOC of the Planner's problem with respect to t_c (when we substitute the budget constraint to write $t_u(t_c)$) when $t_u = R$, $t_c = 0$, noting that $\frac{\partial t_u}{\partial t_c} = -1/2$. Using the shorthand $G(y) \equiv G(u(c^*(y)) + v(10, s^*(y)))$, we can write the planner's FOC as:

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

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

$$\frac{G'(6-\delta)}{G'(6+\delta)} < \frac{(6-\delta+10)^{1/2}}{(6+\delta+3)^{1/2}} < (16/9)^{1/2} \quad (19)$$

we will have:

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

Inequality 19 will be true for sufficiently small δ by continuity of G' (as $\frac{G'(6-\delta)}{G'(6+\delta)} \approx 1$, while $\frac{(6-\delta+10)^{1/2}}{(6+\delta+3)^{1/2}} \approx (16/9)^{1/2}$). This means that for sufficiently small ϵ , the planner's FOC will be positive at $t_c = 0$:

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

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

A.8 Proof of Proposition 3.4

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

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

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

where

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

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

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

Thus our FOC simplifies to

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

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

A.9 Proof of Proposition 3.5

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 1, we can rewrite the FOC as:

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

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

$$\underbrace{\left(1 + \frac{\partial t_u}{\partial t_c}\right) \bar{S} \int_Y u_c(y + t_u + t_c) dF(y|s^* = 1)}_{\text{targeting benefit}} + \underbrace{\frac{\partial t_u}{\partial t_c} (1 - \bar{S}) \int_Y u_c(y + y_c + t_u) dF(y|s^* = 0)}_{\text{exclusion cost}} = 0$$

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 targeting benefit(t_c^*, f_2) \geq targeting benefit(t_c^*, f_1). Moreover, because $F_2(y|s^* = 0)$ (weakly) $FOSD$ $F_1(y|s^* = 0)$, $\frac{\partial t_u}{\partial t_c}(t_c^*, f_2) < 0$, and u_c is decreasing we also get that exclusion cost(t_c^*, f_2) \geq exclusion cost(t_c^*, f_1), with one of these two inequalities being strict. Hence, $FOC(t_c^*, f_2) > 0$, so that increasing t_c increases welfare; hence the optimal local reform is to increase t_c .

□

A.10 Borrowing across generations

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

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

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.³² This assumption implies schooling is weakly increasing in ability μ . We can now derive the following proposition.

Proposition A.2. *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.³³ 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

³²Increasing 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.

³³ $\frac{\partial W}{\partial \mu} = \beta v'(c_2(s)) \frac{\partial y_2(\mu, s)}{\partial \mu}$ 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.

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

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:

$$-\frac{1}{\int_Y \int_{\tilde{\mu}} f(\mu, y) d\mu dy} \int_Y \int_{\tilde{\mu}(y)} u_c(c^*(1; y, \mu)) f(\mu, y) d\mu dy + \frac{1}{\int_Y \int_{\tilde{\mu}} f(\mu, y) d\mu dy} \int_Y \int_0^{\tilde{\mu}(y)} u_c(c^*(0; y, \mu)) f(\mu, y) d\mu dy$$

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 $c^*(1; y, \mu) > c^*(0; y, \mu) \forall y, \mu$.³⁴ Thus, marginal utility is higher for those who do not enroll. Thus, $t_c = 0$ is optimal. \square

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

³⁴Household'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 $c^*(1) \geq c^*(0)$.

A.11 Borrowing within a generation

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

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

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

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

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

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

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

This is functionally equivalent to the original problem without any borrowing. As such, the Proof of Proposition 3.3 still holds in this world, so that it can be optimal to have a CCT.

A.12 Transfers in multiple generations

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

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

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:

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

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

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

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

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

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

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

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

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

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

Notably, the targeting benefit and exclusion cost are identical to Equation 1; 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 A.1 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.13 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 3, but now we consider a modified household problem with labor supply decisions:

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

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

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

Lemma A.3. *Schooling, $s^*(n, \mu)$, is weakly increasing in parent productivity n , so that there exists a cutoff \tilde{n} such that those parents with $n < \tilde{n}$ do not send to school and those with $n \geq \tilde{n}$ send to school.*

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

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

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

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

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

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

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

Lemma A.4. *There still exists a discontinuity in household income $z(n) \equiv y(n) + (1 - s(n))y_c$, where $y(n) = nl(n)$. In particular, around \tilde{n} , we have that:*

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

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

Lemmas A.3 and A.4 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 2 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 A.5. *If utility is concave in consumption and parents can freely adjust their labor supply, a pure UCT is not necessarily optimal.*

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

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

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

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

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

A.14 Annual Schooling Decisions Model

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

A.14.1 Parent Problem

We begin with the parent problem:

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

where $V(t+1, \mu) > V(t, \mu) \forall t \in 0, \dots, T-1$, $V_2(t, \mu) > 0$, and $V_2(t+1, \mu) > V_2(t, \mu)$, i.e., child ability and total schooling are complements. To make the planner's problem tractable, we have imposed that both the conditional and unconditional transfers are constant across the 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 1. *If $s_t^* = 0$ then $s_{t+n}^* = 0 \forall n \in 1, \dots, T-t$. In words, if you stop sending to school for a year, you stop sending to school for all remaining years.*

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

Total utility from $s_m^* = 0$, $s_{m+1}^* = 1$ is given by:

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

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

Total utility from $s_m^* = 1$, $s_{m+1}^* = 0$ is given by:

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

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

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

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

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

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

Because we assume ability and total years of schooling are complements, we know that all households with $\mu > \tilde{\mu}_m$ will strictly prefer stopping in year $m+1$ to stopping in year m , and vice versa. Moreover, because we assume V has decreasing differences in total years of schooling, we know that if $\mu \geq \tilde{\mu}_m \implies \mu > \tilde{\mu}_i$ for $i \in 1, \dots, m-1$ and $\mu \leq \tilde{\mu}_m \implies \mu < \tilde{\mu}_i$ for $i \in m+1, \dots, T$.³⁵ Thus, the set of households sending to school in year m will have $\mu \geq \tilde{\mu}_m$. We now describe the planner's problem.

A.14.2 Planner's Problem

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

Before we write the planner's problem, it is first useful to re-write the parents' optimal

³⁵If $\mu > \tilde{\mu}_{m+1} \implies u(y + t_u + t_c) + V(m+1, \mu) > u(y + t_u + y_{cm+1}) + V(m, \mu)$. 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. Hence $u(y + t_u + y_{cm}) - u(y + t_u + t_c) < V(m, \mu) - V(m-1, \mu) \implies \mu > \tilde{\mu}_m$. But if $\mu > \tilde{\mu}_m$, by the same logic, $\mu > \tilde{\mu}_{m-1}$ etc. Hence, if $\mu \geq \tilde{\mu}_{m+1} \implies \mu > \tilde{\mu}_i$ for $i \in 1, \dots, m$. One can easily repeat a similar exercise to show $\mu \leq \tilde{\mu}_{m+1} \implies \mu < \tilde{\mu}_i$ for $i \in m+2, \dots, T$.

stopping problem explicitly in terms of calendar years and child cohort, i.e., parents of a child who starts school in year b get to pick a year $m \in \{b, \dots, T+b\}$ in which they stop sending their child to school:

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

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

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

By the same logic as before, households with $\mu \geq \tilde{\mu}(y, m-b)$ will send to school in year m and vice versa. Again, for ease of notation we have omitted that $\tilde{\mu}$ is also a function of the transfer schedule (t_u, t_c) . We write the planner's problem as follows³⁶

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

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

$$\begin{aligned} \sum_{b=b_0}^{b_0+B} \sum_{t=b}^{T+b} & \left(\left(1 + \frac{\partial t_u}{\partial t_c} \right) \int_Y u_c(y + t_u + t_c) S(y, t, b) dF(y, t, b) + \right. \\ & \left. \frac{\partial t_u}{\partial t_c} \int_Y u_c(y + t_u + y_c(t-b)) (1 - S(y, t, b)) dF(y, t, b) \right) = 0 \end{aligned} \quad (20)$$

³⁶Note, 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.

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 20 is simply the average targeting benefit across all cohorts and years plus the average exclusion cost across all cohorts and years.

A.15 Sufficient Statistics for Annual Schooling Decisions Model

As with our baseline model, we can show in our annual schooling decisions model 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 20). 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:

$$\frac{1}{TB} \sum_{b=b_0}^{b_0+B} \sum_{t=b}^{T+b} \left(\left(1 + \frac{\partial t_u}{\partial t_c} \right) \int_Y u_c(y + t_u + t_c) S(y, t - b) dF(y) + \frac{\partial t_u}{\partial t_c} \int_Y u_c(y + t_u + y_c(t - b)) (1 - S(y, t - b)) dF(y) \right) = 0 \quad (21)$$

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 $S(y, t - b) = \int_{\bar{\mu}(y, t - b)} dF(\mu|y)$ 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 income, child age, and child gender, e.g., $I(y, t - b, boy)$ etc.

A.16 The Effect of Schooling on Consumption

In this subsection we investigate the effect schooling has on household consumption.³⁷ To do so, we construct measures of household consumption using the November 1998 survey, as this survey has detailed information on food consumption and purchases, allowing us to determine area level food prices.³⁸ It is critical for us to be able to value food given it makes up a substantial proportion of household consumption. Of course, by November 1998, the Progresa grants had started in treatment localities, so we restrict our analysis to households living in control localities. We run the two following regression specifications for all households in treatment localities with at least one 12-15 year-old child:

$$y_{il} = \beta_0 + \beta_1 poverty_score_i + \beta_2 share_insch_i + \theta X_{il} + \epsilon_{il} \quad (22)$$

$$y_{il} = \beta_0 + \beta_1 poverty_score_i + \beta_2 \mathbb{1}(1 \text{ child in sch})_i + \beta_3 \mathbb{1}(2 \text{ or more children in sch})_i + \theta X_{il} + \epsilon_{il} \quad (23)$$

where y_{il} includes three outcome variables of interest for household i living in locality l : 1) weekly household consumption excluding school related consumption (in pesos); 2) weekly household food consumption (in pesos); and 3) household spending in the previous year on seeds, fertilizers, pesticides, machinery, and labor (in pesos). $poverty_score_i$ denotes the poverty index for household i where a lower value implies a greater level of poverty. This index, created by Progresa officials, was used to predict whether per-capita household income, less income of any children, is above 80 pesos, per-capita, per-week. We therefore interpret this variable as a proxy for parental income. We use this variable over parental income given we suspect parental income is measured with a high degree of error. $share_insch_i$ measures the share of children aged 12-15

³⁷We focus only on consumption and not savings given we suspect savings are minimal in these poor communities. Supporting this, when households were asked in the March 1998 survey what they would do if they had extra spending money, less than 5% rank saving the additional money as a first priority.

³⁸We follow Angelucci and De Giorgi (2009) on constructing area level food prices.

in household i enrolled in school, $\mathbb{1}(1 \text{ child in sch})$ takes value 1 if a household has 1 child aged 12-15 enrolled, $\mathbb{1}(2 \text{ or more children in sch})$ takes value 1 if a household has 2 or more children aged 12-15 enrolled in school, and X_{it} denotes a vector of household and area characteristics including: mom's age, dad's age, mom's years of education, dad's years of education, dummy variables for family size, dummy variables for number of children under the age of 7, dummy variables for number of children aged 12-15, marginality index of the locality, and state fixed effects. Results from regressions 22 and 23 are shown in Table 10 below.

Table 10: Household Consumption by Schooling, 1998

	(1)	(2)	(3)	(4)	(5)	(6)
poverty index	0.109*** (0.027)	0.0391* (0.020)	0.248 (0.315)	0.109*** (0.027)	0.0389* (0.020)	0.250 (0.315)
share in school	-17.60** (7.408)	-13.92** (5.601)	-114.8 (84.293)			
1 in school				-22.10*** (7.623)	-17.34*** (5.665)	-92.12 (93.554)
2 or more in school				-22.99** (10.227)	-20.57** (8.293)	-258.0** (121.591)
Mean Dependent	412.8	355.0	771.3	412.8	355.0	771.3
Observations	1303	1328	632	1303	1328	632
R-squared	0.194	0.126	0.147	0.198	0.130	0.151

Note: Standard errors are presented in brackets and clustered at the locality level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample restricted to those living in control localities, with two parents in the household and one parent reporting to be the head, and at least one child aged between 12-15 (in November 1998 survey). Dependent variable by columns: (1) and (4): weekly household consumption less consumption of school goods/services (in pesos); (2) and (5): weekly household food consumption (in pesos); (3) and (6): household spending in the previous year on agricultural inputs (in pesos). Controls not shown here included in all four regressions: mom's education in years, dad's education in years, mom's age in years, dad's age in years, family size dummies (family size calculated as both parents plus all children aged 18 years and younger in a household), dummies for number of children under the age of 7, dummies for number of children aged 12-15, marginality index of the locality, and state fixed effects.

Our results suggest that, conditional on household poverty, households sending a larger share of their secondary school age children to school consume less in terms of consumption of all items (see columns 1 and 4 of Table 10), consumption of food (columns 2 and 5 of Table 10), and consumption of agricultural investments (see columns 3 and 6 of Table 10). Of course, it is difficult to interpret these OLS estimates as causal given potential omitted variable bias. The most concerning omitted variable is an accurate measure of parental income. More specifically, we do not expect our poverty score to capture shocks to parental income. However, we suspect this omitted variable problem to bias the estimates on schooling upwards (i.e. to be more positive). Why? Conditional on a level of poverty (which one could view as the stable component of parental income), those households that experience a positive shock to parental income would be more likely to send their children to school (given we expect schooling to be a normal investment), and have higher consumption. Thus, we take the above results as suggestive evidence that schooling is a lumpy investment that induces a discrete loss of household consumption.

A.17 Robustness of Regressions 9 and 10 to additional sibling controls

Table 11: Constant Income and Price Effects

	(1)	(2)
tc	0.00824*** (0.001)	0.00756*** (0.001)
tu	0.00398* (0.002)	0.00394* (0.002)
tc (sibs∈[12,15])		0.00225 (0.001)
Observations	21570	21570
R-squared	0.157	0.157

Note: Robust standard errors are presented in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $Enroll_{it}$. Sample for columns (1)-(3): 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 to child i (using the 1998 semester 1 transfer schedule). tu denotes the weekly, per-capita transfer to all siblings under the age of 12. tc (sibs [12-15]) denotes weekly, per-capita transfer to all siblings aged 12-15 years (using the 1998 semester 1 transfer schedule). Columns (1) and (2) include child's highest grade completed dummies, age dummies, grade interacted with age, number of siblings aged 0-5, number of siblings aged 6-7, number of siblings aged 8-9, number of siblings aged 10-11, number of siblings aged 12-13, number of siblings aged 14-15, number of siblings aged 16-18, number of siblings in grades 0-5, number of siblings in grades 6-8, number of siblings in grades 9 plus, year dummies, and a child fixed effect.

A.18 Predicting Child Income

We estimate the following regression on all individuals aged 12-18 (inclusive) in 1997, 1998, and 1999:

$$\log(hw_{it}) = \alpha_1 + \alpha_2 age_{it} + \alpha_3 age_{it}^2 + \alpha_4 boy_i + \alpha_5 margindex_l + \delta_{st} + e_{it}$$

where hw denotes the hourly wage, δ_{st} denotes year-state fixed effects and $margindex_l$ denotes the locality level marginality index score created by Progresa officials. For each child in our sample, we predict their hourly wage, and construct their weekly earnings, y_c , by assuming all children work 40 hours per-week (=mean hours worked for those children in the labor force - note median hours worked is 48 so we think this will provide an underestimate on earnings which will therefore provide a conservative estimate of the targeting benefit). Regression results are presented in Table 14, and average predicted incomes for 12-15 year-old boys and girls are presented in Table 13 below.

Table 12: Predicting Child Income: Mincer Regression

	(1)
age	0.208*** (0.061)
age2	-0.00582*** (0.002)
boy	0.0922*** (0.013)
margindex	-0.0979*** (0.007)
Observations	9146
R-squared	0.160

Note: Robust standard errors are presented in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: log hourly wage. Regressors not show: year \times state fixed effects. Margindex denotes the locality level marginality index score created by Progresa officials. Estimated on all individuals aged 12-18 years inclusive reporting positive hourly wages over 1997, 1998, and 1999.

Table 13: Summary Statistics: Predicted Child Incomes

Children aged 12-15	Mean	Std. Dev.
Predicted income (boy)	126.3	26.7
Predicted income (girl)	115.8	24.2

Note: Weekly, predicted incomes in pesos.

A.19 First Stage Regressions: IV Estimation of Equation 11

	(1)	(2)	(3)
hw	1.757*** (0.331)	-15.95*** (1.843)	-4.620*** (0.763)
tc \times hw	-0.0802 (0.068)	6.245*** (0.717)	-0.190 (0.172)
tu \times hw	0.139 (0.108)	-0.845 (0.990)	5.756*** (0.590)
Observations	17691	17691	17691
R-squared	0.0194	0.606	0.655
Fstat on instruments	11.81	80.52	57.52

Note: Robust standard errors presented in parentheses. Dependent variable by column: column (1) father's weekly income, in pesos, per-capita, y^f ; column (2) $t_c \times y^f$; column (3) $t_u \times y^f$; Regressors not shown: $t_c, t_u, t_c^2, t_u^2, t_c \times t_u, t_c \times age, t_u \times age, t_c \times boy, t_u \times boy$, age dummies, grade dummies, age interacted with grade, year and child fixed effects.

A.20 Heterogeneous Income and Price Effects: Equation 11

Table 14: Heterogeneous Income and Price Effects

	(1)	(2)
tc	0.0416** (2.09)	0.0440** (2.08)
tu	0.00868 (0.26)	0.0159 (0.44)
tc × tc	-0.000158 (-0.47)	0.000328 (0.73)
tu × tu	-0.00144* (-1.75)	-0.00134 (-1.51)
tc × tu	-0.000317 (-0.43)	-0.000177 (-0.23)
tc × y^f	-0.0000503 (-1.33)	-0.000370** (-2.07)
tu × y^f	0.0000920 (1.00)	-0.0000676 (-0.18)
tc × age	-0.00202 (-1.43)	-0.00184 (-1.24)
tu × age	0.000283 (0.12)	-0.000115 (-0.05)
tc × boy	-0.00113 (-0.47)	-0.00110 (-0.44)
tu × boy	-0.000516 (-0.11)	-0.0000871 (-0.02)
y^f	0.000162 (0.50)	0.00346 (0.89)
Observations	17365	17279
R-squared	0.146	0.131
Pvalue on interactions	0.354	0.0566

Note: Robust standard errors 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 weekly father's income per capita; tc (tu) denotes the weekly, offered per-capita CCT (UCT). Column (1): OLS; Column(2): instrument father's income per-capita and all interaction terms with father's income per-capita with average locality wage (not including father's wage). Regressors not shown: highest grade completed dummies, age dummies, grade interacted with age, year dummies, and a child fixed effect.