Extrapolation of Treatment Effect Estimates Across Contexts and Policies: An Application to Cash Transfer Experiments*

Kensuke Maeba

Northwestern University

December 26, 2022

Click here for the latest version.

Abstract

Predicting the effects of a new policy often relies on existing evidence of the same policy in other contexts. However, when those contexts are not comparable, one might want to make predictions based on similar policies in one's own context. This paper compares the performance of these approaches in the case of conditional cash transfers (CCTs). Using cash transfer programs in Malawi and Morocco, I predict the average treatment effects of Moroccan CCTs on school enrollment rates based on either Malawi CCTs or Moroccan labeled cash transfers (LCTs). I show that predictions based on the Moroccan LCTs (across policies) are more accurate than the Malawi CCTs (across contexts). To shed light on what causes the difference, I estimate a dynamic model of schooling decisions under each of these interventions separately and compare the estimated parameters across the interventions. I find that the perceived returns to schooling relative to outside options explain the differential predictions. I suggest that the perceived relative returns to schooling are context-dependent so that it is difficult to predict schooling decisions across the contexts.

*I am very grateful to Seema Jayachandran, Lori Beaman, Christopher Udry, and Vivek Bhattacharya for their invaluable guidance. I thank Eduardo Campillo Betancourt, Paul Kim, Ola Paluszynska, and seminar participants at Northwestern University and Development Day for helpful comments. Juliana Sánchez Ariza provided excellent research assistance. All errors are my own. Contact: kensukemaeba2022@u.northwestern.edu

1 Introduction

Predicting the effects of a new policy is a central challenge for policymakers. Such predictions are often made by extrapolating from existing evidence. Whether one can make accurate predictions via extrapolation depends on the internal and external validity of evidence. While the internal validity of existing evidence has improved significantly due to methodological development in causal inference, the external validity has been still concerned. In particular, previous literature on out-of-sample predictions has examined the extrapolation of evidence within policies across contexts (i.e., across-contexts extrapolation).[1] However, little is known about whether one can also learn from the extrapolation of evidence within contexts across policies (i.e., across-policies extrapolation). This paper sheds light on the usefulness of across-policies extrapolation in the case of conditional cash transfer programs intended to improve education outcomes.

Accurate predictions about the effects of a new policy help policymakers make informed policy decisions through a precise calculation of its cost-effectiveness. Conditional cash transfers (CCTs) illustrate that cost-effectiveness is a key component of policymaking. CCTs have been a widespread anti-poverty policy in developing countries, which provide cash to poor households conditional on behavioral requirement. The most common version requires children to attend school regularly, which theoretically increases school enrollment through an income and substitution effect by lowering the price of schooling.[2] CCTs are, however, not necessarily a cost-effective policy because of high administrative costs arising from frequent monitoring of school attendance (Fiszbein and Schady, 2009). Therefore, for policymakers in developing countries who consider CCTs to improve domestic education, it is particularly important to quantify the effects of CCTs beforehand to determine whether they are worth implementation.

When microdata are available, a typical approach to predict the effects of CCTs in a new context is to extrapolate from CCTs in neighboring contexts, assuming that conditional on various covariates, treatment effects are the same across contexts.[3] However, geographic proximity does not guarantee that these contexts are comparable in other aspects. Furthermore, finding CCTs in neighboring contexts is not always plausible as they have been implemented predominantly in Latin America. Therefore, it might be necessary to extrapolate from CCTs

[1] I use the word "contexts" interchangeably with locations.

[2] The income effect of CCTs is positive on schooling if education is a normal good. The substitution effect is also positive as CCTs raise the opportunity costs of non-schooling. Overall, CCTs should increase school enrollment.

[3] An even simpler approach without microdata is to make predictions based on a collection of treatment effect estimates of CCTs in other places. I show its prediction performance in the Appendix.

in different continents, where contexts are likely to be dissimilar. These reasons motivate us to extrapolate from a different policy in the context of interest. If this alternative policy induces behavioral responses we expect under CCTs, predictions based on that policy would be informative. For example, positive shocks to household income arguably increase school enrollment through a similar income effect even if the source of the income shocks is quite different (Foster and Gehrke, 2017; Shah and Steinberg, 2017). An open question is whether these two extrapolations - across-contexts extrapolation and across-policies extrapolation - result in different predictions.

I tackle this question by analyzing cash transfer experiments in Malawi and Morocco (Baird et al., 2011; Benhassine et al., 2015). Both experiments implemented CCTs that provided cash to households conditional on regular school attendance. In addition, the Moroccan experiment had another cash transfers (LCTs), which provided cash irrespective of school attendance. The two experiments also differ in target populations. While the Malawi experiment targeted girls entering secondary school, the Moroccan one studied both boys and girls in primary education.

Despite the differences in target populations, I choose these experiments for my research setting for several reasons. First, both experiments were done in the form of RCTs, providing reliable estimates of the treatment effects with a credible source of identifying variation. The exogenous variation allows my extrapolation methods to derive causal interpretations so that I can focus on the external validity of the experiments. Second, these microdata are publicly available. Finally, the Moroccan experiment is a rare example that offered alternative cash transfers adjacent to CCTs, which provides an opportunity to do the across-policies extrapolation. I then select the Malawi experiment because it satisfies the aforementioned two conditions and serves as the closest comparison to the Moroccan experiment in terms of the intervention year and geographical proximity.

By combining these experiments, I conduct a prediction exercise. Specifically, I predict the average treatment effect (ATE) of the Moroccan CCTs on school enrollment rates by extrapolating from the other two interventions. The across-contexts extrapolation corresponds to extrapolation from the Malawi CCTs while the across-policies extrapolation is based on the Moroccan LCTs. After making these predictions, I then evaluate them against the actual estimate of the ATE. In my data, I estimate that the Moroccan CCTs increased enrollment rates by 5.7 percentage points from a base of 89.4 percent.

The across-contexts extrapolation often uses reduced-form approaches. I use two off-the-shelf approaches that are commonly used: heterogeneous treatment effects approach and propensity score weighting. The first approach estimates a linear heterogeneous treatment effect model using the Malawi CCTs' sample and applies it to the Moroccan CCTs' sample. The second approach reweights the Malawi CCTs' sample by the inverse of the probability of being in the Malawi CCTs' sample. A key assumption behind these approaches is that conditional on observables, potential outcomes are independent of contexts so that policy effects are the same across contexts. To the extent this assumption is satisfied, the across-contexts extrapolation makes accurate predictions.

The across-policies extrapolation, on the other hand, requires a structural model to describe how CCTs and LCTs affect schooling decisions. I thus construct a dynamic discrete choice model of schooling decisions in which a child (or their parents) chooses whether to go to school over a finite time horizon. When the child chooses schooling, then he or she pays school costs and obtains a year of education. When not, then the child consumes income. At the terminal period, the child receives lump-sum returns based on years of education accumulated by that time. The model treats CCTs and LCTs as exogenous shocks that relax flow budget constraints but through different channels: CCTs lower school costs while LCTs increase (per-capita) household income. When extrapolating from the Moroccan LCTs, I estimate two parameters that govern schooling decisions in my model - the flow utility cost of schooling and the returns to schooling relative to outside options - by leveraging the shocks to flow budget constraint and simulate schooling decisions for the Moroccan CCTs' sample. The predictions would be accurate if the substitution effect of CCTs is small relative to the income effect and the income effect of CCTs is similar to that of LCTs.

My main extrapolation results show that the across-policies extrapolation outperforms the across-contexts extrapolation. The across-contexts extrapolation makes diverging predictions. While the heterogeneous treatment effects approach predicts that the ATE is greater than 20 percentage points, the propensity score weighting approach finds a null effect. On the other hand, the across-policies extrapolation makes accurate predictions. It yields a 95% confidence interval containing the target estimate, although it overpredicts the base enrollment rates by 5 percentage points.

Next, I investigate what causes the differential predictions between the two extrapolations. First, I show that the conclusion is not driven by the extrapolation methods. By applying the same structural approach to the across-contexts extrapolation, I compare the two extrapolations fixing the extrapolation methods. I find that while the across-contexts extrapolation predicts the ATE within a 1.5 percentage points difference using the structural approach, the predicted ATE is still statistically different from the estimated one at a significance level of 0.05. Moreover, the base enrollment rates are underpredicted by around 20 percentage points. Thus, the across-policies extrapolation still outperforms the across-contexts extrapolation.

Second, I investigate the prediction differences through the lens of the structural model. By estimating my model separately under each of the interventions, I obtain three sets of the parameter estimates and compare them across the interventions to determine which parameter accounts for the extrapolation results. I find that the estimates of the utility cost of schooling are numerically closer across the contexts than across the policies. This implies that the perceived relative returns to schooling should vary similarly across the policies to explain the outperformance of the across-policies extrapolation.

I verify this similarity across the policies by showing that the perceived relative returns to schooling under the Moroccan CCTs are more accurately recovered across the policies than across the contexts. In particular, while the predicted values from the Moroccan LCTs are decreasing in years of education and are parallel to the estimated values under the Moroccan CCTs, those from the Malawi CCTs are increasing. I argue that this difference stems from the attractiveness of outside options during primary education in each context. When outside options become more and more beneficial as children advance to higher grades so that they consider dropout, the perceived relative returns to schooling could be decreasing in years of education. In other words, when dropout becomes a realistic option, the perceived relative returns to schooling can be a downward sloping curve. Consistent with that, children in the Moroccan experiment were at increasing risk of dropping out from primary education. In contrast, because children in the Malawi experiment mostly completed primary education, the perceived relative returns to schooling estimated under the Malawi CCTs are increasing rather than decreasing. Because the across-contexts extrapolation fits this increasing part to the Moroccan CCTs' sample, the perceived relative returns to schooling extrapolated from the Malawi CCTs become increasing. It is worth noting that while the perceived relative returns to schooling for the Malawi CCTs' sample are increasing in primary education years, they start decreasing after primary education, coinciding with the timing when the dropout rates started to rise.

Finally, I discuss whether the across-contexts extrapolation can make better predictions by improving the extrapolation of the perceived relative to schooling. I first redo the extrapolation by rescaling years of education of the Malawi children so that the support of years of education overlaps across the contexts. This modification allows me to use the decreasing part of the perceived relative returns to schooling when extrapolating them across the contexts. While this modification substantially improves the prediction of the base enrollment rates, the prediction of the ATE is still statistically different from the target estimate. I then explore further improvement by restricting my samples in terms of children's age and sex based on the idea that the subpopulations might have similar expectation about the relative returns to schooling. However, this does not improve predictions additionally.

I provide two suggestive explanations for the remaining performance gap between the two extrapolations. First, the effects of the two Moroccan interventions on school enrollment rates might be largely driven by a non-pecuniary mechanism so that two interventions would be more similar than the two CCTs across the contexts. Benhassine et al. (2015) discuss that through the endorsement of the experiment by the Ministry of Education, a signaling effect was attached to the interventions that education was important. If it was a main driver of the education effects of the two interventions, then they were similar in policy characteristics. Therefore, it would be easy to extrapolate between these interventions. Second, outside options meaningfully differ across the contexts. Because the perceived relative returns to schooling are defined relative to outside options, differences in outside options could explain why it is difficult to extrapolate them across the contexts. By looking at primary reasons for school dropout in each context, I find that the Malawi sample quit school mostly due to teenage pregnancy or marriage while the Moroccan sample was more likely to do so due to other reasons such as domestic work. This difference suggests that children would follow different trajectories when they choose dropout and is reflected in the perceived relative returns to schooling.

This paper is broadly related to the literature on the empirical investigation of the validity of out-of-sample predictions. Previous research in this literature tends to focus on extrapolation from one context to another for the same policies and discuss various approaches to account for differences between the contexts (e.g. Hotz et al. 2005; Stuart et al. 2011; Andrews and Oster 2019; Meager 2019; Rosenzweig and Udry 2020; Vivalt 2020; Bandiera 2021; Dehejia et al. 2021; Gechter 2022; Meager 2022).[4] Those approaches, however, may not perform well

[4]Several papers scrutinize extrapolation in different directions. Banerjee et al. (2017) summarize difficulties in extrapolation from local contexts to global ones. DellaVigna and Linos (2022) compare nudge experiments conducted by academic researchers with those by non-academic institutions.

because the target context could be sufficiently different from the sample contexts, even after controlling for various observable characteristics. For example, Hotz et al. (2005) find inaccurate predictions of the effects of job training programs for workers without working experience based on observationally similar workers in training sites. Allcott (2015) provides evidence on site selection bias that earlier experimental sites are positively selected. In that case, exploiting variation within contexts could be an alternative way to make accurate predictions. This paper is the first study to provide one such example and discusses when extrapolation across policies outperforms the one across contexts.

The closest comparison to this paper is Gechter et al. (2018). They evaluate various extrapolation methods based on a welfare measure in the case of CCTs. In particular, they compare reduced-form approaches for extrapolation from experimental data in a different context with structural ones for extrapolation from pre-treatment data in the target context, alluding to a trade-off between internal validity and external validity. Pritchett and Sandefur (2015) also focus on this trade-off in the case of microcredit and quantify it using the root mean squared error of the treatment effects. This paper is different in that evidence used in the two extrapolations is internally valid as it comes from experimental data. Instead, I compare the external validity of the evidence and demonstrate the usefulness of the evidence about the adjacent policy.

This paper also makes methodological contributions to two strands of literature. First, it presents another application to the literature on structural models estimated with RCTs (Todd and Wolpin, 2006; Attanasio et al., 2012; Duflo et al., 2012). Those papers use randomized treatment assignment to evaluate the fitness of their models or to identify experiment-specific models. This paper deviates from them by using RCTs to identify a general model that is applicable to any cash transfers intended to improve school enrollment. This flexibility allows me to use an identical model for both extrapolations to examine prediction differences due to the selection of extrapolation methods.

Second, this paper builds on recent developments in the literature about dynamic discrete choice models. The literature on dynamic discrete choice models started from Rust (1987) and has developed new estimation methods that ease computational burdens by avoiding solving full models (Hotz and Miller, 1993; Hotz et al., 1994; Aguirregabiria and Mira, 2002; Arcidiacono and Ellickson, 2011; Arcidiacono and Miller, 2011). Recently, Scott (2014) introduced a new estimation method emphasizing identification, which was later formalized by Kalouptsidi et al.

(2021). I take this approach to discuss causal identification of my model as well as relax rational expectations assumption by exploiting randomized treatment assignment. Relaxing this assumption makes my model more realistic, especially because poor households in developing countries might have biased perceptions about returns to schooling due to information frictions (for example, Jensen 2010). Finally, among a few papers using this identification strategy, such as De Groote and Verboven (2019); Diamond et al. (2019); Traiberman (2019), my paper is the first application to schooling decisions. Since human capital investment decisions are one of the most popular topics analyzed with dynamic discrete choice models, my paper is a benchmark that achieves a new identification for a potentially large number of subsequent papers on this topic.

The remainder of the paper is structured as follows. Section 2 briefly describes the Malawi and Moroccan experiments. Section 3 introduces extrapolation methods. Section 4 shows estimation results. Section 5 presents extrapolation results. Section 6 concludes.

2 Two Cash Transfer Experiments

2.1 Malawi experiment

Baird et al. (2011) implemented a cash transfer experiment in 176 subdistricts (enumeration areas or EAs) of Zomba district in Malawi. The experiment offered two types of cash transfers, conditional or unconditional on regular attendance (CCTs and UCTs, respectively).[5] The goal of the paper was to examine the effects of the conditionality attached to cash transfers on schooling. The target population of the experiment was the never-married girls of age 13 to 22 at school who were at the risk of dropping out of secondary school due to pregnancy or early marriage, or both. The authors sampled 16.5 school girls per EA on average and obtained 2087 school girls in total for analysis. They conducted a baseline survey between October 2007 and January 2008, the first follow-up survey between October 2008 and February 2009, an endline survey between February and June 2010.

The randomization for the interventions was conducted at the EA level. Out of 176 EAs, 46 EAs were in the CCTs arm (470 girls), 27 EAs in the UCTs arm (261 girls), 88 EAs in the control arm (1356 girls), and 15 EAs to examine spillover effects were dropped. Cash transfers

[5]The conditionality of the CCTs was to attend more than 80% of schooling days. Cash transfers were made monthly if children satisfied this requirement in the previous month.

were offered at both households and individuals level. The household amount varied across the EAs while the individual amount did within the EAs. To benchmark the size of the transfer, based on the authors' calculation, the household amount for 10 months was on average about 10% of the average household expenditure. The cash transfers were distributed every month for two years (2008 and 2009).

The main empirical results in Baird et al. (2011) showed that students who received CCTs were more likely to stay in school by 4.5 percent (or 4.1 percentage points) than those in the control group after 1 year and by 7.9 percent (or 6.1 percentage points) after 2 years, based on the self-reported school enrollment in the household surveys. The authors also estimated the effects of UCTs, but found that they were biased by misreporting: when using teacher-reported school enrollment as an outcome, the effects of UCTs were not statistically different from 0 while the effects of CCTs remained largely the same.

In what follows, due to this misreporting bias, I use only the CCTs treatment group and the control group for my analysis. As I will discuss in the next subsection, the misreporting behaviors were not observed in the Moroccan experiment. Hence, the underlying mechanisms behind the effects of UCTs on self-reported enrollment rates are not applicable across the policies and the contexts, which makes the Malawi UCTs inappropriate for the following extrapolation exercise.

2.2 Moroccan experiment

Benhassine et al. (2015) implemented a cash transfer experiment covering 600 poorest rural municipalities in 5 poorest regions of Morocco. This experiment also had two treatment arms, CCTs and labeled cash transfers (LCTs).[6] Unlike the CCTs, the LCTs were not tied to regular school attendance. However, like the CCTs, the LCTs were presented as an education program because children were registered to the experiment at schools in their residential areas, which implicitly encouraged school enrollment. The authors investigated whether this nudging was sufficient to change schooling decisions or not. The target population was children of age 6 to 15 who were at the risk of dropping out of primary school. At each school, the authors sampled 8 households that had at least one child in that school past three years, resulting in 4385 households in total. They conducted a baseline survey in June 2008 and an endline survey in June 2010.

[6] The conditionality of CCTs was not to miss school more than 4 times every month.

The randomization for the interventions was conducted at the school area level. Out of 320 school areas, 260 were assigned to the treatment group and 60 to the control group. The treatment group was further divided into 4 subgroups based on whether the cash transfers were the CCTs or the LCTs and whether the recipient was a father or mother. Since the sex of the recipients is not the focus of my analysis, I pool the subgroups and create the CCTs and LCTs treatment groups. The average monthly transfer amount was about 5 percent of the average monthly household consumption. Thus, the average transfer size was, roughly speaking, smaller than the one in the Malawi experiment. The cash transfers were made every 3 or 4 months through 2009 and 2010.

The main empirical findings in Benhassine et al. (2015) showed that children who received the LCTs were more likely to enroll in school approximately by 10 percent (or 7.4 percentage points) than those in the control group after 2 years. In contrast, those who received the CCTs were more so roughly by 7.3 percent (or 5.4 percentage points). Thus, the LCTs were more effective in increasing enrollment rates than the CCTs. The authors discussed two reasons for the similar effect size of the two interventions. One is confusion about the conditionality of the CCTs. Specifically, less than 15% of the parents of the treated children under the CCTs believed that the cash transfers were conditional on some attendance measures, which made the CCTs similar to the LCTs. The other reason was that the effects were driven by the signaling effect that education was important. The authors then suggested that in addition to these factors, the LCTs were relatively successful because they induced schooling among marginal children who were not confident about regular school attendance so that they would not choose schooling under the CCTs.

2.3 Quasi-prediction problem

Given that the Malawi UCTs are not ideal for extrapolation due to the misreporting bias, I set up the prediction exercise as follows. The target object is the average treatment effect (ATE) of the Moroccan CCTs on enrollment rates. The across-contexts extrapolation uses the Malawi CCTs while the across-policies extrapolation does the Moroccan LCTs. In both extrapolations, I use standard approaches to extrapolate from those interventions. Specifically, for the across-contexts extrapolation, I use two reduced-form approaches: heterogeneous treatment effects approach and propensity score weighting. On the other hand, for the across-policies extrapolation, I use a structural model of schooling decisions and make predictions by running counterfactual simulations. In the next section, I will explain those extrapolation methods in detail. Finally, I evaluate the two predictions against the estimated ATE of the Moroccan CCTs. The evaluation is possible because I have data on the Moroccan CCTs.

3 Extrapolation Methods

3.1 Reduced-form approaches for across-contexts extrapolation

3.1.1 Covariates selection

I choose two standard approaches for the across-contexts extrapolation: heterogeneous treatment effects approach and propensity score weighting. The key assumption for these approaches to be able to transfer treatment effect estimates across contexts is that potential outcomes are independent of contexts conditional on a set of observables (Hotz et al., 2005). To satisfy this assumption in my research setting, I first choose children's age and sex for the conditional variables as the target populations in the two experiments differ in them. Additionally, motivated by my structural model below, I also include pre-treatment variables that are likely related to schooling decisions; years of education, per-capita income, school costs, and cash transfer amount. If the assumption is satisfied, then conditional on those variables, the ATE of the Moroccan CCTs should coincide with that of the Malawi CCTs.[7]

3.1.2 Heterogeneous treatment effects approach

A common approach to extrapolate across contexts is the linear projection of heterogeneous treatment effects. I first estimate heterogeneous treatment effects based on the pre-selected covariates, denoted by $W_i = (w_{i1}, \ldots, w_{iK})'$, using the second round of the Malawi CCTs data:

$$d_i = W_i'\beta^{\text{HTE}} + \beta_0^{\text{HTE}}\text{Treatment}_i + \sum_{k=1}^{K} \gamma_k^{\text{HTE}}\text{Treatment}_i \times w_{ik} + \omega_i,$$

where d_i is a dummy variable taking 1 if the child i enrolls in school and 0 otherwise. With the estimated parameters, enrollment decisions are predicted for the Moroccan CCTs' sample,

[7]Another standard approach is entropy balancing, where new weights for the training data are computed to match key moments of selected covariates in the target data (Hainmueller, 2012). Allcott (2015) uses this approach in addition to the heterogeneous treatment effects one. One problem about entropy balancing is that the computation may not converge, which is the case in my data with those variables.

denoted by $\tilde{d}_i$. Then the ATE is predicted by regressing them on the treatment assignment and the sampling strata fixed effects:

$$\tilde{d}_i = \alpha_0^{\text{HTE}} + \alpha_1^{\text{HTE}} \text{Treatment}_i + \text{Stratum}_i + \nu_i^{\text{HTE}}.$$

The standard errors are clustered at the randomization units.

3.1.3 Propensity score weighting

Another common approach is propensity score weighting (Stuart et al., 2011). This approach first pools the Malawi CCTs' and Moroccan CCTs' second-round data and estimates the probabilities of being in the Malawi CCTs' sample as a function of the selected covariates:

$$\mathbf{1}\{i \in \text{Malawi CCTs}\} = W_i' \beta^{\text{PSW}} + \beta_0^{\text{PSW}} \text{Treatment}_i + u_i.$$

I assume that u_i is a logit error. Then, using the Malawi CCTs' sample weighted by the inverse of the propensity scores, I estimate the predicted ATE by regressing the enrollment decisions on the treatment assignment and the sampling strata fixed effects:

$$d_i = \alpha_0^{\text{PSW}} + \alpha_1^{\text{PSW}} \text{Treatment}_i + \text{Stratum}_i + \nu_i^{\text{PSW}}.$$

The standard errors are clustered at the randomization units.

3.2 Structural approach for across-policies extrapolation

3.2.1 Dynamic model of schooling decisions

To extrapolate across the policies, I use a structural model that describes schooling decisions under the CCTs and the LCTs. I construct a dynamic discrete choice model where a child i of school age (or a household i with a school-age child) decides whether to go to school ($d = 1$) or not ($d = 0$) every period t. If the child chooses schooling, then he pays school costs s out of his per-capita income y, accumulates one year of education, and consumes the rest of the income. If not, then the child consumes all of the income. I define the school costs as payments for tuition fees, textbooks, uniforms, and any other necessary goods, varying by grade. Unlike life-cycle models, the income is determined outside the model, although I treat it as potentially endogenous when estimating the model. The model is dynamic because at the terminal period $(t=T)$, the child receives lump-sum returns to education based on the number of years of education e he has had, which creates an inter-temporal trade-off of consumption.[8]

The child maximizes the discounted sum of utilities by making a series of schooling decisions. Formally, the maximization problem is defined as follows:

$$\begin{aligned}
&\max_{\{d_{i\tau}\}_{\tau=t}^{T-1}} E\left[\sum_{\tau=t}^{T-1} \beta^{\tau-t}\left\{\theta \ln\left(c_{i\tau}\right)+\varepsilon_{i\tau}\left(d_{i\tau}\right)\right\}+\beta^{T-t} R\left(e_{i,T}; s_{iT}, y_{iT}\right) | e_{i\tau}, y_{i\tau}, s_{i\tau}, \varepsilon_{i\tau}\right] \\
&\quad \text{s.t. } c_{i\tau}=y_{i\tau}-d_{i\tau} s_{i\tau} \\
&\qquad\quad e_{i\tau}=e_{i,\tau-1}+d_{i,\tau-1},
\end{aligned}$$

where β is a discount factor, c is total consumption, and $\varepsilon(d)$ is a choice-specific preference shock.

The model parameters to estimate are θ, the marginal utility of consumption at a given consumption level, and $R(e;y,s)$, (perceived) lump-sum returns to education the child receives at the terminal period.[9] The returns to education are indexed by (y,s) as I can allow them to vary by the state values.

For future convenience, I rewrite this problem for $t<T$ using the Bellman equation:

$$V\left(e_{it}, x_{it}, \varepsilon_{it}\right)=\max_{d} \ \theta \ln\left(y_{it}-d s_{it}\right)+\varepsilon_{it}(d)+\beta E_{\varepsilon,x}\left[V\left(e_{i,t+1}, x_{i,t+1}, \varepsilon_{i,t+1}\right) | e_{it}, x_{it}, \varepsilon_{it}(d), d\right],$$

where $x_{it}=\left(y_{it}, s_{it}\right)'$. The state variables in this model are (e,x,ε), where an econometrician can observe (e,x) while the child can observe (e,x,ε).

[8] I assume the child does not save. This is consistent with low rates of having any saving technologies in both contexts at baseline.

[9] I call $R(e;y,s)$ perceived returns to education to distinguish from actual returns to education, which I cannot estimate with my data.

3.2.2 Identification of model parameters

To identify the parameters, θ and $R(e;x)$, I first make parametric assumptions on the preference shocks and the discount factor.

Assumption 1. *The state transition function satisfies conditional independence:*

$$
\begin{aligned}
F(x_{t+1},\varepsilon_{t+1}|e_t,x_t,\varepsilon_t,d_t) &= F(y_{t+1},\varepsilon_{t+1}|e_t,x_t,\varepsilon_t,d_t) \\
&= F_\varepsilon(\varepsilon_{t+1})\,F_y(y_{t+1}|y_t)\,.
\end{aligned}
$$

Assumption 2. *$\varepsilon_{it}(d)$ is i.i.d across (i,t,d) and follows the type-I extreme value distribution.*

Assumption 3. *β is set at 0.95 exogenously.*

Assumption 1 makes the preference shocks serially uncorrelated and independent of the observed state variables. This means that the preference shocks are no longer a state variable. Assumption 2 gives me convenient expressions for some objects that appear later in my identification arguments.[10] Finally, Assumption 3 sets the discount factor outside the model as it is generally not identified in the standard dynamic discrete choice models (Rust, 1994; Magnac and Thesmar, 2002).

With these assumptions, I first identify θ via the Euler Equations in Conditional Choice Probabilities (ECCP) approach (Scott, 2014; Kalouptsidi et al., 2021). The ECCP approach combines the Hotz-Miller inversion with the finite dependence property to eliminate continuation values and derive a linear equation to identify θ. A key identification challenge via the ECCP approach is that errors in future expectations are correlated with state variables, which is usually addressed by assuming rational expectations. I deal with it instead by exploiting randomization in the cash transfer experiments. Since it is likely in my context that children or households have biased beliefs about the returns to education, removing the rational expectations assumption makes my model more realistic.

To derive the linear equation for θ, I first define several objects for notational convenience. The ex-ante value function is the value function integrated over the preference shocks:

[10] Note that the key part of Assumption 2 is that an econometrician knows the distribution. Thus, the following arguments will hold as long as the preference shocks are drawn from a known distribution.

$$\overline{V}(e_{it}, x_{it} : \theta) \equiv E_{\varepsilon}[V(e_{it}, x_{it}, \varepsilon_{it} : \theta) | e_{it}, x_{it}].$$

The conditional value function is defined with the ex-ante value function and Assumption 1:

$$v(e_{it}, x_{it}, d : \theta) \equiv \theta \ln(y_{it} - ds_{it}) + \beta E_x \left[\overline{V}(e_{i,t+1}, x_{i,t+1} : \theta) | e_{it}, x_{it}, d\right].$$

The conditional choice probability takes a logit form because of Assumption 2:

$$P(d = 1 | e_{it}, x_{it} : \theta) = \frac{\exp(v(e_{it}, x_{it}, 1 : \theta))}{\exp(v(e_{it}, x_{it}, 0 : \theta)) + \exp(v(e_{it}, x_{it}, 1 : \theta))}.$$

Finally, Assumption 2 simplifies the expression for the ex-ante value function:

$$\begin{aligned}
\overline{V}(e_{it}, x_{it} : \theta) &= \ln \sum_d \exp(v(e_{it}, x_{it}, d : \theta)) + \gamma \\
&= v(e_{it}, x_{it}, 0 : \theta) + \gamma - \ln P(d = 0 | e_{it}, x_{it} : \theta) \\
&= v(e_{it}, x_{it}, 1 : \theta) + \gamma - \ln P(d = 1 | e_{it}, x_{it} : \theta),
\end{aligned}$$

where γ is the Euler's constant. In what follows, I drop θ inside these objects for notational simplicity.

My identification argument starts with the Hotz-Miller inversion:

$$\ln \frac{P(d = 1 | e_{it}, x_{it})}{P(d = 0 | e_{it}, x_{it})} = v(e_{it}, x_{it}, 1) - v(e_{it}, x_{it}, 0). \tag{1}$$

The next step is to decompose the ex-ante value functions into the realized value functions and the residuals: for each $d \in \{0, 1\}$,

$$v(e_{it}, x_{it}, d) = \theta \ln(y_{it} - ds_{it}) + \beta \left(\overline{V}(e_{i,t+1}, x_{i,t+1}) + \eta_{it}(d)\right),$$

where

$$\eta_{it}(d) \equiv E_x\left[\overline{V}(e_{i,t+1}, x_{i,t+1})\,|e_{it}, x_{it}, d\right] - \overline{V}(e_{i,t+1}, x_{i,t+1}).$$

$\eta_{it}(d)$ is called expectation errors (Scott, 2014; Kalouptsidi et al., 2021). As I discuss later, the expectation errors cause endogeneity problems.

Then I use the finite dependence property (Arcidiacono and Miller, 2011). In my model, for any levels of education in period t, two sequences of choices, $(d_{it}, d_{i,t+1}) = \{(1,0),(0,1)\}$, lead to $e_{i,t+2} = e_{it} + 1$. Based on this idea, I rewrite the ex-ante value functions in the following way: if the child chooses schooling in period t,

$$\overline{V}(e_{i,t+1}, x_{i,t+1}) = v(e_{it}+1, x_{i,t+1}, 0) + \gamma - \ln P(d=0|e_{it}+1, x_{i,t+1}),$$

and similarly, if the child chooses non-schooling,

$$\overline{V}(e_{i,t+1}, x_{i,t+1}) = v(e_{it}, x_{i,t+1}, 1) + \gamma - \ln P(d=1|e_{it}, x_{i,t+1}).$$

Notice that this procedure eliminates the continuation values in period $t+1$ when subtracting one from the other. To see this, I expand the conditional value functions in the above expressions:

$$\begin{aligned} v(e_{it}+1, x_{i,t+1}, 0) &= \theta\ln(y_{i,t+1}) + \beta E_x\left[\overline{V}(e_{it}+1, x_{i,t+2})\,|x_{i,t+1}\right], \\ v(e_{it}, x_{i,t+1}, 1) &= \theta\ln(y_{i,t+1} - s(e_{it})) + \beta E_x\left[\overline{V}(e_{it}+1, x_{i,t+2})\,|x_{i,t+1}\right]. \end{aligned}$$

Thus, both conditional value functions have the same continuation values. Finally, by substituting back all of the derived expressions into the Hotz-Miller inversion in Equation (1), I obtain a regression equation to identify θ :

$$
\begin{aligned}
\ln \frac{P\left(d=1 \mid e_{it}, y_{it}, s\left(e_{it}\right)\right)}{P\left(d=0 \mid e_{it}, y_{it}, s\left(e_{it}\right)\right)}= & \theta\left\{\ln \left(y_{it}-s\left(e_{it}\right)\right)-\ln \left(y_{it}\right)\right\} \\
& +\beta \theta\left\{\ln \left(y_{i,t+1}\right)-\ln \left(y_{i,t+1}-s\left(e_{it}\right)\right)\right\} \\
& +\beta \ln \frac{P\left(d=1 \mid e_{it}, y_{i,t+1}, s\left(e_{it}\right)\right)}{P\left(d=0 \mid e_{it}+1, y_{i,t+1}, s\left(e_{it}+1\right)\right)} \\
& +\beta\left(\eta_{it}(1)-\eta_{it}(0)\right) .
\end{aligned}
$$

To tidy the equation, I define several choice probabilities:

$$
\begin{aligned}
P_{it}^{1} & \equiv P\left(d=1 \mid e_{it}, y_{it}, s\left(e_{it}\right)\right), \\
P_{i,t+1}^{2} & \equiv P\left(d=1 \mid e_{it}, y_{i,t+1}, s\left(e_{it}\right)\right), \\
P_{i,t+1}^{3} & \equiv P\left(d=1 \mid e_{it}+1, y_{i,t+1}, s\left(e_{it}+1\right)\right) .
\end{aligned}
$$

I rewrite the equation using the expressions:

$$
\ln \frac{P_{it}^{1}}{1-P_{it}^{1}}-\beta \ln \frac{P_{i,t+1}^{2}}{1-P_{i,t+1}^{3}}=\theta\left\{\ln \left(1-\frac{s\left(e_{it}\right)}{y_{it}}\right)-\beta \ln \left(1-\frac{s\left(e_{it}\right)}{y_{i,t+1}}\right)\right\}+\beta\left(\eta_{it}(1)-\eta_{it}(0)\right) . \quad (2)
$$

Assuming that I can construct the dependent variable, running an OLS regression on this equation does not provide a consistent estimator of θ as $\left(\frac{s(e_{it})}{y_{it}}, \frac{s(e_{it})}{y_{i,t+1}}\right)$ are correlated with $\eta_{it}(1)-\eta_{it}(0)$. To see this, I first decompose the expectation error into two parts. One is heterogeneity in the perceived returns to education across households. For example, high and low income households may have systematically different expectations about the future returns to education at baseline because of information frictions. The other is forecasting errors, such that if the income is higher than expected, then the forecasting errors are also larger, even when all households have the same perceived returns to education.

I argue that the treatment assignment serves as an IV to resolve this endogeneity problem.[11] The relevance condition is likely satisfied because the cash transfers are a part of the income or the school costs. The exclusion restriction for the heterogeneous perceived returns to education

[11] The cash transfer amount can also serve as an IV. However, I choose the treatment assignment to account for the nonlinear effects of cash transfers on enrollment rates.

requires the treatment assignment should be orthogonal to households' expectations about the future at baseline. This would hold true because the randomization does not allow them to select the treatment assignment based on their beliefs about the returns to education. The treatment assignment also satisfies the exclusion restriction with respect to the forecasting error as the randomization guarantees that households in the treatment and the control group, on average, make equally inaccurate predictions about the future value functions at baseline.[12]

Importantly, I can make the above arguments under any cash transfer interventions under consideration. In my model, the CCTs exogenously shift the independent variable through the school costs while the LCTs through the income. Thus, all of the cash transfers can serve as an IV to identify θ consistently. Moreover, this difference in how the cash transfers work in my model can lead to varying θ across the interventions, even if the amount of the cash transfers is the same. This is because the CCTs make larger relative changes in the school costs than the LCTs do in the income. Thus, varying estimates of θ across the experiments thus reflect the differences in policy characteristics.

After identifying θ, I proceed to identify the returns to education, $R(e;x)$. Instead of identifying this object, which requires an assumption on the functional form, I choose to identify the differential returns to education across the choices nonparametrically. The identification of the differential returns to education depends on the terminal period as they are the terminal payoffs in my model. I set the terminal period at $T = 3$, one period after the cash transfer interventions, mainly because I do not know from my data when children usually enter labor markets in both contexts, which is another candidate for the terminal period.

By setting $T = 3$, I can take a shortcut because I can use Equation (1) at $t = 2$:

$$\ln \frac{P_{i,2}^1}{1 - P_{i,2}^1} = \theta \ln \left(1 - \frac{s(e_{i,2})}{y_{i,2}}\right) + \beta \Delta R(e_{i,2}; x_{i,2}), \tag{3}$$

$$\Delta R(e_{i,2}; x_{i,2}) \equiv R(e_{i,2} + 1; x_{i,2}) - R(e_{i,2}; x_{i,2}).$$

$\Delta R(e_{i,2}; x_{i,2})$ represents the differential returns to education or the relative returns to

[12] To argue this point in detail, suppose the treatment assignment is yet to be announced. Then, if households believe they are in the treatment group and are assigned to the treatment group, they make no forecasting errors. If they are, in contrast, assigned to the control group, then they overestimate the income or underestimate the school costs, depending on the cash transfer experiments they are in. Similar reasoning can be made when they believe they are in the control group. Because the randomization makes them believe they are assigned to a certain group by chance and it does so indeed, the forecasting error would be balanced across the treatment status.

schooling, at a given level of education $e_{i,2}$. Intuitively, this parameter is identified through the residual variation of enrollment rates not explained by changes in the relative utility of schooling due to the cash transfers. Because the choice probabilities are a function of the state variables, $\Delta R(e_{i,2}; x_{i,2})$ varies across the state values.

3.2.3 Two-step estimation

To operationalize the identification arguments, I estimate my model in two steps. The first step is to estimate the choice probabilities from my data. This can be done nonparametrically if data have sufficient variation in each state, which is not the case with my data. Therefore, I estimate the choice probabilities with the flexible logit of the state variables to smooth the probabilities across the state space.[13] After obtaining the choice probabilities estimates, I run a 2SLS estimation on Equation (2) to estimate θ. Finally, I recover $\Delta R(e; x)$ from Equation (3) with the estimates of the choice probabilities and θ.

In the choice probabilities estimation, I select an estimation method based on the ability to replicate the ATEs using the estimates. Specifically, I first estimate the choice probabilities with the flexible logit of the state variables via MLE. Then I compute the shares of children choosing schooling for the treatment and the control group in each survey round and compare them with the true ones. To replicate the ATEs through my model, these shares need to be sufficiently close to each other, given the magnitude of the ATE in percentage points. This suggests that a-few-percentage-points errors could lead to a failure of the replication. If the MLE estimates of the choice probabilities are not close enough in this sense, then I estimate them using the same logit but via GMM, where I directly match those shares as the moment conditions. Because the GMM estimator might distort the individual choice probabilities to match the shares, I will check whether simulated schooling decisions based on the GMM estimates are significantly different from those based on the MLE estimates. More details about the choice probabilities estimation are available in Appendix.

3.2.4 Prediction

Having estimated θ and $\Delta R(e; x)$, the prediction of the ATE of the Moroccan CCTs is based on the parameter estimates from the Moroccan LCTs combined with $(e_{i,2}, y_{i,2}, s_{i,2})$ from the

[13]Another common way of smoothing the choice probabilities is to weigh nonparametric estimates across states. One advantage of my smoothing approach is that I do not have to discretize my state space. My approach is used, for example, in a numerical example of Arcidiacono and Miller (2011).

Moroccan CCTs. I first predict the choice probabilities:

$$\hat{P}(d=1|e_{i,2},y_{i,2},s_{i,2}) = \frac{\exp\left(\widehat{\theta^{\text{LCTs}}}\ln\left(1-\frac{s_{i,2}}{y_{i,2}}\right)+\beta\widehat{\Delta R^{\text{LCTs}}}(e_{i,2}:x_{i,2})\right)}{1+\exp\left(\widehat{\theta^{\text{LCTs}}}\ln\left(1-\frac{s_{i,2}}{y_{i,2}}\right)+\beta\widehat{\Delta R^{\text{LCTs}}}(e_{i,2}:x_{i,2})\right)}, \tag{4}$$

Using them as the dependent variable, I run an OLS regression:

$$\hat{P}(d=1|e_{i,2},y_{i,2},s_{i,2}) = \delta_1+\delta_2\text{Treatment}_i+\text{Stratum}_i+\nu_i, \tag{5}$$

where δ_1 and δ_2 represent the simulated control enrollment rates and ATE, respectively.

While extrapolating $\hat{\theta}$ from the Moroccan LCTs is easy, extrapolating $\widehat{\Delta R}(e;x)$ is not straightforward as the state space may not overlap across the policies. For example, school costs for the Moroccan CCTs' treated children can be negative if the transfers cover more than the actual costs while they are always positive for the Moroccan LCTs' sample. Thus, I need to extrapolate $\widehat{\Delta R}(e;x)$ for the Moroccan CCTs' children from the empirical distribution of $\widehat{\Delta R}(e;x)$ under the Moroccan LCTs. To do this, I take both parametric and nonparametric approaches. The parametric approach is that I construct a linear projection of $\widehat{\Delta R}(e;x)$ on $\left(e,e^2\right)$. The nonparametric one uses the Random Forest algorithm to figure out the relationship between $\widehat{\Delta R}(e;x)$ and e. Because I have no prior on which approach is superior to the other, I show all results for both approaches.

4 Estimation Results

4.1 Data

My main datasets are constructed from the publicly available data of the Malawi and Moroccan experiments (Baird et al., 2012; Ozler et al., 2015a,b; Benhassine et al., 2019).[14] I have access to several rounds of household surveys in each experiment. I use the baseline, the first follow-up, and the endline survey from the Malawi data and the baseline and the endline survey from the

[14]The Malawi data are published at Microdata Library of The World Bank. For instance, the link to the baseline survey is here: https://microdata.worldbank.org/index.php/catalog/1005. The Moroccan data are available at OPEN ICPSR: https://www.openicpsr.org/openicpsr/project/114579/version/V1/view?path=/openicpsr/114579/fcr:versions/V1&type=project.

Moroccan data. The subscript t in my notation corresponds to survey rounds of each household survey.

Next, I construct key variables, schooling decisions (d_{it}), years of education (e_{it}), per-capita income (y_{it}), school costs (s_{it}), and cash transfers (z_{it}), from the household surveys as follows. Schooling decisions are based on whether an individual is currently in school or not, which I observe in the data. I also observe years of education as current grades children are in. However, they are subject to reporting errors such as they increase by 2 years between the survey rounds. Hence, instead of using the reported grades, I construct years of education from schooling decisions and years of education at baseline as the model describes $(e_{i,t+1} = e_{it} + d_{it})$. The resulting years of education are highly correlated with the reported grades, verifying this construction.[15] The amount of the cash transfers is also observed. To convert it into an annual term, I multiply the monthly amount by 10, the number of months school was open. For the control group children, I impute 0 for their cash transfer amount.

Per-capita income is not observed in both data. I use an adjusted per-capita household expenditure as a proxy. I first construct annual household expenditures by multiplying by 12 monthly household expenditures of the month prior to the surveys. Second, I subtract the amount of the cash transfers from the annual expenditures.[16] Then I divide it by household size adjusted by the OECD equivalence scale, where the household head's consumption is twice as much and the other adult members' consumption is 1.4 times as much as a child's consumption. Finally, if a child is in the LCTs treatment group, I add the cash transfers. This assumes the cash transfers are all consumed by the eligible child in my model. Because cash transfers were provided to cover school costs, I assume that they are spent entirely for the child even if not used for schooling.

School costs are constructed from the annual individual school expenditures reported by the control group children who were in school. In particular, I create the menu of the school costs by taking the median of the school expenditures, separately for each grade, weighted by

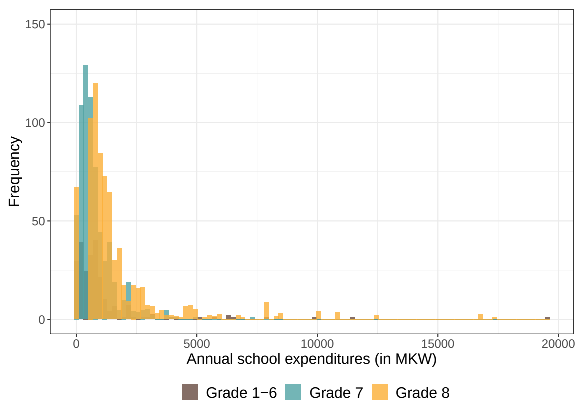
[15] Constructing years of education in this way is not straightforward in the Moroccan data as the household surveys were 2 years apart. In the data, differences in years of education between the surveys vary from 0 to 2, which requires me to choose whether in my model years of education increase by 1 or 2 if children choose schooling. I assume that years of education increase by 1 at endline if schooling is chosen at baseline. While the resulting years of education show a high correlation with the reported grades, this assumption may not be innocuous because the amount of the cash transfers is increasing in grades. Thus, not only does the average years of education become lower, but the size of the cash transfers relative to the per-capita income or the school costs becomes smaller.

[16] If children are in the LCTs treatment group, I subtract the cash transfers for everyone. If children are in the CCTs treatment group, I subtract only for those choosing schooling.

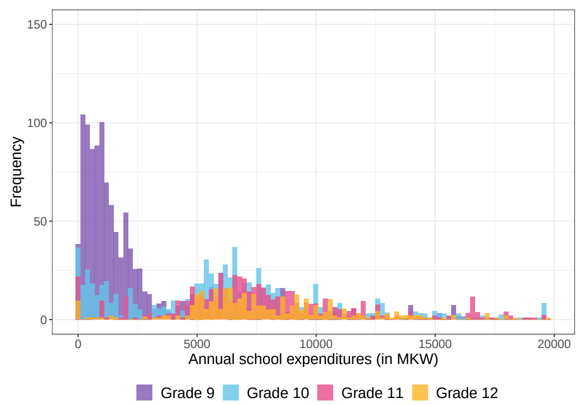
sampling weights.[17] A concern is that the school expenditures might include private investment in children's education, which would vary by households' income level. If the cross-sectional variation is large, then the same school costs for all children in the same grades are not a reasonable assumption. To mitigate this concern, Figure 1-a, 1-b, and 1-c show the distributions of the school expenditures. I find the skewed distributions for all grades, suggesting that the school expenditures did not vary largely across households.[18] Finally, I subtract the cash transfers from the school costs if children are in the CCTs treatment group.

Figure 1: Distributions of school expenditures

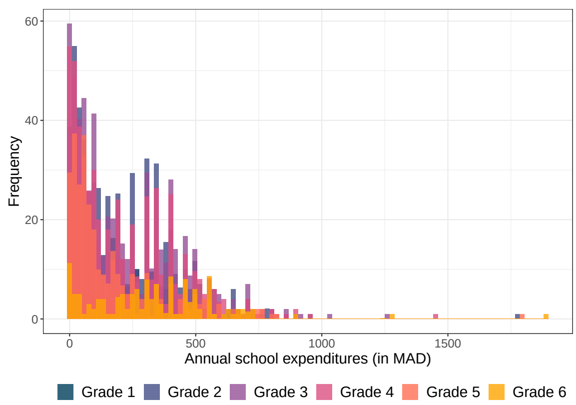
1-a: Malawi - Primary school



1-b: Malawi - Secondary school



1-c: Morocco - Primary school



Note: I use the children in the control group and pool them across the survey rounds. For the Malawi figures, I drop the outliers with the expenditures higher than 20000 Malawi Kwacha (≈ 5%). For the Moroccan figure, I drop the outliers with the expenditures higher than 2000 Moroccan Dirham (≈ 0.1%).

A few more remarks about the construction of the datasets are as follows. First, I measure the

[17] I group grade 1 to grade 6 into one in the Malawi data because of the small number of observations in those grades. Thus, children in those grades face the same school costs.

[18] Because there are outliers in the reported school expenditures, I use the median instead of the mean.

income and the school costs in 100 USD in 2008 to keep the same units across the experiments. One USD in 2008 was 140 Malawian Kwacha and 7.75 Moroccan Dirham.[19] Second, I remove observations who were in the final grade of secondary school in the Malawi data and primary school in the Moroccan data at baseline as their school cost estimates were noisy due to a small number of observations. Third, I make balanced panel data for each experiment with no missing values in any of the variables I created above.

I show the summary statistics of the created variables in Table 1 to check whether they are balanced at baseline for each experiment. Within the contexts, most of the variables are balanced across the treatment status at baseline. Across the contexts, I notice several differences. First, the average years of education are higher in the Malawi experiment than the Moroccan one because of the different targeting populations. That is, the Malawi sample is only girls who were about to enter secondary school and is older than the Moroccan one. For the same reason, the sex ratio and children's age are different across the experiments. Second, the income and the school costs are on average higher in the Moroccan experiments, suggesting that the Moroccan sample is richer. Finally, while the average cash transfer amount is similar across the experiments, its relative size to the income or school costs is different. The relative size of the cash transfers is larger in the Malawi experiment.

Finally, I report the ATE of each intervention on enrollment rates in my data. The ATEs are estimated by a reduced-form regression that is similar to the specifications used in Baird et al. (2011); Benhassine et al. (2015). I use the second round of the household surveys and regress the schooling decisions on a dummy variable for the treatment group and the sampling strata fixed effects:

$$d_{i,2} = \alpha_1 + \alpha_2 \text{Treatment}_i + \text{Stratum}_i + \nu_i. \tag{5'}$$

The estimation results in Table 2 are aligned with the original findings. In my samples, the Malawi CCTs increased enrollment rates by 4.1 percent (or 3.7 percentage points). In contrast, the Moroccan CCTs did by 6.4 percent (or 5.7 percentage points), and the Moroccan LCTs did by 8.2 percent (or 7.3 percentage points). As I mentioned in Section 2, the estimates of the ATEs in the original papers varied from 4 percentage points to 7 percentage points. Moreover,

[19]The conversion rates are from The World Bank data: https://data.worldbank.org/indicator/PA.NUS.FCRF?end=2021&locations=MA-MW&start=1960.

Table 1: Summary statistics of key variables at baseline

	Malawi		Morocco		
	(1) Control	(2) CCTs	(3) Control	(4) CCTs	(5) LCTs
= 1 if enrollment	1.000	1.000	0.909	0.921	0.920*
Years of education	8.046	7.960	2.755	2.776*	2.764
Per-capita income (in 100 USD)	1.173	1.571	5.368	5.335	5.345
School costs (in 100 USD)	0.123	0.124	0.213	0.212	0.212
Cash transfers (in 100 USD)	NA	1.006	NA	1.054	1.057
=1 if girls	1.000	1.000	0.448	0.471*	0.486**
Age	14.964	14.740	9.889	9.910	9.912
Obs.	1145	412	1276	3706	1740
Joint F-test		0.153		0.250	0.106

Note: Standard errors are clustered at randomization units. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. The cash transfer amount is taken from the second round of household survey. The stars indicate t-tests on whether the treatment and control groups are different on average. Joint F-test reports the p-values from F-tests on whether the variables in the table except for the cash transfer amount are balanced across the groups jointly.
*** p<0.01 ** p<0.05 * p<0.1

in the Moroccan experiment, their estimates were higher under LCTs than CCTs. Although I made different sample restrictions, the estimated ATEs with my sample are able to replicate these features of the original estimates of the ATEs. In the following analysis, I evaluate extrapolation results based on these estimated ATEs as well as the estimated enrollment rates of the control group in Table 2.

Table 2: Estimates of ATEs on enrollment rates

	Malawi	Morocco	
	(1) CCTs	(2) CCTs	(3) LCTs
ATE	0.0369*	0.0567***	0.0726***
	(0.0200)	(0.0106)	(0.0107)
Control mean	0.896***	0.894***	0.893***
	(0.0154)	(0.00951)	(0.00833)
Obs.	1490	4982	3018

Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights.
*** p<0.01 ** p<0.05 * p<0.1

4.2 Reduced-form approaches

I show the estimated coefficients of the two reduced-form approaches in Table 3. The first column is the estimates of the coefficients from the heterogeneous treatment effects using the Malawi CCTs' sample. The interaction terms are not statistically significantly different from 0 except for the one interacted with the pre-treatment school costs, which shows stronger treatment effects for children with higher school costs. The second column is the estimated coefficients from the logit in the propensity score weighting. The coefficients reflect differences in the baseline characteristics across the experiments. For example, since the Malawi CCTs' sample was about to enter secondary school, children with more years of education are more likely to be observed in the Malawi CCTs.

4.3 Structural approach

4.3.1 Parameter estimates

I show the estimates of θ and the empirical distributions of $\Delta R(e;x)$ under the Moroccan LCTs. To construct the dependent variable in Equation (2), I use the GMM estimates of the choice probabilities.[20] Table 4 shows the estimate of θ. The first-stage F-statistics for weak instruments is greater than the conventional threshold of 10, and the estimated value is statistically different from 0 at a significance level of 0.01. To interpret the size of the parameter, I compute the elasticity of schooling with respect to the cash transfer amount and find that the average elasticity is 0.142. That is, a 1% increase in cash transfer amount on average increases the probability of schooling by 0.142%.[21]

The distribution of the estimates of $\Delta R(e;x)$ is presented in Figure 2. The estimates are mostly positive, which means the returns to schooling relative to outside options are positive for the Moroccan LCTs' sample. The positive relative returns to schooling make sense. Given that the LCTs would not affect the flow utility differentially across the choices substantially, schooling decisions should be driven by the positive differential returns. The average size of the

[20]The full estimation results are available in Appendix.

[21]The formula to compute the elasticity is as follows:

$$\frac{\partial P^1_{i,2}}{\partial z_{i,2}}\frac{z_{i,2}}{P^1_{i,2}} = \begin{cases} \frac{\partial P^1_{i,2}}{\partial s_{i,2}}\frac{\partial s_{i,2}}{\partial z_{i,2}}\frac{z_{i,2}}{P^1_{i,2}} = P^1_{i,2}\theta\frac{z_{i,2}}{y_{i,2}-s_{i,2}} & \text{for CCTs} \\ \frac{\partial P^1_{i,2}}{\partial y_{i,2}}\frac{\partial y_{i,2}}{\partial z_{i,2}}\frac{z_{i,2}}{P^1_{i,2}} = P^1_{i,2}\theta\frac{z_{i,2}}{y_{i,2}-s_{i,2}}\frac{s_{i,2}}{y_{i,2}} & \text{for LCTs.} \end{cases}$$

Table 3: Estimates of coefficients in reduced-form approaches

	HTE	PSW
	(1)	(2)
	= 1 if enrollment (OLS)	= 1 if in Malawi CCTs (Logit)
Treatment	0.364	12.22*
	(0.253)	(7.256)
Years of education	0.0167	4.785***
	(0.0241)	(1.114)
Per-capita income	0.00630	-0.471**
	(0.0142)	(0.201)
School costs	-0.0156	-36.06***
	(0.132)	(12.72)
Cash transfers	0.0275	-11.35*
	(0.0377)	(5.818)
Age	-0.0580***	2.677***
	(0.00774)	(0.927)
Treatment × Years of education	-0.0373	
	(0.0270)	
Treatment × Per-capita income	-0.00847	
	(0.0144)	
Treatment × School costs	0.304*	
	(0.167)	
Treatment × Age	-0.00477	
	(0.0133)	
Constant	1.660***	-60.07***
	(0.182)	(17.14)
Obs.	1490	6472

Clustered standard errors (randomization units) in parentheses. HTE indicates the heterogeneous treatment effects approach while PSW does the propensity score weighting. Observations are weighted by sampling weights for the HTE only. The interaction between the treatment dummy and the cash transfer amount is omitted because of collinearity with the cash transfer amount itself.
*** p<0.01 ** p<0.05 * p<0.1

Table 4: Estimates of θ under Moroccan LCTs

	(1)
θ	38.90***
	(11.30)
Obs.	3016
1st stage F statistics	25.483
CCP estimation	GMM

Note: Clustered standard errors (randomization units) in parentheses. Observations are weighted by sampling weights. I report the Kleiberge-Paap F statistics for weak identification.
*** p<0.01 ** p<0.05 * p<0.1

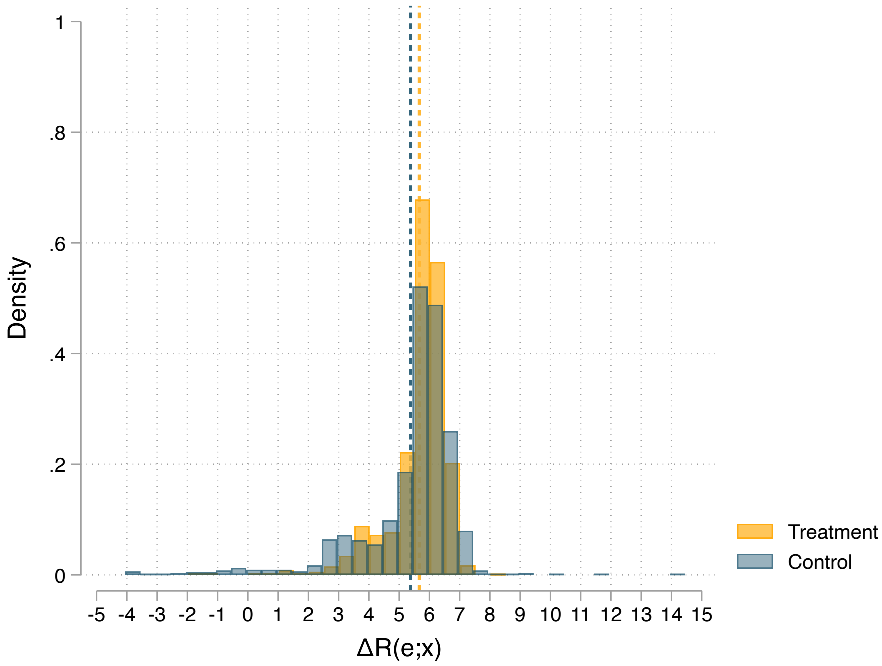
relative returns to schooling is slightly greater for the treatment group than the control one, which is counterintuitive. Theoretically, the relative returns to schooling should be larger for the control group children as their schooling decisions are entirely due to the relative returns to schooling. Thus, the fact that the relative returns are estimated greater for the treatment group indicates that the LCTs affected them directly. This can be interpreted as the signaling mechanism highlighted in Benhassine et al. (2015). That is, the LCTs increased enrollment partly because they were perceived as a signal that education was important for children's future.

To reinforce the above discussion, Table 5 compares the average size of differential utility and differential returns across the treatment status. The mean differential utility is negative for both groups because schooling provides no instantaneous returns. The differential utility is statistically significantly larger for the treatment group because of the cash transfers. As shown in Figure 2, the differential returns are also larger for the treatment group. Finally, the differential returns are greater in absolute terms than the differential utility as the former is the discounted sum of a series of the differential utility.

4.3.2 Model fit

With the estimated parameters above, I check the fitness of my model to the Moroccan LCTs' data. I evaluate the model fit based on how close the ATEs as well as enrollment rates for the control group simulated in my model are to the estimated values in column (2) of Table 2. Following the procedures in Section 3.2.4, I obtain $\hat{\delta}_1$ and $\hat{\delta}_2$ and run statistical tests separately:

Figure 2: Empirical distributions of $\Delta R(e;x)$ under Moroccan LCTs



Note: The dashed lines indicate $E[\Delta R_i(e)]$ for each group, weighted by sampling weights.

$$H_0^k : \delta_k = \hat{\alpha}_k \text{ for } k \in \{1,2\},$$

where $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are from the estimation of Equation (5').[22]

Table 6 presents the predicted values of the ATE of the Moroccan LCTs and the control group's enrollment rates. I predict that the Moroccan LCTs increase enrollment rates by 5.39 percentage points from a base of 90 percent. The results of the statistical tests show that for both objects, the null hypotheses about the equality between the predicted and estimated values are not rejected at a significance level of 0.05. These results support the good fitness of my model to the data in which it is estimated. It is worth noting that the good fitness of my model is ex-ante expected as I choose the estimation method to match these moments when estimating the choice probabilities.

[22] Another way of estimating the ATEs is to draw the preference shocks from the type-I extreme value distribution for all observations and use the simulated schooling decisions for the dependent variable in Equation (5). This approach is theoretically identical to mine, except that the standard errors would be larger. This is because the simulated decisions are binary while the choice probabilities are continuous between 0 and 1. This difference would affect the results of the statistical tests. I do not simulate the decisions because they may vary by the draws of the preference shocks.

Table 5: Estimated size of differential utility and returns under the Moroccan LCTs

	(1) Control	(2) Treatment
$E[\theta\Delta u]$	-1.706	-1.390***
$E[\beta\Delta R_i(e)]$	5.109	5.383***

Note: I use the second round of the household survey to create this table. Standard errors are clustered at randomization units. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. The stars indicate t-tests on whether the treatment and control groups are different on average.
*** p<0.01 ** p<0.05 * p<0.1

I also check how accurately my model can predict individual decisions. Since I choose the GMM estimator to estimate the choice probabilities, which does not maximize the likelihood of the data, it may be that the choice probabilities estimates are not as accurate at the individual level. To investigate this empirically, I simulate individual decisions by drawing preference shocks and compute the share of children with the simulated decisions matched with the actual decisions. The results in Table 7 show my model correctly simulates the choices for nearly 90% of the Moroccan LCTs' sample. I also check if my model is systematically more likely to overestimate $\left(d_{it}^{\text{Data}}=0, d_{it}^{\text{Model}}=1\right)$ or underestimate $\left(d_{it}^{\text{Data}}=1, d_{it}^{\text{Model}}=0\right)$ schooling decisions or not and find that it makes prediction errors in both directions at similar rates.

5 Extrapolation Results

5.1 Comparison of two extrapolations

5.1.1 Standard extrapolation methods

I first show the prediction results of the two extrapolations using the standard approaches. That is, I use the reduced-form approaches for the across-contexts extrapolation and the structural approach for the across-policies extrapolation. Table 8 present the predicted ATE of the Moroccan CCTs on enrollment rates as well as the control group's enrollment rates. The first two columns are the across-contexts extrapolation, and the next two columns are the across-policies extrapolation.

I find that the across-contexts extrapolation fails to predict the ATE. Column (1) is about

Table 6: Replication of ATE on enrollment rates for Moroccan LCTs

	(1)
ATE	0.0539***
	(0.00954)
Control mean	0.900***
	(0.00889)
Obs.	3018
Target ATE	0.073
= Target ATE	0.051
95% CI of ATE	[0.035, 0.073]
Target control mean	0.893
= Target control mean	0.476
95% CI of control mean	[0.882, 0.917]

Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. I report p-values from F-tests on the null hypothesis that the predicted value is equal to the estimated one, separately for the ATE and the control mean..
*** p<0.01 ** p<0.05 * p<0.1

Table 7: Model fit at individual level under Moroccan LCTs

	(1)
Correct	0.887
Overestimation	0.057
Underestimation	0.056

Note: Observations are weighted by sampling weights.

the heterogeneous treatment effects approach, which predicts that the Moroccan CCTs increase enrollment rates statistically significantly by 21.2 percentage points. In contrast, in column (2), the propensity score weighting predicts the null treatment effect. Both predictions have a 95% confidence interval of the predicted ATE that does not contain the estimated ATE. If anything, the propensity score weighting makes more reliable predictions as it predicts the control group's enrollment rates accurately: the null hypothesis on the equality between the predicted and estimated enrollment rates is not rejected at a significance level of 0.05.

In contrast, I find that the across-policies extrapolation successfully predicts the ATE. Column (3) and (4) differ in how I extrapolate $\Delta R(e;x)$ when computing the predicted choice probabilities. Regardless of that, the across-policies extrapolation predicts that the Moroccan CCTs increase enrollment rates by 5.77 or 5.90 percentage points, less than 0.5 percentage points away from the estimated value. The null hypotheses on the equality are also not rejected at a significance level of 0.05. However, the across-policies extrapolation overpredicts the control groups' enrollment rates. I will discuss this prediction bias when interpreting the differential extrapolation results with my structural model.[23]

Table 8: Prediction of ATE on enrollment rates
Across-contexts (reduced-form) vs Across-policies (structural)

Target: Morocco CCTs	Across-contexts		Across-policies	
	(1)	(2)	(3)	(4)
	HTE	PSW	Linear	RF
ATE	0.212***	0.00660	0.0590***	0.0577***
	(0.00442)	(0.0184)	(0.00545)	(0.00542)
Control mean	1.127***	0.895***	0.941***	0.942***
	(0.00373)	(0.0128)	(0.00531)	(0.00529)
Obs.	4982	1490	4982	4982
Target ATE	0.057	0.057	0.057	0.057
= Target ATE	0.000	0.007	0.674	0.863
95% CI of ATE	[0.203, 0.221]	[-0.030, 0.043]	[0.048, 0.070]	[0.047, 0.068]
Target control mean	0.894	0.894	0.894	0.894
= Target control mean	0.000	0.927	0.000	0.000
95% CI of control mean	[1.120, 1.134]	[0.869, 0.920]	[0.931, 0.952]	[0.932, 0.953]

Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. HTE indicates the heterogeneous treatment effects approach while PSW does the propensity score weighting. Linear indicates the linear extrapolation of $\Delta R(e;x)$ using the second order polynomials of years of education while RF does the Random Forest algorithm on years of education. I report p-values from F-tests on the null hypothesis that the predicted value is equal to the estimated one, separately for the ATE and the control mean.
*** p<0.01 ** p<0.05 * p<0.1

5.1.2 Structural approaches

One might concern that the results in Table 8 are driven by the differences in the extrapolation methods. That is, the structural approach might be better able to transfer the treatment effect estimates than the reduced-form approaches.[24] To isolate the differences in the external validity between the two interventions and seek an improvement in the across-contexts predictions, I do the across-contexts extrapolation using the structural approach. Since my model is causally identified under any types of cash transfers that shift the utility cost of schooling exogenously,

[23]In the Appendix, I show the extrapolation results using higher order polynomials of e for the linear projection. The results are robust to this modification.

[24]Fudenberg et al. (2022) define the predictive power of economic models conditional on unpredictable variation by any empirical methods. Using their terminology, my structural approach might be more complete than the reduced-form approaches.

it can be applied to the Malawi CCTs. Moreover, I can estimate the model under the Moroccan CCTs as well. As I will discuss in the following sections, the comparison of the estimated parameters in the identical model allows me to investigate whether θ or $\Delta R(e; x)$ or both is key for successful predictions.

The new extrapolation results are presented in Table 9. The first two columns are the across-contexts extrapolation results using the structural approach while the last two columns reproduce the across-policies extrapolation results reproduced from Table 8. The predicted ATE is at best 4.3 percentage points from a base of 67.6 percent. Compared to the results using the reduced-form approaches, the predicted ATE is closer to the estimated value, although the 95% confidence interval still does not contain it. In terms of the predictions of the levels of enrollment rates, the structural approach underpredicts the control group's enrollment rates by 20 percentage points. While this prediction is more sensible than the heterogeneous treatment effects approach, it is worse than the propensity score weighting. Thus, the overall performance of the across-contexts extrapolation is not necessarily improved by using the structural approach. However, in the subsequent analysis, I will use the structural approach for the across-contexts extrapolation because of its predictive power of the ATE.[25]

5.2 Interpretation of extrapolation results

5.2.1 Comparison of estimated models

So far, my results show that the across-policies extrapolation outperforms the across-contexts one and that the structural approach improves the prediction of the ATE of the across-contexts extrapolation. To understand what causes the prediction differences, I obtain the estimated models under all of the cash transfers studied in this paper. Then, I compare the extrapolated values of θ and $\Delta R(e; x)$ from each intervention with the estimated ones under the Moroccan CCTs. This analysis allows me to understand which parameter is key for accurate predictions.

[25] While the comparison of extrapolation methods for the across-contexts extrapolation is beyond the scope of this paper, one explanation for the improvement of the prediction of ATE via the across-contexts extrapolation with the structural method relative to the reduced-form ones is the normalization of variables within the interventions. That is, the structural method measures cash transfers and school costs relative to per-capita income while the reduced-form ones use the absolute values in prediction. Given that per-capita income and school costs are on average substantially higher for the Moroccan experiment, ignoring the level differences could harm the performance of the across-contexts extrapolation. In the Appendix, however, I show that the across-contexts extrapolation results do not improve when cash transfers and school costs are defined relative to per-capita income, but do improve when years of education and children's age are additionally standardized, which is consistent with my analysis in Section 5.2.

Table 9: Prediction of ATE on enrollment rates
Across-contexts (structural) vs Across-policies (structural)

Target: Morocco CCTs	Across-contexts		Across-policies	
	(1) Linear	(2) RF	(3) Linear	(4) RF
ATE	0.0431***	0.0412***	0.0590***	0.0577***
	(0.00465)	(0.00644)	(0.00545)	(0.00542)
Control mean	0.702***	0.676***	0.941***	0.942***
	(0.00390)	(0.00556)	(0.00531)	(0.00529)
Obs.	4982	4982	4982	4982
Target ATE	0.057	0.057	0.057	0.057
= Target ATE	0.004	0.016	0.674	0.863
95% CI of ATE	[0.034, 0.052]	[0.028, 0.054]	[0.048, 0.070]	[0.047, 0.068]
Target control mean	0.894	0.894	0.894	0.894
= Target control mean	0.000	0.000	0.000	0.000
95% CI of control mean	[0.694, 0.710]	[0.665, 0.687]	[0.931, 0.952]	[0.932, 0.953]

Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. Linear indicates the linear extrapolation of $\Delta R(e;x)$ using the second order polynomials of years of education while RF does the Random Forest algorithm on years of education. I report p-values from F-tests on the null hypothesis that the predicted value is equal to the estimated one, separately for the ATE and the control mean.
*** p<0.01 ** p<0.05 * p<0.1

Table 10 shows the estimates of θ under each of the cash transfers. First of all, all of the estimates are positive and statistically significantly different from 0, and the 1st stage F-statistics are greater than the conventional threshold. This proves that all of the interventions serve as a strong IV to identify θ. The size of θ varies substantially across the interventions, although it is more similar within the policies than within the contexts. This is because the size depends on the extent to which the cash transfers shift the relative utility of schooling, $\ln\left(\frac{y-s}{y}\right)$, which differs across the types of the cash transfers. Therefore, the estimated θ under the Moroccan CCTs is better approximated by the one under the Malawi CCTs than the Moroccan LCTs.

Figure 3 shows the overall empirical distributions of $\Delta R(e;x)$. Under all of the interventions, the relative returns to schooling are largely positive. Like the estimates of θ, the average size of the relative returns is more similar within the policies than within the contexts. This makes sense provided that the relative returns are identified as the remaining variation in the odds of schooling. That is, higher enrollment rates for the treatment group under the CCTs are explained by a higher relative utility of schooling rather than higher relative returns to schooling. In other words, since the LCTs do not increase the relative utility more than the CCTs do, an increase in enrollment rates under the LCTs should be driven more by the relative returns.

Table 10: Comparison of estimates of θ across interventions

	Malawi	Morocco	
	(1) CCTs	(2) CCTs	(3) LCTs
θ	1.008***	2.670***	38.90***
	(0.256)	(0.454)	(11.30)
Obs.	1479	4981	3016
1st stage F statistics	113.011	3843.510	25.483
CCP estimation	MLE	GMM	GMM
= different policy θ		0.000	
= different context θ		0.000	

Note: Clustered standard errors (randomization units) in parentheses. Observations are weighted by sampling weights. I report the Kleiberge-Paap F statistics for weak identification.
*** p<0.01 ** p<0.05 * p<0.1

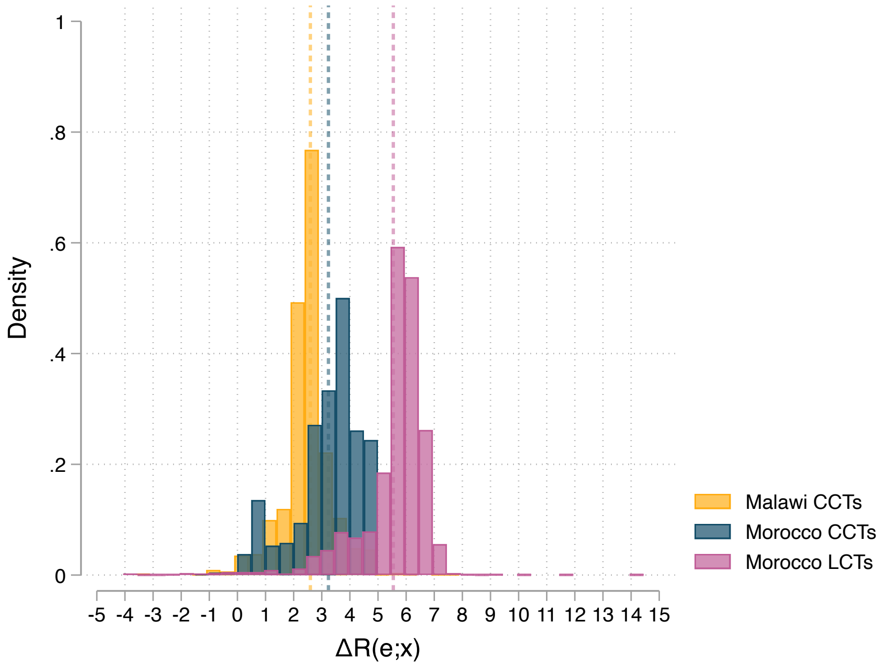
Although the average size of $\Delta R(e;x)$ is more similar within the policies, this does not imply that the estimated $\Delta R(e;x)$ under the Moroccan CCTs is better approximated by the one under the Malawi CCTs. This is confirmed by the extrapolation results that the across-policies extrapolation outperforms the across-contexts extrapolation. One explanation that reconciles these two facts is that the variation in $\Delta R(e;x)$ across the treatment status is more similar across the policies than across the contexts.

Table 11 shows the average relative utility and returns of schooling for the treatment and control groups separately for each intervention. The average relative utility of schooling in column (1) and (2) presents a larger variation between the groups under the CCTs than the LCTs, consistent with the reasoning behind the varying size of θ. On the other hand, the average relative returns to schooling in column (3) and (4) show similar patterns across the policies. While the average returns are estimated to be smaller for the treatment group under the Malawi CCTs, they are, if anything, greater for the treatment group under the two Moroccan cash transfers, which can be explained by the signaling effect that altered the participants' perception about education. This similarity in the across-groups variation of $\Delta R(e;x)$ is an advantage that the across-policies extrapolation has when predicting the ATE.

5.2.2 Varying predictions of enrollment rates

I discuss how the varying estimates of θ and $\Delta R(e;x)$ across the interventions translate into the varying performance of the two extrapolations. I first argue that the overestimation and

Figure 3: Comparison of empirical distributions of $\Delta R(e;x)$ across interventions



Note: The dashed lines indicate $E[\Delta R_i(e)]$ under each intervention, weighted by sampling weights.

underestimation of enrollment rates are partly explained by the differential sizes of the parameter estimates. Specifically, when extrapolating with a larger θ than the one under the Moroccan CCTs, I am likely to overpredict the relative utility of schooling for the treatment group while underpredicting it for the control group. This is because the treatment children tend to have positive values for $\ln\left(\frac{y-s}{y}\right)$ as the CCTs covered more than the school costs while the control group has negative values. Thus, by multiplying by a larger value, the relative utility for the treatment group becomes larger while that for the control group gets smaller. In contrast, when extrapolating with a distribution of $\Delta R(e;x)$ that is on average greater than the one under the Moroccan CCTs, I am likely to overpredict the relative returns for both groups as I tend to assign larger values for a given year of education than the estimated values. Because the probabilities of schooling, as in Equation (5), are increasing in the sum of the relative utility and the relative returns, the sizes of the parameter values can determine the direction of bias in predicting the treatment group's enrollment rates.

Table 12 summarizes the prediction bias for each extrapolation. My estimation results show the following relationships:

$$\hat{\theta}^{\text{Malawi CCTs}} < \hat{\theta}^{\text{Morocco CCTs}} < \hat{\theta}^{\text{Morocco LCTs}}$$

Table 11: Comparison of average relative utility and returns across interventions

	$E[\theta\Delta u]$		$E[\beta\Delta R_i(e)]$	
	(1) Control	(2) Treatment	(3) Control	(4) Treatment
Malawi CCTs	-0.292	0.594***	2.526	2.373*
Morocco CCTs	-0.117	0.409***	3.045	3.085
Morocco LCTs	-1.706	-1.390***	5.109	5.383***

Note: I use the second round of the household surveys to create this table. Standard errors are clustered at randomization units. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. The stars indicate t-tests on whether the treatment and control groups are different on average.
*** p<0.01 ** p<0.05 * p<0.1

and

$$E\left[\widehat{\Delta R}(e;x)^{\text{Malawi CCTs}}\right] < E\left[\widehat{\Delta R}(e;x)^{\text{Morocco CCTs}}\right] < E\left[\widehat{\Delta R}(e;x)^{\text{Morocco LCTs}}\right].$$

Therefore, the across-contexts extrapolation is subject to a downward bias while the across-policies extrapolation is to an upward bias when predicting the treatment group's enrollment rates. Consistent with this, in Table 9, I find that the across-contexts extrapolation underpredicts them by 20 percentage points while the across-policies extrapolation overpredicts by 6 percentage points.

Table 12: Direction of prediction bias for enrollment rates

	Across-contexts		Across-policies	
	Treatment	Control	Treatment	Control
$\hat{\theta}^{\text{Extrapolated}} \ln\left(1-\frac{s_{i,2}}{y_{i,2}}\right)$	$\Downarrow$	$\Uparrow$	$\Uparrow$	$\Downarrow$
$\widehat{\Delta R}^{\text{Extrapolated}}(e_{i,2};x_{i,2})$	$\Downarrow$	$\Downarrow$	$\Uparrow$	$\Uparrow$
$\hat{\theta}^{\text{Extrapolated}} \ln\left(1-\frac{s_{i,2}}{y_{i,2}}\right)+\beta\widehat{\Delta R}^{\text{Extrapolated}}(e_{i,2};x_{i,2})$	$\Downarrow$	?	$\Uparrow$	?

In contrast, the direction of the prediction bias for the control group's enrollment rates is ambiguous in both extrapolations. Therefore, the discussion based on the average size of the parameter values does not speak to the prediction bias for the ATE of the Moroccan CCTs. To understand that, in the next subsection, I examine more closely the extrapolated values of the parameters.

5.2.3 Varying predictions of ATE

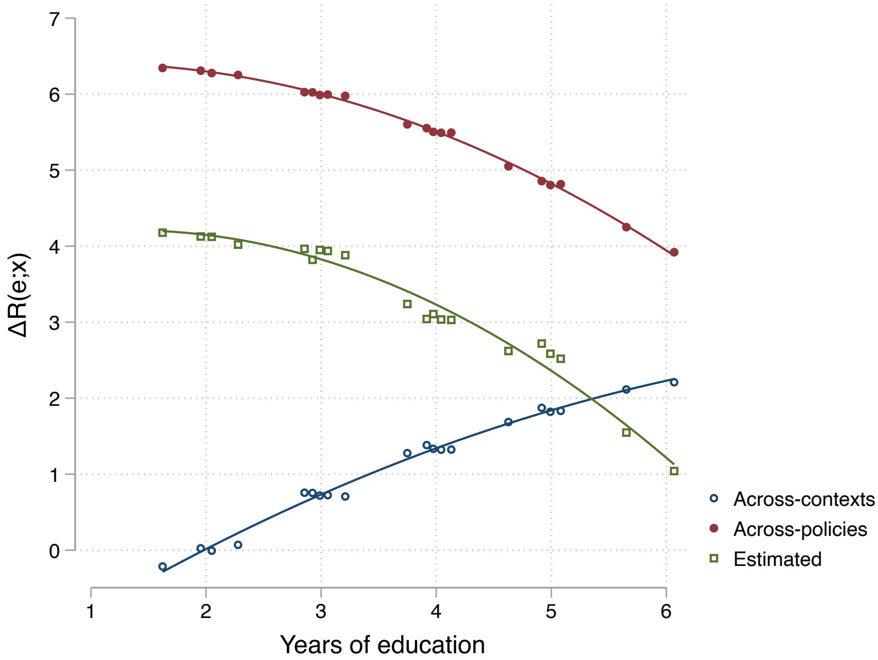
I proceed to explore why the across-policies extrapolation predicts the ATE of the Moroccan CCTs more accurately than the across-contexts extrapolation. First of all, given that $\hat{\theta}^{\text{Malawi CCTs}}$ is closer to $\hat{\theta}^{\text{Morocco CCTs}}$ than $\hat{\theta}^{\text{Morocco LCTs}}$ is, $\widehat{\Delta R}(e;x)$ extrapolated from the Moroccan LCTs should be closer to the estimated values than those extrapolated from the Malawi CCTs are. To visualize this point, I examine the correlation between the extrapolated $\widehat{\Delta R}(e;x)$ and the estimated $\widehat{\Delta R}(e;x)$. Because the performance of both extrapolations does not differ by how I extrapolate $\widehat{\Delta R}(e;x)$ as in Table 9, I show subsequent results using the linear projection unless indicated otherwise.

Figure 4 plots both the extrapolated and estimated values of $\widehat{\Delta R}(e;x)$ on years of education for the Moroccan CCTs' sample. The green square line represents the estimated values and is a downward sloping curve over years of education. The blue empty circle line represents $\widehat{\Delta R}(e;x)$ extrapolated from the Malawi CCTs. Unlike the estimated values, the line is an upward sloping curve, indicating an opposite correlation between $\widehat{\Delta R}(e;x)$ and years of education. In contrast, the red filled circle line for the extrapolated $\widehat{\Delta R}(e;x)$ from the Moroccan LCTs is parallel to the estimated values. Therefore, although the levels are different, the across-policies extrapolation predicts within-intervention variation of $\widehat{\Delta R}(e;x)$ more accurately than the across-contexts extrapolation.

One explanation for the opposite correlation pattern in $\widehat{\Delta R}(e;x)$ extrapolated across the contexts is differences in the education levels of the targeting populations across the experiments. As seen in Figure 4, the estimated values of $\widehat{\Delta R}(e;x)$ for the Moroccan CCTs' sample is decreasing in years of education over primary education, which can be explained by increasing opportunity costs of schooling. As children obtain more education or get older, they become more valuable in non-schooling activities. If returns to those activities increase more rapidly than returns to schooling, then the relative returns to schooling would be decreasing. The more beneficial outside options are, the more likely children choose to drop out.

In contrast, Figure 5 shows the estimated values of $\widehat{\Delta R}(e;x)$ for the Malawi CCTs' sample have two parts: increasing up to grade 8, which is the final grade in primary school, then decreasing. In other words, the relative returns are increasing over primary education and decreasing over secondary education for the Malawi CCTs' sample. The increasing relative returns mean that the returns to schooling grow faster than non-schooling, which suggests that outside options do not appear as attractive as schooling during that period. Consistent

Figure 4: Comparison of $\widehat{\Delta R}(e;x)$ for Moroccan CCTs' sample



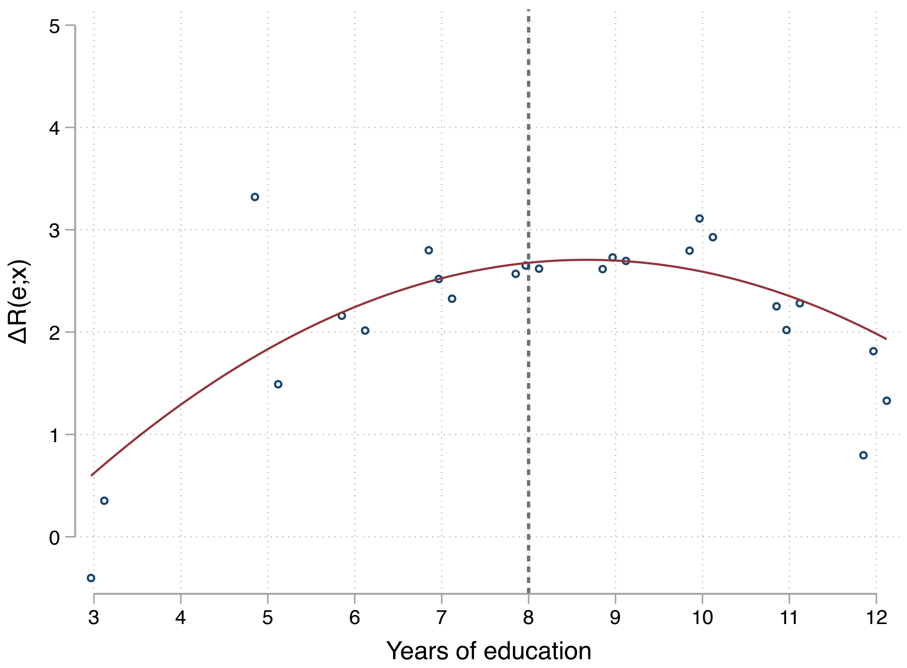
Note: The figure shows the bin scatter plots for the Moroccan CCTs' sample. I use residualized values by controlling for sampling strata fixed effects. Observations are weighted by sampling weights. I draw quadratic fit lines.

with that, the Malawi CCTs' sample largely completed primary education, as the experiment targeted girls who were at risk of dropping out of secondary school. Therefore, the tipping point about the relative returns corresponds to when the Malawi children started choosing dropout. When I extrapolate from the Malawi CCTs, I predict $\widehat{\Delta R}(e;x)$ based on the increasing part, leading to the upward sloping curve of the extrapolated $\widehat{\Delta R}(e;x)$.

Finally, to confirm that the poor performance of the across-contexts extrapolation stems from $\widehat{\Delta R}(e;x)$, I run the across-contexts extrapolation by replacing either $\hat{\theta}$ or $\widehat{\Delta R}(e;x)$ with the estimated values. Results in Table 13 show that the performance improves only when I replace $\widehat{\Delta R}(e;x)$. When replacing $\hat{\theta}$, column (2) and (3) show that the ATE is overpredicted while the control group's enrollment rates are underpredicted as much as when not replacing it. In contrast, when replacing $\widehat{\Delta R}(e;x)$, while the ATE is similarly underpredicted as before, the underprediction of the enrollment rates is corrected. As a result, the overall prediction accuracy is improved.[26]

[26] When I do the same exercise for the across-policies extrapolation, I find that the performance becomes worse off when replacing $\widehat{\Delta R}(e;x)$ than $\hat{\theta}$, which is coherent to the argument in the main text. However, I also find that the across-policies extrapolation underpredicts the ATE when replacing $\hat{\theta}$. This is because if I use the estimated values of θ, then the relative utility becomes smaller compared to the relative returns. As a result, the

Figure 5: $\widehat{\Delta R}(e;x)$ on years of education: Malawi CCTs' sample



Note: The figure shows the bin scatter plots for the Malawi CCTs sample. I use residualized values by controlling for sampling strata fixed effects. Observations are weighted by sampling weights. I draw a quadratic fit line. The dashed line indicates the final grade of primary education in Malawi.

5.3 How to improve across-context extrapolation

5.3.1 Normalization of years of education

I have shown that the across-policies extrapolation outperforms the other because the former predicts the distribution of $\widehat{\Delta R}(e;x)$ more accurately. I have also shown that the poorer prediction of that by the across-contexts extrapolation is due to the difference in the timing when children started choosing school dropout. A natural question is whether eliminating such a difference would lead to a better performance of the across-contexts extrapolation.

To answer this question, I redo the across-contexts extrapolation with normalizing years of education. In particular, I first recenter the years of education of the Malawi CCTs' sample by subtracting 7. This modification makes the Malawi children similar to the Moroccan ones in the sense that both could choose dropout from grade 1. I also drop the Malawi children in primary education levels when predicting $\widehat{\Delta R}(e;x)$. This sample restriction eliminates the upward sloping part of $\widehat{\Delta R}(e;x)$ in Figure 5 and thus makes the extrapolated relative returns

choice probabilities are largely dependent on the relative returns. Since the exponential of the relative returns makes the choice probabilities close to 1 for everyone, the predicted ATE becomes smaller than estimated. The estimation results are shown in Appendix.

Table 13: Across-contexts extrapolation with replacement of $\hat{\theta}$ or $\widehat{\Delta R}(e;x)$

	Across-contexts			
	Linear	Linear	RF	
	(1)	(2)	(3)	(4)
ATE	0.0431***	0.0953***	0.0909***	0.0367***
	(0.00465)	(0.00486)	(0.00655)	(0.00721)
Control mean	0.702***	0.688***	0.663***	0.901***
	(0.00390)	(0.00411)	(0.00570)	(0.00689)
Obs.	4982	4982	4982	4982
Replace $\hat{\theta}$		✓	✓	
Replace $\widehat{\Delta R}(e;x)$				✓
Target ATE	0.057	0.057	0.057	0.057
= Target ATE	0.004	0.000	0.000	0.006
95% CI of ATE	[0.034, 0.052]	[0.086, 0.105]	[0.078, 0.104]	[0.022, 0.051]
Target control mean	0.894	0.894	0.894	0.894
= Target control mean	0.000	0.000	0.000	0.276
95% CI of control mean	[0.694, 0.710]	[0.680, 0.696]	[0.651, 0.674]	[0.888, 0.915]

Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. Linear indicates the linear extrapolation of $\Delta R(e;x)$ using the second order polynomials of years of education while RF does the Random Forest algorithm on years of education. Replace $\hat{\theta}$ and Replace $\widehat{\Delta R}(e;x)$ mean I replace each parameter with the estimated value under the Moroccan CCTs. I report p-values from F-tests on the null hypothesis that the predicted value is equal to the estimated one, separately for the ATE and the control mean.
*** p<0.01 ** p<0.05 * p<0.1

look more comparable to the estimated values.

I present the results of the across-contexts extrapolation with the normalization in Table 14. First, I notice that the ATE is still underestimated in all results, even more than without the normalization. Second, in contrast, the estimated enrollment rates are closer to the estimated levels than without the normalization. Third, the sample restriction does not improve the performance additionally, probably because it removes only a small number of observations. Fourth, the results are robust to how I extrapolate $\widehat{\Delta R}(e;x)$. The normalization improves the overall performance of the across-contexts extrapolation while it still underestimates the ATE of the Moroccan CCTs.

5.3.2 Heterogeneity across age and sex

It is plausible that the across-contexts extrapolation can make more accurate predictions for a subset of the Moroccan CCTs' sample. Because the Malawi CCTs' sample differs from

Table 14: Across-contexts extrapolation with normalized years of education

	Across-contexts			
	Linear		RF	
	(1)	(2)	(3)	(4)
ATE	0.0150***	0.0147***	0.0188***	0.0189***
	(0.000788)	(0.00265)	(0.00215)	(0.00227)
Control mean	0.903***	0.875***	0.870***	0.868***
	(0.000674)	(0.00221)	(0.00186)	(0.00195)
Obs.	4982	4982	4982	4982
Normalization	✓	✓	✓	✓
Sample restriction		✓		✓
Target ATE	0.057	0.057	0.057	0.057
= Target ATE	0.000	0.000	0.000	0.000
95% CI of ATE	[0.013, 0.017]	[0.010, 0.020]	[0.015, 0.023]	[0.014, 0.023]
Target control mean	0.894	0.894	0.894	0.894
= Target control mean	0.000	0.000	0.000	0.000
95% CI of control mean	[0.901, 0.904]	[0.871, 0.880]	[0.866, 0.873]	[0.864, 0.871]

Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. Linear indicates the linear extrapolation of $\Delta R(e;x)$ using the second order polynomials of years of education while RF does the Random Forest algorithm on years of education. Normalization means the normalized years of education are used and Sample restriction does the elimination of the Malawi sample with less than 8 years of schooling when doing extrapolation. I report p-values from F-tests on the null hypothesis that the predicted value is equal to the estimated one, separately for the ATE and the control mean.
*** p<0.01 ** p<0.05 * p<0.1

Moroccan CCTs' one in children's age and sex, adjusting for these differences makes the samples observationally more alike. As a result, it might be easier to extrapolate the distribution of the relative returns to schooling across the contexts. This is more likely the case if boys and girls have differential returns to education in Morocco. Benhassine et al. (2015) provide suggestive evidence that parental beliefs over children's returns to education in labor markets varied across children's sex. Hence, I examine heterogeneity in the across-contexts extrapolation results across children's age and sex.

I show in Table 15 the performance of the across-contexts extrapolation separately for boys and girls in the Moroccan CCTs' sample. In all of the results, I use the normalized years of education as it improves the overall prediction accuracy substantially. I find no differential performance across children's sex. For girls, I also look at the predictions separately for older cohorts. Column (3) and (6) show that as the ATE of the Moroccan CCTs for the older girls is larger than for girls of all ages, the predicted ATE is also larger for that age group. However, I still underpredict the ATE (for example, in column (3), 2.2 percentage points compared to 13.9 percentage points). Thus, the across-contexts extrapolation does not improve predictions by focusing on the Moroccan CCTs' children of the same sex in the same age cohorts.

Table 15: Across-contexts extrapolation across age and sex of Moroccan CCTs' sample

	Across-contexts					
	Linear			RF		
	Boys	Girls		Boys	Girls	
	(1)	(2)	(3) Age $\geq$ 12	(4)	(5)	(6) Age $\geq$ 12
ATE	0.0151***	0.0148***	0.0221***	0.0206***	0.0165***	0.0346***
	(0.00109)	(0.00129)	(0.00258)	(0.00296)	(0.00343)	(0.00623)
Control mean	0.902***	0.903***	0.880***	0.867***	0.872***	0.809***
	(0.000913)	(0.00112)	(0.00222)	(0.00250)	(0.00297)	(0.00552)
Obs.	2666	2313	858	2666	2313	858
Normalization	✓	✓	✓	✓	✓	✓
Target ATE	0.048	0.068	0.139	0.048	0.068	0.139
= Target ATE	0.000	0.000	0.000	0.000	0.000	0.000
95% CI of ATE	[0.013, 0.017]	[0.012, 0.017]	[0.017, 0.027]	[0.015, 0.026]	[0.010, 0.023]	[0.022, 0.047]
Target control mean	0.912	0.871	0.709	0.912	0.871	0.709
= Target control mean	0.000	0.000	0.000	0.000	0.839	0.000
95% CI of control mean	[0.900, 0.904]	[0.901, 0.905]	[0.876, 0.884]	[0.863, 0.872]	[0.866, 0.878]	[0.798, 0.820]

Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. Linear indicates the linear extrapolation of $\Delta R(e;x)$ using the second order polynomials of years of education while RF does the Random Forest algorithm on years of education. Normalization means the normalized years of education are used. I report p-values from F-tests on the null hypothesis that the predicted value is equal to the estimated one, separately for the ATE and the control mean..
*** p<0.01 ** p<0.05 * p<0.1

I also look at the heterogeneous performance of the across-contexts extrapolation across age cohorts of the Malawi CCTs' sample. In particular, when extrapolating $\widehat{\Delta R}(e;x)$, I use children of age from 13 to 16, the highest age of the Moroccan CCTs' sample.[27] Table 16 shows the results separately for boys and girls in the Moroccan CCTs' sample. I find that the sample restriction on the Malawi CCTs' sample does not improve the predictions of the across-contexts extrapolation either. Thus, while adjusting for differences in years of education narrows the performance gap between the two extrapolations, the across-contexts extrapolation does not improve additionally by restricting the samples in terms of children's age and sex.

[27] To do this analysis, I need to re-estimate the model with children's age as a new state variable to make $\Delta R(e;x)$ vary by age. The new parameter estimates are available in Appendix.

Table 16: Across-contexts extrapolation using young cohorts of Malawi CCTs' sample

	Across-contexts			
	Linear		RF	
	(1) Boys	(2) Girls	(3) Boys	(4) Girls
ATE	0.00376***	0.00402***	0.00664***	0.00658***
	(0.000254)	(0.000318)	(0.000113)	(0.000134)
Control mean	0.980***	0.980***	0.963***	0.963***
	(0.000211)	(0.000270)	(0.0000839)	(0.000108)
Obs.	2666	2313	2666	2313
Normalization	✓	✓	✓	✓
Target ATE	0.048	0.068	0.048	0.068
= Target ATE	0.000	0.000	0.000	0.000
95% CI of ATE	[0.003, 0.004]	[0.003, 0.005]	[0.006, 0.007]	[0.006, 0.007]
Target control mean	0.912	0.871	0.912	0.871
= Target control mean	0.000	0.000	0.000	0.000
95% CI of control mean	[0.980, 0.980]	[0.979, 0.980]	[0.963, 0.963]	[0.963, 0.963]

Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. Linear indicates the linear extrapolation of $\Delta R(e;x)$ using the second order polynomials of years of education while RF does the Random Forest algorithm on years of education. Normalization means the normalized years of education are used. I report p-values from F-tests on the null hypothesis that the predicted value is equal to the estimated one, separately for the ATE and the control mean.
*** p<0.01 ** p<0.05 * p<0.1

5.4 Why $\widehat{\Delta R}(e;x)$ varies similarly across policies?

5.4.1 Signaling effects

As I have discussed, one channel through which CCTs and LCTs in the Moroccan experiment improved school enrollment is signaling effects that parents perceived the cash transfer program supported by the Ministry of Education as the importance of education for children. While I do not model this effect explicitly, it is reflected in $\Delta R(e;x)$ as residual variation in schooling decisions unexplained by the contemporaneous effects of the cash transfers on the utility cost of schooling (see Section 5.2.1 for more detailed explanations). This additional effect of the cash transfers, or lack thereof, is not transferable across the contexts and thus the across-contexts extrapolation does not predict $\widehat{\Delta R}(e;x)$ accurately as the across-policies extrapolation.

5.4.2 Confusion about conditionality

Another reason why the CCTs and LCTs in the Moroccan experiment have comparable effect sizes is that the parents of the CCTs' treated children misunderstood that the cash transfers were not tied to school attendance. Table 17 shows that 11.1% of them in my data correctly understood the conditionality attached to the CCTs. Moreover, only 14.4% of them thought that they could receive the transfers by enrolling children in school. Therefore, a large fraction of the Moroccan CCTs' sample treated the CCTs as unconditional, which made it easy to extrapolate from the LCTs. In addition, although my data do not have information about how the Malawi CCTs' sample perceived the conditionality, the misunderstanding seems less prevalent in the Malawi experiment as Baird et al. (2011) show that the UCTs implemented along with the CCTs did not increase school enrollment.

Table 17: Knowledge about conditionality in Moroccan experiment

	(1) CCTs	(2) LCTs
Know program	0.999	1.000
Think transfers are conditional on enrollment	0.144	0.121
Think transfers are conditional on 5 absences	0.111	0.085

Therefore, I examine whether the across-contexts extrapolation can make more accurate predictions if the Moroccan CCTs' sample understood the conditionality correctly. To do this analysis, I compute the ATE of the Moroccan CCTs under the perfect understanding. Specifically, I first estimate my model under the Moroccan CCTs incorporating the knowledge about the conditionality: if a respondent answered that the cash transfers were conditional on school enrollment or regular attendance, they are subtracted from school costs, and if not, they are added to per-capita income. After estimating the model, I simulate schooling decisions assuming that everyone understands the conditionality so that the cash transfers are a subsidy to school costs, and estimate the ATE.[28] If the confusion explained the success of the across-policies extrapolation, the new ATE should be more accurately predicted by the across-contexts extrapolation.

Column (2) in Table 18 shows the counterfactual ATE of the Moroccan CCTs. If the conditionality is perfectly understood, then the Moroccan CCTs would increase enrollment rates

[28] To obtain counterfactual relative returns to schooling, I train the Random Forest algorithm on the state variables with the confusion and generate predicted values based on the same set of variables without the confusion.

Table 18: Prediction of ATE on enrollment rates under perfect understanding
Across-contexts (structural) vs Across-policies (structural)

	Estimation		Across-contexts	Across-policies
	(1)	(2)	(3)	(4)
	Original	Counterfactual	Linear	Linear
ATE	0.0567***	0.122***	0.0431***	0.0590***
	(0.0106)	(0.00836)	(0.00465)	(0.00545)
Control mean	0.894***	0.868***	0.702***	0.941***
	(0.00951)	(0.00814)	(0.00390)	(0.00531)
Obs.	4982	4982	4982	4982
Target ATE			0.122	0.122
= Target ATE			0.000	0.000
95% CI of ATE			[0.034, 0.052]	[0.048, 0.070]
Target control mean			0.868	0.868
= Target control mean			0.000	0.000
95% CI of control mean			[0.694, 0.710]	[0.931, 0.952]

Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. Original indicates the original estimates of the ATE and the control group's enrollment rates using my data while Counterfactual does the simulated ones when I assume that the conditionality is perfectly understood. Linear indicates the linear extrapolation of $\Delta R(e;x)$ using the second order polynomials of years of education while RF does the Random Forest algorithm on years of education. I report p-values from F-tests on the null hypothesis that the predicted value is equal to the estimated one, separately for the ATE and the control mean.
*** p<0.01 ** p<0.05 * p<0.1

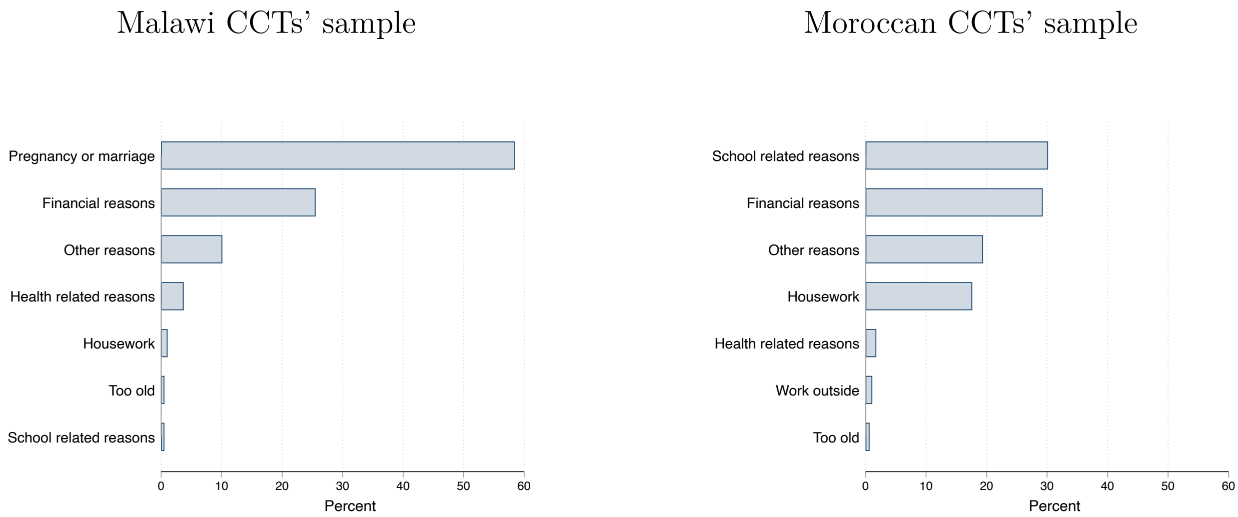
by 12.2 percentage points from a base of 86.8 percentage points, which is more than twice as large as the originally estimated effect size in column (1). Moreover, it is greater than the ATE of the Moroccan LCTs by 5 percentage points. Column (3) and (4) show that neither of the extrapolations predicts the counterfactual ATE accurately. However, the predicted values of the across-policies extrapolation are numerically closer to that as well as the control group's enrollment rates. Therefore, removing the degree of confusion about the conditionality among the Moroccan CCTs' sample does not make the across-contexts extrapolation perform better. In other words, ignoring the confusion levels does not explain the poorer performance of the across-contexts extrapolation.

5.4.3 Outside options

One explanation for the remaining performance gap between the two extrapolations is that outside options are meaningfully different across the contexts. Because $\Delta R(e;x)$ is measured relative to outside options, differences in outside options would affect the shape of $\Delta R(e;x)$.

More specifically, the values of outside options are reflected in the amount of cash transfers and thus affects $\Delta R(e;x)$ through the relative utility of schooling. While I have limited information about children's trajectories after the experiments, I can provide suggestive evidence on varying outside options by looking at primary reasons for school dropout in each context. Figure 6 shows that nearly 60% of the Malawi CCTs' sample who dropped out mentioned pregnancy or marriage as the main reason, both of which often stemmed from financial hardship. On the other hand, the major reasons for school dropout among the Moroccan CCTs' sample are school quality and liquidity constraints. In addition, around 20% of them reported housework as a reason to drop out, which less than 5% of the Malawi CCTs' dropout did. Therefore, what children choose instead of schooling varies across the contexts, which would affect the returns to non-schooling differentially. Consequently, it is more difficult to extrapolate $\widehat{\Delta R}(e;x)$ across the contexts than the policies.

Figure 6: Primary reasons for school dropout



Note: I use the samples in the first and second rounds of household surveys to create these graphs. Financial reasons are no money for school-related fees. Health related reasons are such as illnesses and disabilities. School related reasons include poor school infrastructure, poor quality of teaching, and bad access to school. The total of the bars in the graphs may not be 100 due to observations with the missing primary reasons.

6 Conclusion

Predicting the effects of a CCT program on educational outcomes is important for policymakers with limited program budgets. Provided that the program may put strain on administrative capacity because of continuous monitoring of school attendance, the quantification helps policymakers understand whether the program is worth implementation. Existing studies propose various approaches that enable us to transfer treatment effect estimates of CCTs across contexts. The key assumption they make is that conditional on observables, potential outcomes are independent of context characteristics. It would be, however, challenging to satisfy this assumption if we have to rely on CCT programs implemented in other contexts that are not necessarily comparable. In that case, an alternative way of predicting effects is by extrapolating from an adjacent policy in the same context that resembles how CCTs affect educational outcomes. Little is known, however, about in what ways the two extrapolations differ in making predictions.

This paper empirically studies this question using cash transfer experiments in Malawi and Morocco. My analysis shows that the across-policies extrapolation dominates the across-contexts extrapolation in predicting the ATE on enrollment rates. Through the lens of the structural model, I find that the driver of the differential predictions is the returns to schooling measured against outside options. I also find that by controlling for children's levels of education, sex, and age, the prediction gap is moderately narrowed. Finally, I suggest that the remaining gap should be explained by outside options, which seemingly vary substantially across the contexts.

While this paper investigates the two extrapolations in the case of CCTs, the underlying problem that contexts may be too different to make reliable predictions is observed in other settings (for example, heterogeneous treatment effects of microcredit studied by Meager 2019, 2022). Compared to extrapolation across contexts, however, the predictive power of extrapolation across policies is understudied in previous literature. Since this paper is the first to conduct a comparative analysis of the two extrapolations, the generalizability of my findings is an interesting avenue for future research.[29]

Another research avenue is how to aggregate evidence from different policies to make predictions. I show that the utility costs of schooling vary similarly across the contexts while the relative returns to schooling do across the policies. Thus, it is natural to think that a better prediction can be made by borrowing the best parameters from each intervention. This idea, however, may not immediately work. In the Appendix, I find no improvement of the across-policies extrapolation by using either the benchmark $\hat{\theta}$ or $\widehat{\Delta R}(e;x)$ because it distorts the ratio of the relative utility of schooling to the relative returns to schooling. Therefore, it is necessary to maintain the size balance of parameters when borrowing them from multiple policies in order to make accurate predictions. How to do that is beyond the scope of this paper

[29]Theoretical work on the predictive power of extrapolation across contexts is emerging. For example, Andrews et al. (2022) measure the ability of economic models to transfer evidence across contexts.

and hence requires future work.

Finally, this paper has implications for policy designs. As Hendren and Sprung-Keyser (2020) suggest, policymakers are willing to pay for precise estimates of policy effects.[30] Globally, this suggests potential benefits of coordination in policy designs across countries so that extrapolation across contexts makes accurate predictions. Locally, on the other hand, this suggests that a welfare-maximizing policymaker wants to design a policy so that it is informative about future policies. My comparative analysis of the two extrapolations speaks to both points empirically in the case of predicting the treatment effects of CCTs.

[30]Relatedly, Hjort et al. (2021) show that Brazilian mayors are willing to pay for research findings about policy evaluations.

References

Aguirregabiria, Victor, and Pedro Mira. 2002. "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models." *Econometrica* 70 (4): 1519–1543.

Allcott, Hunt. 2015. "Site Selection Bias in Program Evaluation." *The Quarterly Journal of Economics* 130 (3): 1117–1165. 10.1093/qje/qjv015.

Andrews, Isaiah, Drew Fudenberg, Annie Liang, and Chaofeng Wu. 2022. "The Transfer Performance of Economic Models." *Working paper.* 10.2139/ssrn.4175591.

Andrews, Isaiah, and Emily Oster. 2019. "A Simple Approximation for Evaluating External Validity Bias." *Economics Letters* 178 58–62. 10.1016/j.econlet.2019.02.020.

Arcidiacono, Peter, and Paul B. Ellickson. 2011. "Practical Methods for Estimation of Dynamic Discrete Choice Models." *Annual Review of Economics* 3 (1): 363–394. 10.1146/annurev-economics-111809-125038.

Arcidiacono, Peter, and Robert A. Miller. 2011. "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity." *Econometrica* 79 (6): 1823–1867. 10.3982/ECTA7743.

Attanasio, Orazio P., Costas Meghir, and Ana Santiago. 2012. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA." *The Review of Economic Studies* 79 (1): 37–66. 10.1093/restud/rdr015.

Baird, Sarah, Francisco HG Ferreira, Berk Özler, and Michael Woolcock. 2014. "Conditional, Unconditional and Everything in between: A Systematic Review of the Effects of Cash Transfer Programmes on Schooling Outcomes." *Journal of Development Effectiveness* 6 (1): 1–43.

Baird, Sarah, Craig McIntosh, and Berk Özler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *The Quarterly journal of economics* 126 (4): 1709–1753.

Baird, Sarah, Craig McIntosh, and Berk Özler. 2012. "Schooling, Income, and Health Risk Impact Evaluation Household Survey 2007-2008, Round I (Baseline)." May. 10.48529/XP7Y-3K93.

Bandiera, Oriana. 2021. “Do Women Respond Less to Performance Pay? Building Evidence from Multiple Experiments.” 3 (4): 20.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2017. “From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application.” *Journal of Economic Perspectives* 31 (4): 73–102. 10.1257/jep.31.4.73.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen. 2015. “Turning a Shove into a Nudge? A “Labeled Cash Transfer” for Education.” *American Economic Journal: Economic Policy* 7 (3): 86–125. 10.1257/pol.20130225.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen. 2019. “Replication Data for: Turning a Shove into a Nudge? A "Labeled Cash Transfer" for Education.” October. 10.3886/E114579V1.

De Groote, Olivier, and Frank Verboven. 2019. “Subsidies and Time Discounting in New Technology Adoption: Evidence from Solar Photovoltaic Systems.” *American Economic Review* 109 (6): 2137–2172. 10.1257/aer.20161343.

Dehejia, Rajeev, Cristian Pop-Eleches, and Cyrus Samii. 2021. “From Local to Global: External Validity in a Fertility Natural Experiment.” *Journal of Business & Economic Statistics* 39 (1): 217–243. 10.1080/07350015.2019.1639407.

DellaVigna, Stefano, and Elizabeth Linos. 2022. “RCTs to Scale: Comprehensive Evidence From Two Nudge Units.” *Econometrica* 90 (1): 81–116. 10.3982/ECTA18709.

Diamond, Rebecca, Tim McQuade, and Franklin Qian. 2019. “The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco.” *American Economic Review* 109 (9): 3365–3394. 10.1257/aer.20181289.

Duflo, Esther, Rema Hanna, and Stephen P Ryan. 2012. “Incentives Work: Getting Teachers to Come to School.” *American Economic Review* 102 (4): 1241–1278. 10.1257/aer.102.4.1241.

Fiszbein, Ariel, and Norbert R. Schady. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty.* World Bank Publications.

Foster, Andrew D, and Esther Gehrke. 2017. “Start What You Finish! Ex Ante Risk and Schooling Investments in the Presence of Dynamic Complementarities.” Working Paper 24041, National Bureau of Economic Research. 10.3386/w24041.

Fudenberg, Drew, Jon Kleinberg, Annie Liang, and Sendhil Mullainathan. 2022. “Measuring the Completeness of Economic Models.” *Journal of Political Economy* 130 (4): 956–990.

Gechter, Michael. 2022. “Generalizing the Results from Social Experiments: Theory and Evidence from Mexico and India.” *manuscript, Pennsylvania State University.*

Gechter, Michael, Cyrus Samii, Rajeev Dehejia, and Cristian Pop-Eleches. 2018. “Evaluating Ex Ante Counterfactual Predictions Using Ex Post Causal Inference.” *arXiv preprint arXiv:1806.07016.*

Hainmueller, Jens. 2012. “Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies.” *Political Analysis* 20 (1): 25–46. 10.1093/pan/mpr025.

Hendren, Nathaniel, and Ben Sprung-Keyser. 2020. “A Unified Welfare Analysis of Government Policies*.” *The Quarterly Journal of Economics* 135 (3): 1209–1318. 10.1093/qje/qjaa006.

Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini. 2021. “How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities.” *American Economic Review* 111 (5): 1442–1480. 10.1257/aer.20190830.

Hotz, V. J., R. A. Miller, S. Sanders, and J. Smith. 1994. “A Simulation Estimator for Dynamic Models of Discrete Choice.” *The Review of Economic Studies* 61 (2): 265–289. 10.2307/2297981.

Hotz, V. Joseph, Guido W. Imbens, and Julie H. Mortimer. 2005. “Predicting the Efficacy of Future Training Programs Using Past Experiences at Other Locations.” *Journal of Econometrics* 125 (1-2): 241–270. 10.1016/j.jeconom.2004.04.009.

Hotz, V. Joseph, and Robert A. Miller. 1993. “Conditional Choice Probabilities and the Estimation of Dynamic Models.” *The Review of Economic Studies* 60 (3): 497–529. 10.2307/2298122.

Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling *." *Quarterly Journal of Economics* 125 (2): 515–548. 10.1162/qjec.2010.125.2.515.

Kalouptsidi, Myrto, Paul T. Scott, and Eduardo Souza-Rodrigues. 2021. "Linear IV Regression Estimators for Structural Dynamic Discrete Choice Models." *Journal of Econometrics* 222 (1): 778–804. 10.1016/j.jeconom.2020.03.016.

Magnac, Thierry, and David Thesmar. 2002. "Identifying Dynamic Discrete Decision Processes." *Econometrica* 70 (2): 801–816. 10.1111/1468-0262.00306.

Meager, Rachael. 2019. "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments." *American Economic Journal: Applied Economics* 11 (1): 57–91. 10.1257/app.20170299.

Meager, Rachael. 2022. "Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature." *American Economic Review* 112 (6): 1818–1847. 10.1257/aer.20181811.

Ozler, Berk, Sarah Baird, Craig McIntosh, and Ephraim Chirwa. 2015a. "Schooling, Income, and Health Risk Impact Evaluation Household Survey 2008-2009, Round 2 (Midline)." August. 10.48529/P47C-7345.

Ozler, Berk, Sarah Baird, Craig McIntosh, and Ephraim Chirwa. 2015b. "Schooling, Income, and Health Risk Impact Evaluation Household Survey 2010, Round 3 (Midline)." August. 10.48529/7W21-DJ26.

Pritchett, Lant, and Justin Sandefur. 2015. "Learning from Experiments When Context Matters." *American Economic Review* 105 (5): 471–475. 10.1257/aer.p20151016.

Rosenzweig, Mark R, and Christopher Udry. 2020. "External Validity in a Stochastic World: Evidence from Low-Income Countries." *The Review of Economic Studies* 87 (1): 343–381. 10.1093/restud/rdz021.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55 (5): 999. 10.2307/1911259.

Rust, John. 1994. "Chapter 51 Structural Estimation of Markov Decision Processes." In *Handbook of Econometrics*, Volume 4. 3081–3143, Elsevier, . 10.1016/S1573-4412(05)80020-0.

Scott, Paul. 2014. "Dynamic Discrete Choice Estimation of Agricultural Land Use."

Shah, Manisha, and Bryce Millett Steinberg. 2017. "Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital." *Journal of Political Economy* 125 (2): 527–561. 10.1086/690828.

Stuart, Elizabeth A., Stephen R. Cole, Catherine P. Bradshaw, and Philip J. Leaf. 2011. "The Use of Propensity Scores to Assess the Generalizability of Results from Randomized Trials: Use of Propensity Scores to Assess Generalizability." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 174 (2): 369–386. 10.1111/j.1467-985X.2010.00673.x.

Todd, Petra E, and Kenneth I 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* 96 (5): 1384–1417. 10.1257/aer.96.5.1384.

Traiberman, Sharon. 2019. "Occupations and Import Competition: Evidence from Denmark." *American Economic Review* 109 (12): 4260–4301. 10.1257/aer.20161925.

Vivalt, Eva. 2020. "How Much Can We Generalize From Impact Evaluations?" *Journal of the European Economic Association* 18 (6): 3045–3089. 10.1093/jeea/jvaa019.

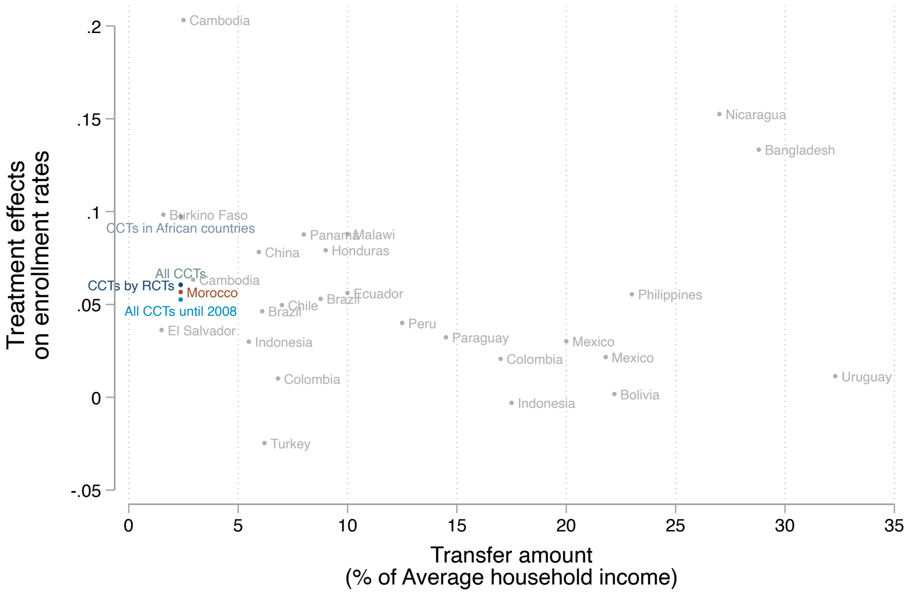
Vivalt, Eva, Aidan Coville, and K. C. Sampada. 2022. "Weighing the Evidence: Which Studies Count?".

Appendix

Details about predictions based on treatment effect estimates

The simplest approach to predict the treatment effect of CCTs in a new context is to use a collection of treatment effect estimates of CCTs in other settings. While the approaches examined in the main text are relevant to researchers, this approach may be more realistic for policymakers. Using the information summarized in Baird et al. (2014), I demonstrate the performance of this approach.

Reported treatment effect estimates in Baird et al. (2014)



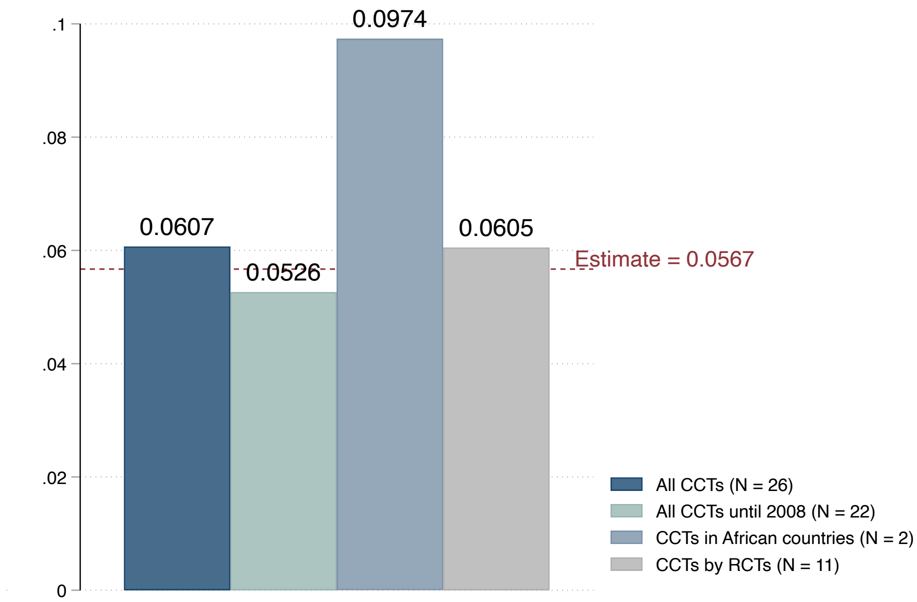
Specifically, the predictions are based on a linear relationships between ATEs and transfer amount. To do this, I first collect from the article the following information about CCT programs; where and when they were implemented, whether they were RCTs or not, ATEs on the odds of schooling, control group's baseline enrollment rates, and transfer amount as a share of average household income. Then, I recover ATEs on enrollment rates for each program using those on the odds of schooling and the control group's enrollment rates:

$$\text{ATE}^{\text{enrollment}} = \frac{\text{ATE}^{\text{odds}} \times \text{Odds}^{\text{Control}}}{1+\text{ATE}^{\text{odds}} \times \text{Odds}^{\text{Control}}} - \text{Enrollment}^{\text{Control}}.$$

Finally, I regress ATEs on enrollment rates on transfer amount separately for each sample and predict based on the average ratio of the transfer amount to household income for the Moroccan CCTs in my data, which is 2.3%.

I make 4 predictions using different sets of CCTs: (1) using all of the CCTs, (2) using the CCTs up to 2008 (when the Moroccan CCTs were implemented), (3) using the CCTs in African countries, and (4) using the CCTs implemented in the form of RCTs.[31] The bar chart shows the prediction results. While the predictions are numerically close to the estimated ATE, they vary across the input CCTs. To understand what causes the variation and potentially improve the predictions, one can focus on a subset of the CCTs that offers more granular data, which motivates my analysis in the main text.

Simple predictions of ATE of Moroccan CCTs on enrollment rates



[31]The third and fourth predictions are based on findings in Vivalt et al. (2022), which show that policymakers appreciate evidence from similar contexts while researchers do evidence with internally valid evaluation methods.

Details about choice probabilities estimation

In the first step of my estimation procedures, I estimate the choice probabilities with the flexible logit of the state variables either via MLE or GMM. The choice of the estimation method depends on whether they can replicate the shares of children choosing schooling for the treatment and control group in each survey round. Both estimations empirically yield similar estimates of the parameters of the logit.

Why the GMM estimates sometimes can replicate the ATEs that the MLE ones cannot is because of a trade-off of whether to prioritize the individual choice probabilities or the aggregate shares. The GMM estimator targets the aggregate shares directly while the MLE maximizes the likelihood of individual choices. This trade-off is salient when I estimate the choice probabilities under unconditional cash transfers because relative changes in the income by the cash transfers are small. Therefore, the MLE estimator may not be able to sufficiently differentiate the choice probabilities across the treatment status, leading to imprecise estimates of the ATEs at this stage. This can be fixed by the GMM estimator, although it does not necessarily maximize the likelihood of the data. I use the GMM estimates in some cases to replicate the ATEs through my model. In particular, I choose the GMM estimator for both of the Moroccan interventions. The replication of the ATE under the Moroccan CCTs is only slightly better with the GMM estimates than the MLE ones. On the other hand, the GMM estimator benefits largely the replication under the Moroccan LCTs.

The flexible logit consists of the second-order polynomials of (e_{it}, y_{it}, s_{it}) fully interacted with survey round dummy variables. The number of parameters to estimate is 30 for the Malawi experiments and 20 for the Moroccan ones. When I estimate the choice probabilities via GMM, I need to have more moment conditions than the number of parameters. Thus my moments are the shares of children choosing schooling across grades, the treatment status, and survey rounds. The resulting number of moment conditions are 49 for the Malawi experiment and 24 for the Moroccan one.[32]

When I estimate the choice probabilities either via MLE, I maximize the likelihood weighted by sampling weights. When estimating via GMM, I also weigh observations with sampling weights. In addition, I use the sample size of each moment as a weighting matrix to prioritize

[32]Since children at baseline were all in school in the Malawi data, I do not observe children in grade 12 at baseline. I also have few children in grade 1 to 6. As a result, the number of moment conditions is 49, less than 72 (12 grades $\times$ 2 groups $\times$ 3 rounds). On the other hand, in the Moroccan data, the number of moment conditions is 24 (6 grades $\times$ 2 groups $\times$ 2 rounds).

the ATEs for larger subgroups. Finally, I solve the minimization problem using the Nelder-Mead algorithm with the MLE estimates as the starting points. I use the Nelder-Mead algorithm because it always returns the same estimates, although the estimation results are robust to different algorithms.

After estimating the parameters of the flexible logit, I obtain the choice probabilities estimates as the predicted values and construct the dependent variable in Equation (2). In this step, I top-code the estimated probabilities higher than 0.99 for two reasons. First, there are observations with the probabilities being 1, which should not happen theoretically, because of the numerical precision of computation software. If $\hat{P}_{it}^1$ or $\hat{P}_{i,t+1}^3$ is numerically 1, then I cannot define the dependent variable. Since this is more likely to occur among the treatment children, I would underestimate the probability of schooling for the treatment group if dropping such observations. The top-coding of the choice probabilities avoids this issue by including them in the subsequent estimations. Second, without the top-coding, I find $\hat{P}_{i,t+1}^2$ and $\hat{P}_{i,t+1}^3$ higher for the treatment group than the control one in all experiments while no difference in $\hat{P}_{it}^1$. This is consistent with the treatment effects taking place just after the first round of household surveys. However, I find a significantly smaller $\frac{\hat{P}_{i,t+1}^2}{1-\hat{P}_{i,t+1}^3}$ for the treatment group. This happens because without the top-coding, observations with estimated probabilities extremely close to 1 can have huge values of $\frac{\hat{P}_{i,t+1}^2}{1-\hat{P}_{i,t+1}^3}$. For instance, $\left(\hat{P}_{i,t+1}^2, \hat{P}_{i,t+1}^3\right) = (0.9999, 0.9999)$ has more than 100 times large values than $\left(\hat{P}_{i,t+1}^2, \hat{P}_{i,t+1}^3\right) = (0.99, 0.99)$, although both can be interpreted as the child almost surely chooses schooling. If these observations are more likely to be in the control group, then I could have a smaller $\frac{\hat{P}_{i,t+1}^2}{1-\hat{P}_{i,t+1}^3}$ for the treatment group. By top-coding the probabilities estimates above 0.99, I do not distort the average of $\left(\hat{P}_{it}^1, \hat{P}_{i,t+1}^2, \hat{P}_{i,t+1}^3\right)$ while I have a significantly larger $\frac{\hat{P}_{i,t+1}^2}{1-\hat{P}_{i,t+1}^3}$ for the treatment group.

Estimates of parameters in flexible logit

	Malawi		Morocco			
	CCTs		CCTs		LCTs	
	MLE	GMM	MLE	GMM	MLE	GMM
e_{it}	0.091	0.330	27.425	27.569	13.049	10.409
y_{it}	0.357	3.174	-0.114	-0.453	-0.351	-0.403
s_{it}	-3.505	2.382	381.601	382.367	256.707	299.118
e_{it}^2	-0.016	-0.014	-1.399	-1.394	-0.577	-0.365
y_{it}^2	-0.008	-0.065	-0.003	0.056	0.001	0.164
s_{it}^2	-15.241	-10.048	27.342	28.698	-173.538	-174.620
$e_{it} \times y_{it}$	-0.023	-0.348	0.053	0.010	0.059	0.300
$y_{it} \times s_{it}$	0.044	-3.783	0.421	-0.166	1.361	-10.152
$s_{it} \times e_{it}$	1.214	1.361	-77.725	-77.687	-39.236	-38.758
r_2	-25.358	-27.056	104.663	101.515	133.165	140.028
r_3	-27.848	-33.179	-	-	-	-
$e_{it} \times r_2$	1.757	1.485	-27.888	-27.751	-33.495	-36.704
$y_{it} \times r_2$	0.411	1.297	0.156	0.811	0.590	3.219
$s_{it} \times r_2$	4.952	7.685	-383.342	-386.879	-504.754	-566.107
$e_{it} \times r_3$	1.517	1.826	-	-	-	-
$y_{it} \times r_3$	1.154	1.723	-	-	-	-
$s_{it} \times r_3$	1.363	-0.280	-	-	-	-
$e_{it}^2 \times r_2$	-0.098	-0.122	1.366	1.304	1.392	1.184
$y_{it}^2 \times r_2$	-0.047	1.566	0.000	-0.035	-0.003	-0.172
$s_{it}^2 \times r_2$	15.513	13.557	-26.469	-26.383	-23.792	-55.242
$e_{it}^2 \times r_3$	-0.059	-0.061	-	-	-	-
$y_{it}^2 \times r_3$	-0.017	0.372	-	-	-	-
$s_{it}^2 \times r_3$	14.575	14.291	-	-	-	-
$e_{it} \times y_{it} \times r_2$	0.022	0.261	-0.054	-0.002	-0.087	0.114
$y_{it} \times s_{it} \times r_2$	-0.546	-3.771	-0.429	1.085	-1.429	-1.853
$s_{it} \times e_{it} \times r_2$	-1.341	-1.919	78.072	77.831	100.196	115.435
$e_{it} \times y_{it} \times r_3$	-0.066	-0.249	-	-	-	-
$y_{it} \times s_{it} \times r_3$	-0.060	-5.051	-	-	-	-
$s_{it} \times e_{it} \times r_3$	-1.038	-0.626	-	-	-	-
Constant	19.652	20.662	-100.129	-99.584	-53.568	-52.544

Note: Observations are weighted by sampling weights.

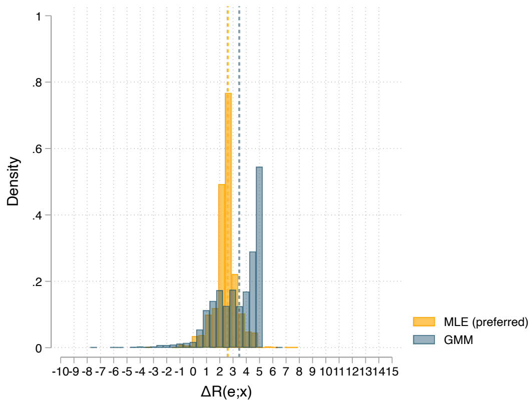
Estimates of θ separately for using MLE and GMM estimates

	Malawi		Morocco			
	CCTs		CCTs		LCTs	
	(1)	(2)	(3)	(4)	(5)	(6)
θ	1.008***	0.834***	5.254***	2.670***	7.884	38.90***
	(0.256)	(0.283)	(0.247)	(0.454)	(4.818)	(11.30)
Obs.	1479	1479	4981	4981	3016	3016
1st stage F-value	113.011	113.011	3843.510	3843.510	25.483	25.483
CCP estimation	MLE	GMM	MLE	GMM	MLE	GMM

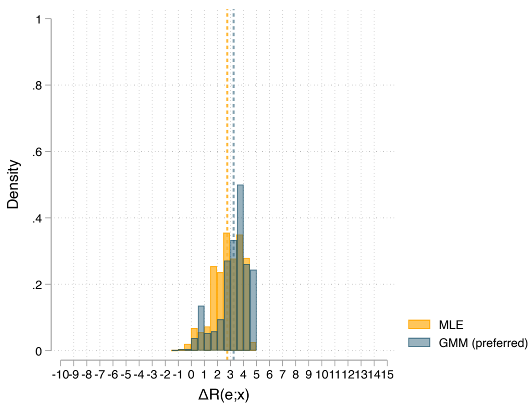
Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. I report the Kleiberge-Paap F statistics for weak identification.
*** p<0.01 ** p<0.05 * p<0.1

Empirical distribution of $\Delta R(e; x)$ separately for using MLE and GMM estimates

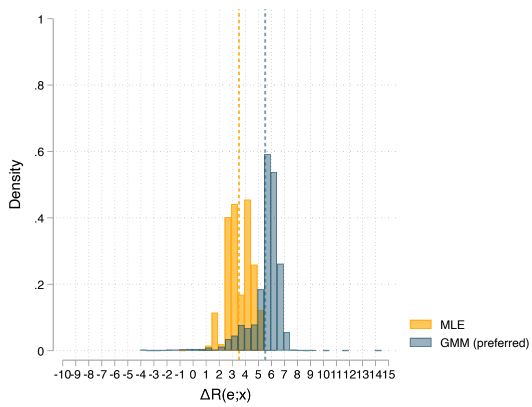
Malawi CCTs



Morocco CCTs



Morocco LCTs



Replication of ATE for all interventions

	Malawi		Morocco			
	MLE (preferred)	GMM	MLE		GMM (preferred)	
	CCTs	CCTs	CCTs	LCTs	CCTs	LCTs
	(1)	(2)	(3)	(4)	(5)	(6)
ATE	0.0317***	-0.00467	0.0518***	0.00318	0.0554***	0.0539***
	(0.00495)	(0.0124)	(0.00235)	(0.00200)	(0.00777)	(0.00954)
Control mean	0.895***	0.909***	0.897***	0.933***	0.894***	0.900***
	(0.00241)	(0.00540)	(0.00201)	(0.00138)	(0.00747)	(0.00889)
Obs.	1490	1490	4982	3018	4982	3018
Target ATE	0.037	0.037	0.057	0.073	0.057	0.073
= Target ATE	0.290	0.001	0.038	0.000	0.869	0.051
95% CI of ATE	[0.022, 0.041]	[-0.029, 0.020]	[0.047, 0.056]	[-0.001, 0.007]	[0.040, 0.071]	[0.035, 0.073]
Target control mean	0.896	0.896	0.894	0.893	0.894	0.893
= Target control mean	0.721	0.013	0.074	0.000	0.981	0.476
95% CI of control mean	[0.890, 0.899]	[0.899, 0.920]	[0.893, 0.901]	[0.930, 0.936]	[0.879, 0.909]	[0.882, 0.917]

Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. I report p-values from F-tests on the null hypothesis that the predicted value is equal to the estimated one, separately for the ATE and the control mean.
*** p<0.01 ** p<0.05 * p<0.1

Model fit at individual level for all interventions

	Malawi	Morocco	
	CCTs	CCTs	LCTs
Correct	0.828	0.888	0.887
Overestimation	0.080	0.057	0.057
Underestimation	0.091	0.055	0.056

Note: I use the second round of household surveys as the estimation samples. Observations are weighted by sampling weights.

Comparison of extrapolation methods in across-contexts extrapolation

As shown in Table 8 and 9, the predictions of ATE via the across-contexts extrapolation vary by the extrapolation methods. Specifically, the reduced-form methods predict at 21.2 percentage points in the heterogeneous treatment effects approach and null effects in the propensity score weighting while the structural methods does at 4.3 or 4.1 percentage points, the latter of which is numerically closer to the target estimate.

One explanation for the better prediction when using the structural method is the normalization of variables. In particular, cash transfer amount and school costs are (implicitly) defined relative to per-capita income in the structural model, the reduced-form methods use the absolute values in prediction. Given that school costs and per-capita income are on average higher in the Moroccan experiment than the Malawi one, ignoring these level differences could lead to poor predictions. For example, the heterogeneous treatment effects approach has a positive coefficient for per-capita income, which could explain the substantial overprediction of enrollment rates under the Moroccan CCTs. The propensity score weighting approach puts a large weight for the Malawi child with high income, who might not benefit from the CCTs. While whether the normalization of these variables improves predictions or not is ex ante ambiguous, it could be the case in my research setting ex post.

The table below shows the performance of the across-contexts extrapolation when normalizing covariates used in the reduced-form methods. Column (1) and (4) reproduce the results in Table 8. Column (2) and (5) show results when cash transfer amount and school costs are defined as a fraction of per-capita income. The predictions do not improve relative to those in column (1) and (4), suggesting that the structural method is better able to predict ATE not because of defining cash transfer amount and school costs in relative terms. Column (3) and (6) show results when years of education and children's age are additionally standardized. Surprisingly, the predictions substantially improve for both extrapolation methods. This is consistent with the analysis that years of education should have a common support across the contexts to better extrapolate the perceived relative returns to schooling when using the structural method for the across-contexts extrapolation.

Reduced-form extrapolation with normalization of variables

	HTE			PSW		
	(1)	(2)	(3)	(4)	(5)	(6)
ATE	0.212***	0.0308***	0.0352***	0.00660	-0.0459**	0.0655***
	(0.00442)	(0.00421)	(0.00346)	(0.0184)	(0.0202)	(0.0140)
Control mean	1.127***	1.111***	0.894***	0.895***	1.000***	0.920***
	(0.00373)	(0.00367)	(0.00302)	(0.0128)	(0.00221)	(0.0112)
Obs.	4982	4982	4982	1490	1490	1490
Target ATE	0.057	0.057	0.057	0.057	0.057	0.057
= Target ATE	0.000	0.000	0.000	0.007	0.000	0.534
95% CI of ATE	[0.203, 0.221]	[0.023, 0.039]	[0.028, 0.042]	[-0.030, 0.043]	[-0.086, -0.006]	[0.038, 0.093]
Target control mean	0.894	0.894	0.894	0.894	0.894	0.894
= Target control mean	0.000	0.000	0.833	0.927	0.000	0.022
95% CI of control mean	[1.120, 1.134]	[1.104, 1.118]	[0.888, 0.900]	[0.869, 0.920]	[0.996, 1.005]	[0.897, 0.942]
Normalization of s, y, z		✓	✓		✓	✓
Normalization of e, age			✓			✓

Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights unless indicated otherwise. HTE indicates the heterogeneous treatment effects approach while PSW does the propensity score weighting. Normalization of s, y, z means cash transfer amount and school costs are defined as a fraction of per-capita income. Normalization of e, age means years of education and children's age are standardized.
*** p<0.01 ** p<0.05 * p<0.1

Supplementary analysis

Robustness of linear extrapolation of $\widehat{\Delta R}(e;x)$

	Across-policies			
	(1) 1st	(2) 2nd (preferred)	(3) 3rd	(4) 4th
ATE	0.0564***	0.0590***	0.0575***	0.0582***
	(0.00550)	(0.00545)	(0.00545)	(0.00547)
Control mean	0.944***	0.941***	0.943***	0.942***
	(0.00536)	(0.00531)	(0.00532)	(0.00534)
Obs.	4982	4982	4982	4982
Target ATE	0.057	0.057	0.057	0.057
= Target ATE	0.947	0.674	0.886	0.793
95% CI of ATE	[0.046, 0.067]	[0.048, 0.070]	[0.047, 0.068]	[0.047, 0.069]
Target control mean	0.894	0.894	0.894	0.894
= Target control mean	0.000	0.000	0.000	0.000
95% CI of control mean	[0.933, 0.954]	[0.931, 0.952]	[0.932, 0.953]	[0.931, 0.952]
	Across-contexts			
	(1) 1st	(2) 2nd (preferred)	(3) 3rd	(4) 4th
ATE	0.0132***	0.0431***	0.0155***	0.0365***
	(0.000220)	(0.00465)	(0.000972)	(0.0111)
Control mean	0.922***	0.702***	0.901***	0.614***
	(0.000177)	(0.00390)	(0.000855)	(0.00927)
Obs.	4982	4982	4982	4982
Target ATE	0.057	0.057	0.057	0.057
= Target ATE	0.000	0.004	0.000	0.070
95% CI of ATE	[0.013, 0.014]	[0.034, 0.052]	[0.014, 0.017]	[0.015, 0.058]
Target control mean	0.894	0.894	0.894	0.894
= Target control mean	0.000	0.000	0.000	0.000
95% CI of control mean	[0.921, 0.922]	[0.694, 0.710]	[0.899, 0.903]	[0.595, 0.632]

Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. I report p-values from F-tests on the null hypothesis that the predicted value is equal to the estimated one, separately for the ATE and the control mean. 1st, 2nd, 3rd, and 4th indicate the order of polynomials I include when using the parametric extrapolation of $\Delta R(e;x)$.

*** p<0.01 ** p<0.05 * p<0.1

Across-policies extrapolation with replacement of $\hat{\theta}$ and $\widehat{\Delta R}(e;x)$

	Across-policies			
	Linear		RF	
	(1)	(2)	(3)	(4)
ATE	0.00290***	0.297***	0.00301***	0.297***
	(0.000175)	(0.0190)	(0.000148)	(0.0190)
Control mean	0.993***	0.702***	0.993***	0.702***
	(0.000157)	(0.0184)	(0.000134)	(0.0184)
Obs.	4982	4982	4982	4982
Replace $\hat{\theta}$	✓		✓	
Replace $\widehat{\Delta R}(e;x)$		✓		✓
Target ATE	0.057	0.057	0.057	0.057
= Target ATE	0.000	0.000	0.000	0.000
95% CI of ATE	[0.003, 0.003]	[0.260, 0.335]	[0.003, 0.003]	[0.260, 0.335]
Target control mean	0.894	0.894	0.894	0.894
= Target control mean	0.000	0.000	0.000	0.000
95% CI of control mean	[0.992, 0.993]	[0.666, 0.739]	[0.992, 0.993]	[0.666, 0.739]

Note: Clustered standard errors (randomization units) in parentheses. Sampling strata fixed effects are included in all regressions. Observations are weighted by sampling weights. Linear indicates the linear extrapolation of $\Delta R(e;x)$ using the second order polynomials of years of education while RF does the Random Forest algorithm on years of education. Replace $\hat{\theta}$ and Replace $\widehat{\Delta R}(e;x)$ mean I replace each parameter with the estimated value under the Moroccan CCTs. I report p-values from F-tests on the null hypothesis that the predicted value is equal to the estimated one, separately for the ATE and the control mean.

*** p<0.01 ** p<0.05 * p<0.1

Estimate of θ when age is added to state variables under Malawi CCTs

	Malawi CCTs	
	(1)	(2)
θ	1.008***	1.031***
	(0.256)	(0.249)
Obs.	1479	1479
1st stage F statistics	113.011	113.011
CCP estimation	MLE	MLE
Age added		✓

Note: Clustered standard errors (randomization units) in parentheses. Observations are weighted by sampling weights. I report the Kleiberge-Paap F statistics for weak identification.
*** p<0.01 ** p<0.05 * p<0.1

Empirical distribution of $\widehat{\Delta R}(e;x,\text{age})$ under Malawi CCTs