

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

Kensuke Maeba

Northwestern University

May, 2023

Abstract

Predicting the effects of a new policy often relies on existing evidence of the same policy in other contexts. However, cross-contexts predictions may fail because those contexts may have distinct characteristics and the same policies may function differently across contexts. In such cases, one might want to make predictions based on similar policies in one's own context to hold local factors constant. This paper compares these approaches using cash transfer programs in Malawi and Morocco. By predicting the treatment effect 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 the sources of the difference, I estimate a dynamic model of schooling decisions under each intervention separately and compare the estimated parameters across the interventions. I find that perceived returns to schooling relative to outside options explain the differential predictions. I suggest that differences in the underlying mechanisms of the Malawi and Moroccan CCTs are reflected in the cross-contexts variation of the relative returns.

*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 at UIUC for helpful comments. Juliana Sánchez Ariza provided excellent research assistance. All errors are my own.

Maeba: kensukemaeba2022@u.northwestern.edu

1 Introduction

Predicting the effects of a new policy is a central challenge for policymakers. The accuracy of such predictions depends on the internal and external validity of the evidence. While methodological developments in causal inference have significantly improved the internal validity of existing evidence, concerns about external validity remain. Previous research on out-of-sample predictions has primarily focused on extrapolating 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 aims to shed light on the usefulness of across-policies extrapolation in the case of conditional cash transfer programs designed to improve educational outcomes.

Accurate predictions about the effect size of a new policy are crucial for policymakers to make informed decisions about its cost-effectiveness. Conditional cash transfers (CCTs) serve as a prime example of the importance of cost-effectiveness in policymaking. CCTs provide cash to poor households conditional on certain behavioral requirements, and have become a popular anti-poverty policy in developing countries. The most common version of CCTs requires children to attend school regularly, which theoretically increases school enrollment through an income and substitution effect by lowering the price of schooling.² The estimated effect size, however, varies across CCT programs, suggesting potential heterogeneity in their cost-effectiveness.³ Therefore, it is important for policymakers in developing countries who consider CCTs to improve domestic education to quantify the effects of CCTs beforehand to determine whether they are a cost-effective policy to implement.⁴

When microdata are available, researchers often rely on extrapolating from CCTs in neighboring contexts to make predictions, assuming that treatment effects remain constant across contexts conditional on various covariates.⁵ However, the assumption of comparability across contexts is not always valid even for geographically proximate regions. Moreover, as CCTs

¹The word “context” includes target populations and locations as well as details about the implementation of policies. Allcott (2015) defined “site” similarly.

²The 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.

³See Baird et al. (2014) for a recent review on the effects of CCTs on education outcomes.

⁴The fiscal and administrative costs also contribute to heterogeneity in the cost-effectiveness of CCTs. See García and Saavedra (2017) for a meta-analysis on the cost-effectiveness of various CCT programs and Caldés et al. (2006) for a comparison of the cost structure of CCTs across three programs in Mexico, Honduras, and Nicaragua.

⁵An even simpler approach without microdata is to make predictions based on a collection of treatment effect estimates of CCTs in other places. I show the prediction performance of this approach in Appendix A.

are predominantly implemented in Latin America, finding comparable neighboring contexts may not always be feasible. Therefore, policymakers may consider extrapolating from different policies in their own contexts if they are informative about the behavioral responses expected under CCTs. For example, positive rainfall shocks to household income arguably increase school enrollment through a similar income effect, despite the source of the income shocks being 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 about the effects of CCTs.

To tackle this question, I examine cash transfer programs in Malawi and Morocco (Baird et al., 2011; Benhassine et al., 2015). The two programs are an ideal setting to study the question for several reasons. First, there are three cash transfer interventions that are necessary to define the two extrapolations. Both programs implemented CCTs that provided cash to households conditional on regular school attendance. In addition, the Moroccan experiment had another cash transfer intervention (Labeled Cash Transfers, or hereafter referred to as LCTs), which provided cash irrespective of school attendance. By using these three interventions, I define the across-contexts extrapolation as extrapolating from the Malawi CCTs to the Moroccan CCTs while the across-policies extrapolation is from the Moroccan LCTs to the Moroccan CCTs. Second, both programs were implemented as randomized controlled trials (RCTs), providing reliable estimates of treatment effects thanks to a credible source of identifying variation. This exogenous variation also allows my extrapolation methods to derive causal interpretations so that I can focus my analysis on the external validity of the interventions. Finally, the microdata of the two programs are publicly available. To the best of my knowledge, the two programs are the only combination that allows me to study the extrapolation problem about the effects of CCTs on education in the same continent. By combining these experiments, I set up a prediction exercise. I predict the average effect, or intention-to-treat effect, of the Moroccan CCTs on school enrollment rates by extrapolating from the other two interventions. After making these predictions, I then evaluate them against the actual estimate of the treatment effect. In my data, I estimate that the Moroccan CCTs increased the enrollment rate of the treatment group by 5.7 percentage points from a base of 89.4 percent.

My across-contexts extrapolation uses two off-the-shelf reduced-form approaches that are commonly used for out-of-sample prediction: heterogeneous treatment effects approach and propensity score weighting. The first approach estimates a heterogeneous treatment effect model

using the sample from the Malawi CCTs and applies the estimated model to the Moroccan CCTs' sample. The second approach combines the two samples to compute the propensity of being in the Malawi CCTs' sample and predict the treatment effect with the Malawi CCTs' sample weighted by the inverse of the propensity score. A key assumption behind these approaches is that conditional on observables, potential outcomes are independent of contexts so that policy effects (or in other words, how CCTs induce schooling) are the same across contexts. To the extent that this assumption is satisfied, the across-contexts extrapolation makes accurate predictions.

My across-policies extrapolation, on the other hand, uses a structural approach. I 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 they pay school costs and obtain a year of education. When the child chooses not schooling, then they consume all of their 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. By leveraging the shocks to flow budget constraints created by the Moroccan LCTs, I causally estimate this model and simulate schooling decisions for the Moroccan CCTs' sample. The resulting predictions would be accurate if the substitution effect of the CCTs is small relative to the income effect and the income effect of the CCTs is similar in magnitude to that of the LCTs, or if both the CCTs and LCTs encourage schooling via common mechanisms different from alleviating flow budget constraints.

My extrapolation results show that the across-policies extrapolation outperforms the across-contexts extrapolation. The across-contexts extrapolation underpredicts the treatment effect. While the heterogeneous treatment effects approach predicts that the treatment effect is 1.56 percentage points, the propensity score weighting approach finds a null effect, neither of which has a 95 percent confidence interval that contains the true estimate. In contrast, the across-policies extrapolation makes accurate predictions. It predicts that the treatment effect is 5.77 or 5.90 percentage points and the equality between the prediction and the true estimate is not rejected at a significance level of 0.05. The results are robust when I use a data-driven way of selecting covariates as well as when I apply the same structural approach in the across-contexts extrapolation. Therefore, the across-policies extrapolation yields both numerically and

statistically accurate predictions, compared to the across-contexts extrapolation.

To understand the sources of the differential predictions of the two extrapolations, I collapse my comparative analysis to the comparison of two parameters in my structural model that are key to illustrate an inter-temporal trade-off regarding schooling decisions - flow utility cost of schooling and perceived returns to schooling relative to outside options - across the interventions. In particular, I estimate my model separately for each intervention, and compare the estimated parameters across the interventions to determine which parameter explains the disparity in the extrapolation results.

I find that the differences between the two extrapolations stem from the perceived relative returns to schooling. I first observe that the estimates of the utility cost of schooling are numerically closer for the same policy (or across the contexts) than the same context (or across the policies). This already implies that the relative returns should be similar across the policies and are essential for accurate predictions, given the better performance of the across-policies extrapolation. I verify this similarity by comparing the estimates of the relative returns under the Moroccan CCTs with the extrapolated values from the two interventions. I find that the predicted values from the Moroccan LCTs are decreasing in years of education and are parallel to the estimated values, whereas those from the Malawi CCTs are increasing over the same domain. I also conduct a decomposition analysis in which I replace one of the parameter values used in the across-contexts extrapolation with the true values estimated with the Moroccan CCTs. I find that the prediction improves only when I use the true values for the relative returns. Finally, I show that the differences in the relative returns across the contexts cannot be fully eliminated by examining observationally similar subsamples across the contexts.

I then argue that the remaining variation in the perceived relative returns to schooling can be attributed to differences in the underlying mechanisms regarding how the cash transfers induce schooling. This is because the relative returns are identified via the residual variation in schooling that is not explained by exogenous changes in the flow utility cost of schooling due to the cash transfers. In other words, the identification of the utility cost is built on the economic intuition that cash transfers induce schooling by relaxing flow budget constraints, and any deviations from that are captured by the relative returns. This means that the estimates of the utility cost are similar between the two CCTs, to the extent that schooling decisions in both contexts are explained by the economic reasoning. On the other hand, the estimated relative returns are similar between the two Moroccan interventions, to the extent that schooling

decisions under those interventions are driven by any other mechanisms. The fact that the relative returns account for the prediction differences is indeed aligned with the prior that the across-policies extrapolation is more successful than the across-contexts extrapolation when the Moroccan CCTs induce schooling in a similar way to the Moroccan LCTs but differently from the Malawi CCTs.

Consistent with this discussion, Benhassine et al. (2015) provided two possible explanations for the similarities in the underlying mechanisms between the CCTs and LCTs: a signaling effect and a misunderstanding of the conditionality. First, the authors argued that the signaling effect about the importance of education was attached to the two cash transfer interventions through the endorsement of the experiment by the Ministry of Education and could have driven the effects of the interventions. Second, they also argued that the treated households in the CCT arm misunderstood that the cash transfers were unconditional. While both explanations suggest similarities in the underlying mechanisms of the two interventions, it is unlikely that the confusion about the conditionality is the main source of similarity, given that the estimates of the flow utility cost of schooling sufficiently differ across the two interventions. Corroborating this, in my counterfactual analysis, I demonstrate that removing the misunderstanding of the conditionality does not alter the relative performance of the two extrapolation approaches. Thus, I conclude that the two interventions may be similar perhaps due to the signaling effect, which may not be the primary driver of the effects of the Malawi CCTs.

Although my analysis does not claim that extrapolation across policies always outperforms extrapolation across contexts in prediction accuracy, it instead shows that policies that are ostensibly similar may have different underlying mechanisms, leading to the failure of naive extrapolation across contexts. My analysis then suggests that a local policy that resembles how the target policy of interest works can be exploited via structural approaches to obtain accurate out-of-sample predictions. In other words, accurate out-of-sample predictions require the extrapolation of policies with the same underlying mechanisms, whether extrapolating across contexts or policies.

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 has focused on extrapolation from one context to another for the same policies and proposed 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)).⁶ Those approaches, however, may fail to yield accurate predictions if the target context differs significantly from the sample contexts, even after controlling for various observable characteristics. For example, Hotz et al. (2005) found that extrapolating the effects of job training programs for workers without working experiences based on observationally similar workers in training datasets led to inaccurate predictions. Allcott (2015) provided evidence of site selection bias, indicating that earlier experimental sites are positively selected. This paper contributes to this literature by demonstrating that differences in underlying mechanisms can also lead to poor predictions via extrapolation across contexts. I also show that when extrapolation across contexts fails to produce reliable predictions, exploiting variation within contexts can serve as an alternative way to do so. This paper presents the first example of such a case and discusses when extrapolation across policies outperforms extrapolation across contexts.

The closest comparison to this paper is Gechter et al. (2018). They evaluated various extrapolation methods based on a welfare measure in the case of CCTs. In particular, they compared reduced-form approaches for extrapolation from experimental data in a different context with structural approaches 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 focused on this trade-off in the case of microcredit and quantified 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 an 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 used 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 estimate an identical model with the three interventions under consideration and compare estimated parameters across the interventions to highlight the sources of prediction differences

⁶Several papers scrutinized extrapolation in different directions. Banerjee et al. (2017) summarized difficulties in extrapolation from local contexts to global ones. DellaVigna and Linos (2022) compared nudge experiments conducted by academic researchers with those by non-academic institutions.

in the two extrapolations.

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 Kalouptsi et al. (2021). I take this approach to discuss the causal identification of my model, and to relax rational expectations assumption by leveraging randomized treatment assignment. Relaxing this assumption makes my model more realistic, especially because low-income households in developing countries may have biased perceptions of 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 a prevalent topic analyzed with dynamic discrete choice models, my paper will serve as a benchmark that achieves a new identification for numerous 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 school attendance (CCTs and UCTs, respectively).⁷ The aim of the study was to investigate the effects of the conditionality attached to cash transfers on schooling. The target population of the experiment was the never-married girls of aged 13 to 22 at school who were at 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 had 2087

⁷The conditionality of the CCTs was to attend more than 80 percent of schooling days. Cash transfers were made monthly if children satisfied this requirement in the previous month.

school girls in total for analysis. They conducted three surveys: a baseline survey between October 2007 and January 2008, the first follow-up survey between October 2008 and February 2009, and 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), and 88 EAs in the control arm (1356 girls). Fifteen EAs to examine spillover effects were dropped in their analysis. Cash transfers were offered at both household and individual levels. The household amount varied across the EAs while the individual amount did within the EAs. To benchmark the size of the transfers, based on the authors' calculation, the household amount for 10 months was on average about 10 percent of the average household annual consumption expenditures. 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 4.5 percent (or 4.1 percentage points) more likely to stay in school than those in the control group after 1 year, and 7.9 percent (or 6.1 percentage points) more likely after 2 years, based on self-reported school enrollment in the household surveys. The authors also estimated the effects of UCTs, but found that the effects 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 unchanged.

Since the misreporting bias observed in the UCTs arm was not observed in the Moroccan experiment, the underlying mechanisms behind the effects of UCTs on self-reported school enrollment rates may not be generalizable across policies and contexts. As a result, for the following extrapolation exercise, I only use the treatment and the control groups in the CCT arm from the Malawi experiment.

2.2 Moroccan experiment

Benhassine et al. (2015) conducted a cash transfer experiment covering 600 poorest rural municipalities in the 5 poorest regions of Morocco. This experiment also had two treatment arms, CCTs and labeled cash transfers (LCTs).⁸ Unlike the CCTs, the LCTs were not tied to regular school attendance. However, like the CCTs, the LCTs were perceived as an education program because children were registered for the experiment at schools in their residential areas and because the program was introduced by the Ministry of Education, both of which implicitly

⁸The conditionality of the CCTs was not to miss school more than 4 times every month.

encouraged school enrollment. The authors investigated whether these nudging effects were sufficient to alter schooling decisions. The target population was children aged 6 to 15 who were at risk of dropping out of primary school. At each school, the authors sampled 8 households that had at least one child in that school over past three years, resulting in 4385 households in total. They conducted a baseline survey in June 2008 and an endline survey in June 2010.

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 gender 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 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 the LCTs increased school enrollment for the treatment group by approximately 10 percent (or 7.4 percentage points) compared to the control group after 2 years. In contrast, the CCTs increased enrollment by roughly 7.3 percent (or 5.4 percentage points). Thus, the LCTs were found to be more effective in increasing enrollment rates than the CCTs. The authors suggested two reasons for the similar effect size of the two interventions. First, there was confusion among parents about the conditionality of the CCTs. Specifically, less than 15 percent 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 functionally similar to the LCTs. Second, both interventions had the signaling effect that education was important through the endorsement of the program by the government. The authors then suggested that in addition to these factors, the LCTs induced schooling among marginal children who were not confident about regular school attendance so that they would not choose schooling under the CCTs, making the LCTs relatively more effective.

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: the heterogeneous treatment effects approach and the 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 gender for the conditioning variables as the target populations in the two experiments differ in these dimensions. Additionally, motivated by my structural model below, I also include pre-treatment variables that are likely to affect schooling decisions; years of education, per-capita income, school costs, and the cash transfer amount. If the assumption holds, then conditional on those variables, the treatment effect of the Malawi CCTs should coincide with that of the Moroccan CCTs.⁹

To test the robustness of the way of selecting the conditioning variables, I also employ a data-driven approach. Specifically, I use the causal forest algorithm proposed by Wager and Athey (2018) with all of the covariates that are available for both interventions. The causal forest algorithm then estimates average treatment effects for various subsamples to identify variables that moderate the treatment effect. More details about this approach are provided in Appendix B.

3.1.2 Heterogeneous treatment effects approach

A common approach to extrapolate across contexts is the heterogeneous treatment effects approach. I first estimate heterogeneous treatment effects based on the pre-selected covariates, denoted by $W_i = (w_{i1}, \dots, w_{iK})'$, 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, school enrollment decisions are predicted for the Moroccan CCTs’ sample, denoted by \tilde{d}_i . In order to make the predicted values lie between 0 and 1, I

⁹Another standard approach is entropy balancing, where sampling weights for the training data are computed to match key moments of selected covariates in the target data (Hainmueller, 2012). Allcott (2015) used this approach in addition to the heterogeneous treatment effects one. One problem with entropy balancing is that the computation of the weights may not converge, which is the case in my data with those variables.

assume that ω_i is a logit error so that \tilde{d}_i is the predicted logit probability of schooling for individual i . Then the treatment effect is predicted by regressing them on treatment assignment and 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 treatment effect by regressing school enrollment decisions on treatment assignment and 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. The purpose of the model is to replicate the effects of the CCTs on school enrollment from those of the LCTs. The effects of the CCTs can be decomposed into the income effect, the substitution effect, and non-pecuniary effects such as the signaling effect highlighted in Section 2.2. While the effects of the LCTs are not informative about the substitution effect, which creates prediction errors in the across-policies extrapolation, the

model estimated under the LCTs can recover the other effects.

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 ($d = 0$) in each period t . If the child chooses schooling, then they pay school costs s out of their per-capita income y , accumulate one year of education, and consume the rest of the income. If the child chooses not to attend school, then they consume all of their income. I define the school costs as payments for tuition fees, textbooks, uniforms, and any other necessary goods, which can vary by the level of education. 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 years of education e they have had, which creates an inter-temporal trade-off of consumption.¹⁰

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

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

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.¹¹ 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:

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

where $x_{it} = (y_{it}, s_{it})'$. 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) .

¹⁰I assume the child does not save. This is consistent with low rates of having any saving technologies in both contexts at baseline.

¹¹I 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 θ and $R(e; x)$, I 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.¹² Finally, Assumption 3 sets the discount factor outside the model as it is generally not identified in standard dynamic discrete choice models (Rust, 1994; Magnac and Thesmar, 2002).

With these assumptions, I begin with identifying θ via the Euler Equations in Conditional Choice Probabilities (ECCP) approach (Scott, 2014; Kalouptsi 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 randomized treatment assignments. Since children or households in my context may have biased beliefs about returns to education, relaxing 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:

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

¹²Note that the essential part of Assumption 2 is that an econometrician knows the distribution of the preference shocks. Thus, the following argument will hold as long as the preference shocks are drawn from a known distribution, although its exposition may be different.

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

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

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 provides a simplified expression for the ex-ante value function:

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

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:

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

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

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

where

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

$\eta_{it}(d)$ is called expectation errors (Scott, 2014; Kalouptsi 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 a given level 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 ,

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

and similarly, if the child chooses non-schooling,

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

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:

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

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 :

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

To tidy the equation, I define several choice probabilities:

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

I rewrite the equation using these expressions:

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

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 decompose the expectation errors into two parts. First, there is heterogeneity in perceived returns to education across households. For example, high- and low-income households may have systematically different expectations about future returns to education at baseline because of information frictions. The other part is forecasting errors, meaning that if 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 randomized treatment assignment of the Moroccan LCTs serves as an IV to resolve this endogeneity problem.¹³ The relevance condition is likely satisfied because the cash transfers are a part of household income. The exclusion restriction for the heterogeneous perceived returns to education requires that the treatment assignment should be orthogonal to households' future expectations at baseline. This would hold true because the randomization did not allow households to select their treatment status based on their beliefs about returns to education and because the history of schooling decisions is not a state variable.¹⁴ 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 groups, on average, make equally inaccurate predictions about future value functions at baseline.¹⁵

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 cash transfers can serve as IVs to identify θ consistently. Moreover, this difference in how the cash transfers work in my model leads to different estimates of θ across the interventions, even if the amount of the cash transfers is the same, due to the substitution effect of the CCTs. In the model, this corresponds to the fact that the CCTs make larger relative changes in the school costs than the LCTs do in the income. Thus, the variation in estimates of θ across the interventions reflects the differences in

¹³The 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 school enrollment rates.

¹⁴The second point allows me to assume that the cash transfers affect schooling decisions only through changes in the flow budget constraints in my identification argument.

¹⁵To argue this point in detail, suppose that 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 their income. Similar reasoning can be made when they believe they are in the control group. If the randomization makes them believe they are assigned to a certain group by chance, then the forecasting error would be balanced across the treatment status.

policy characteristics.

After identifying θ , I proceed to identify the perceived returns to education, $R(e; x)$. Since the identification of this object requires a functional form assumption, I choose to nonparametrically identify differential perceived returns to education across the choices. The identification of the differential returns 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 households receive the cash transfers mainly because of the lack of data on when children usually enter labor markets in both contexts, which could be another candidate for the terminal period.

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

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

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

$\Delta R(e_{i,2}; x_{i,2})$ represents the differential perceived returns to education or the perceived relative returns to schooling, for children with $e_{i,2}$ years of education. Intuitively, this parameter is identified through the residual variation of schooling decisions that is not explained by changes in the relative utility of schooling due to the cash transfers. The parameter varies across the state values because the choice probabilities are a function of the state variables.

My identification argument suggests that θ should differ across the types of the cash transfers while $\Delta R(e_{i,2}; x_{i,2})$ should be similar between them. This is because θ summarizes how the cash transfers relax the flow budget constraints to induce schooling, which is theoretically different between the LCTs and the CCTs. In other words, the model cannot perfectly replicate the effects of the CCTs on the budget constraints from the LCTs, unless the substitution effect of the CCTs is absent. This is the source of prediction errors in the across-policies extrapolation, as θ is assumed to be the same in the extrapolation. On the other hand, $\Delta R(e_{i,2}; x_{i,2})$ captures any deviations from that standard mechanism, such as the psychological effect of the cash transfers.¹⁶ While the model is agnostic about such neoclassical mechanisms reflected in this parameter, it may be reasonable to assume that they are common between the two interventions when doing the across-policies extrapolation, as these mechanisms are likely to be context-dependent. In

¹⁶To see this point more clearly, suppose that children who receive the CCTs or the LCTs choose schooling completely because of non-pecuniary effects. Then, theoretically, their subsidized school costs or complemented income should not explain schooling decisions at all, and hence the probability of schooling and the flow budget constraint in Equation (2) become uncorrelated, leading to θ being 0. Then, the probability of schooling is equal to $\Delta R(e_{i,2}; x_{i,2})$.

summary, the two parameters can be regarded as representing competing mechanisms through which the cash transfers affect schooling.

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 a flexible logit of the state variables to smooth the probabilities across the state space.¹⁷ After obtaining the estimates of the choice probabilities, 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 that can replicate the treatment effect of the Moroccan LCTs with the estimates of the choice probabilities. 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 groups in each survey round and compare them with the true shares. To replicate the treatment effect using my model, these shares need to be sufficiently close to each other, given the effect size of the Moroccan LCTs. If the MLE estimates of the choice probabilities do not meet this criterion, then I estimate the choice probabilities using the same logit but via GMM, where I directly match those shares as the moment conditions. While the GMM estimator might distort the individual choice probabilities to match the shares, in Appendix D, I provide evidence that an estimated model using the GMM estimates can accurately simulate individual schooling decisions. More details about the choice probabilities estimation are available in Appendix C.

3.2.4 Prediction

Having obtained the estimates of θ and $\Delta R(e; x)$ under the Moroccan LCTs, the prediction of the treatment effect of the Moroccan CCTs is made by using $(e_{i,2}, y_{i,2}, s_{i,2})$ from the Moroccan

¹⁷Another common way of smoothing the choice probabilities is to weight 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 provided in Arcidiacono and Miller (2011).

CCTs. I first predict the choice probabilities denoted with a tilde:

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

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

$$\tilde{P}(d = 1|e_{i,2}, y_{i,2}, s_{i,2}) = \delta_1 + \delta_2 \text{Treatment}_i + \text{Stratum}_i + \nu_i, \quad (4)$$

where δ_1 and δ_2 represent the predicted control group's school enrollment rate and treatment effect, respectively.

It is worth noting that extrapolating $\Delta R(e; x)$ is not straightforward as the state space may not overlap across the two interventions. For example, school costs for the treated children under the Moroccan CCTs can be negative if the transfers cover more than the actual costs while they are always positive under the Moroccan LCTs. Thus, I need to extrapolate $\Delta R(e; x)$ for the Moroccan CCTs' children from the empirical distribution of $\Delta R(e; x)$ under the Moroccan LCTs. To do this, I take both parametric and nonparametric approaches. The parametric approach constructs a linear projection of $\Delta R(e; x)$ on (e, e^2) . The nonparametric approach uses the Random Forest algorithm to capture the relationship between $\Delta R(e; x)$ and e . In what follows, I present results for both approaches for robustness.

4 Estimation Results

4.1 Data

My main dataset is 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).¹⁸ I use a baseline, the first follow-up, and an endline surveys from the Malawi data and a baseline and an endline surveys from the Moroccan data. The subscript t in my notation corresponds to a survey round.

Next, I construct the key variables; schooling decisions, years of education, per-capita income, school costs, and cash transfers from the household surveys. The schooling decisions are

¹⁸The 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>.

determined based on whether an individual is currently enrolled in school, which I observe in the data. The years of education can be constructed based on the current grades that children are in. However, due to reporting errors such as an increase of 2 years between survey rounds, 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, validating this construction.¹⁹ The cash transfer amount is also observed in the data. I convert it into an annual term by multiplying the monthly amount by 10, the number of months schools are open in a year. For children in the control group, I impute 0 for their cash transfer amount.

Since children’s income is not measured in both experiments, I use an adjusted per-capita household expenditure to construct the per-capita income as a proxy. I first construct annual household expenditures by multiplying them 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.²⁰ Then I divide the resulting expenditures by the household size adjusted by the OECD equivalence scale, where the household head’s consumption is twice as much as a child’s consumption and each of the other adult members’ consumption is 1.4 times as much. Finally, if a child is in the LCTs treatment group, I add the cash transfers to the per-capita expenditures. This assumes the cash transfers are all consumed by the eligible child in my model. Because the CCTs were provided to cover school costs, I assume that cash transfers are spent entirely for the child even if not used for schooling.

The school costs are constructed from annual individual school expenditures reported by children in the control group who were enrolled in school. I create a menu of the school costs by taking the median of the school expenditures for each grade, weighted by sampling weights.²¹ A concern is that the school expenditures might include private investment in children’s education, which would create variation in the school expenditures by households’ income levels. If the

¹⁹Constructing years of education in this way is not straightforward in the Moroccan data as the household surveys were conducted 2 years apart. In the data, differences in years of education between the surveys vary from 0 to 2, which requires choosing 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.

²⁰If 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.

²¹I pool children in grades 1 to 6 in the Malawi data because of a small number of observations in those grades. Thus, children in those grades face the same school costs.

cross-sectional variation is large, then the same school costs for all children in the same grades are not a reasonable assumption. Figure G.1 in Appendix G shows that the distributions of the individual school expenditures are skewed for all grades, alleviating this concern.²² Finally, I subtract the cash transfer amount from the resulting school costs for children in the CCTs treatment group.

A few more remarks about the construction of the variables are in order. First, I measure the 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.²³ 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 costs when they choose schooling at baseline are not available. 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 each context, these variables are jointly balanced at baseline. Across the contexts, there are several notable differences. First, the average years of education are higher in the Malawi experiment than in the Moroccan experiment because of the differences in the targeting populations. That is, the Malawi sample comprises only girls who were about to enter secondary school and is thus on average older than the Moroccan sample. For the same reason, the gender ratio and children's age differ between the experiments. These demographic differences result in different reasons for school dropout, as illustrated in Figure 1. Specifically, while the Malawi girls quit school mainly due to pregnancy or marriage, the Moroccan children dropped out for various reasons such as poor school quality and domestic work. Second, the income and the school costs are on average higher in the Moroccan experiment, reflecting the fact that Morocco had a higher GDP per capita than Malawi in 2008.²⁴ Third, while the average cash transfer amount is similar across the experiments, its relative size to the income or school costs differs. The relative size of the cash transfers is larger in the Malawi experiment.

Finally, I report the Intention-To-Treat effects of each intervention on the school enrollment

²²Because there are outliers in the reported school expenditures, I use the median instead of the mean.

²³The conversion rates are from The World Bank data: <https://data.worldbank.org/indicator/PA.NUS.FCRF?end=2021&locations=MA-MW&start=1960>.

²⁴According to the World Bank data, the GDP per capita (in PPP) in 2008 was \$1287 for Malawi and \$5890 for Morocco: <https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD?end=2010&locations=MA-MW&start=2000>.

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

rates in my data. The effects 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 sampling strata fixed effects:

$$d_{i,2} = \alpha_1 + \alpha_2 \text{Treatment}_i + \text{Stratum}_i + \nu_i. \quad (5)$$

The estimation results in Table 2 are aligned with the original results reported in the two papers. In my samples, the Malawi CCTs increased the school enrollment rate of the treatment group by 3.7 percentage points (or 4.1 percent). In contrast, the Moroccan CCTs increased it by 5.7 percentage points (or 6.4 percent), and the Moroccan LCTs did so by 7.3 percentage points (or 8.2 percent). As I mentioned in Section 2, the original estimates of the treatment effects varied from 4 to 7 percentage points. Moreover, in the Moroccan experiment, the estimates were higher under LCTs than CCTs. Therefore, although I made different sample restrictions, the estimated effects with my samples are able to capture these features. In the following analysis, I evaluate extrapolation results based on my estimated treatment effect and enrollment rate of the control group in Column (2) of Table 2.

Figure 1: 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.

4.2 Parameter estimates

4.2.1 Reduced-form approaches

I show the estimated coefficients of the two reduced-form approaches in Table 3. The first column displays the estimates of the coefficients from the heterogeneous treatment effects approach using the logit with the Malawi CCTs' sample. The coefficients on the interaction terms suggest that the treatment is more effective for children with less education, higher school costs, or in younger age cohorts at baseline. The second column presents the estimated coefficients from the logit in the propensity score weighting. The coefficients reflect differences in baseline characteristics across the experiments. For example, since the Malawi CCTs' sample was about to enter secondary school, children with more years of education receive a higher propensity score, holding the other variables constant.

4.2.2 Structural approach

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.²⁵ 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

²⁵The full estimation results are available in Appendix C.

Table 2: Estimates of ITT effect on school enrollment rates

	Malawi	Morocco	
	(1)	(2)	(3)
	CCTs	CCTs	LCTs
Treatment	0.0369* (0.0200)	0.0567*** (0.0106)	0.0726*** (0.0107)
Control mean	0.896*** (0.0154)	0.894*** (0.00951)	0.893*** (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

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 percent increase in the cash transfer amount on average increases the probability of schooling by 0.142 percent.²⁶

The empirical distribution of $\Delta R(e; x)$ is presented in Figure 2. The estimates are mostly positive, indicating that the perceived returns to schooling relative to outside options are positive for the Moroccan LCTs' sample. This is reasonable as schooling decisions under the LCTs should be driven by the positive relative returns, as the LCTs would not differentially affect the flow utility cost of schooling across the choices substantially. The average size of the relative returns is slightly greater for the treatment group than the control group, which is counterintuitive because theoretically the relative returns should be larger for children in the control group as their school enrollment is entirely due to the positive relative returns. Thus, the fact that the relative returns are estimated greater for the treatment group indicates that the LCTs impacted the relative returns 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 the signal that education was important for children's future.

²⁶The formula to compute the elasticity is as follows:

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

Table 3: Estimates of coefficients in reduced-form approaches

	HTE	PSW
	(1) = 1 if enrollment (Logit)	(2) = 1 if in Malawi CCTs (Logit)
Treatment	12.06** (5.062)	12.22* (7.256)
Years of education	0.148 (0.235)	4.785*** (1.114)
Per-capita income	0.0795 (0.222)	-0.471** (0.201)
School costs	-0.0972 (1.350)	-36.06*** (12.72)
Cash transfers	0.775 (0.770)	-11.35* (5.818)
Age	-0.513*** (0.0660)	2.677*** (0.927)
Treatment \times Years of education	-1.007** (0.463)	
Treatment \times Per-capita income	0.312 (0.337)	
Treatment \times School costs	7.285** (3.118)	
Treatment \times Age	-0.290* (0.167)	
Constant	9.145*** (1.884)	-60.07*** (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 Kleibergen-Paap F statistics for weak identification.

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

In Appendix D, I check the fitness of my model to the Moroccan LCTs' data by comparing the replicated treatment effect of the Moroccan LCTs with the estimated effect. I find that the equality between the two is not rejected at a significance level of 0.05. I also find that the model correctly simulates 89 percent of the individual schooling decisions with no systematic bias in misprediction. These results indicate that while the model is parsimonious, it successfully describes the schooling decisions under the Moroccan LCTs.

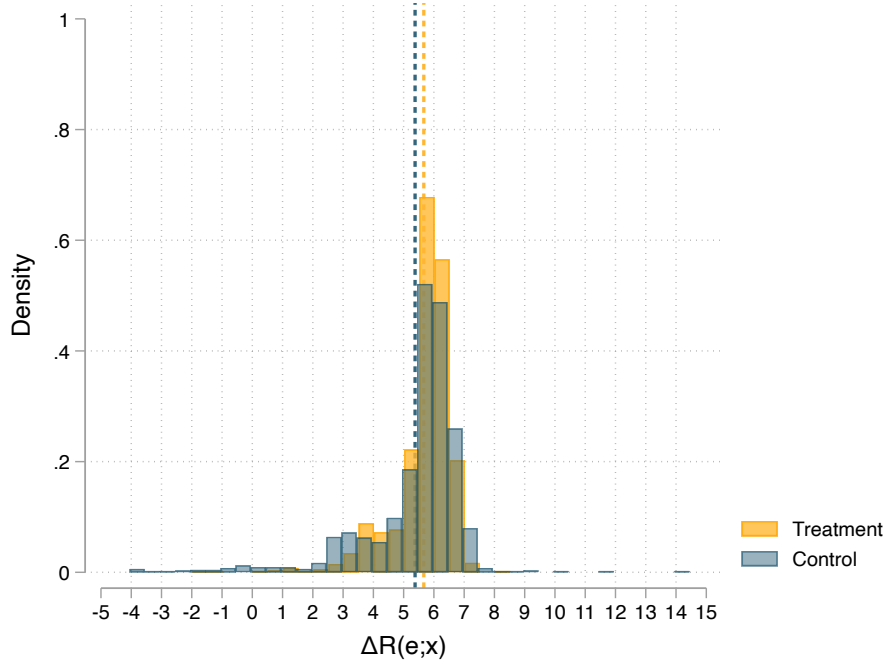
5 Extrapolation Results

5.1 Comparison of two extrapolations

Table 5 present the predicted effect of the Moroccan CCTs on school enrollment rates, as well as the predicted enrollment rates of the control group. The first two columns show the results of the across-contexts extrapolation. I find that the across-contexts extrapolation fails to predict the treatment effect. In column (1), using the heterogeneous treatment effects approach, the across-contexts extrapolation predicts that the Moroccan CCTs increase the enrollment rate of the treatment group statistically significantly by 1.56 percentage points from a base of 98.4 percent. In contrast, in column (2), the propensity score weighting predicts the null treatment effect while it predicts the enrollment rate of the control group as 89.5 percent. Both predictions fail to yield a 95 percent confidence interval that contains the target treatment effect estimate.

As mentioned in Section 3.1.1, I also use the causal forest algorithm, which is a data-driven way to select conditioning variables, to conduct the across-contexts extrapolation. In Appendix B, I show that the causal forest predicts that the Moroccan CCTs increase the

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.

school enrollment rate of the treatment group by 3.7 percentage points, which is statistically significantly different from the target treatment effect estimate. Therefore, the across-contexts extrapolation, regardless of using the reduced-form approaches or the data-driven approach, mispredicts the true treatment effect numerically and statistically.

In contrast, I find that the across-policies extrapolation successfully predicts the treatment effect. Columns (3) and (4) differ in how I extrapolate $\Delta R(e; x)$ when computing the predicted choice probabilities. Despite the difference, the across-policies extrapolation generates similar predictions. It predicts that the Moroccan CCTs increase the school enrollment rate of the treatment group by 5.77 or 5.90 percentage points, which is less than 0.5 percentage points away from the target treatment effect estimate. Furthermore, the null hypotheses on the equality are not rejected at a significance level of 0.05.²⁷ However, the across-policies extrapolation overpredicts the enrollment rate of the control group. In Appendix F, I explain this bias in predicting the enrollment rates with the relative size of the parameter estimates.

One might concern that the results in Table 5 are driven by the extrapolation methods. That is, the structural approach might be better able to transfer the treatment effect estimates than

²⁷Table G.1 in Appendix G shows the extrapolation results using higher order polynomials of e for the linear projection. The results are robust to this modification.

Table 5: Prediction of treatment effect on school enrollment rates
Across-contexts (reduced-form) vs Across-policies (structural)

Target: Morocco CCTs	Across-contexts		Across-policies	
	(1) HTE	(2) PSW	(3) Linear	(4) RF
Treatment	0.0156*** (0.000535)	0.00660 (0.0184)	0.0590*** (0.00545)	0.0577*** (0.00542)
Control mean	0.984*** (0.000515)	0.895*** (0.0128)	0.941*** (0.00531)	0.942*** (0.00529)
Obs.	4982	1490	4982	4982
Target ITT	0.057	0.057	0.057	0.057
= Target ITT	0.000	0.007	0.674	0.863
95% CI of ITT	[0.015, 0.017]	[-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	[0.983, 0.985]	[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 treatment effect and the control mean.

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

the reduced-form approaches.²⁸ To address this concern, I do the across-contexts extrapolation using the structural approach. This is plausible because my model is causally identified under any type of cash transfers that shift the flow utility cost of schooling exogenously, thus it can be applied to the Malawi CCTs (as well as to the Moroccan CCTs).

The new extrapolation results are presented in Table 6. The first two columns show the results of the across-contexts extrapolation results using the structural approach while the last two columns reproduce the across-policies extrapolation results from Table 5. The predicted effect 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 effects are closer to the target estimate, although the 95 percent confidence intervals still do not contain it. In terms of the predictions of the levels, the structural approach underpredicts the school enrollment rate of the control group by 20 percentage points, which is worse than the reduced-form approaches. Thus, the overall performance of the across-contexts extrapolation is not necessarily improved by using the

²⁸Fudenberg et al. (2022) defined 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.

structural approach, and the conclusion about the relative performance of the two extrapolations remains unchanged.²⁹

Table 6: Prediction of treatment effect on school enrollment rates
Across-contexts (structural) vs Across-policies (structural)

Target: Morocco CCTs	Across-contexts		Across-policies	
	(1) Linear	(2) RF	(3) Linear	(4) RF
Treatment	0.0431*** (0.00465)	0.0412*** (0.00644)	0.0590*** (0.00545)	0.0577*** (0.00542)
Control mean	0.702*** (0.00390)	0.676*** (0.00556)	0.941*** (0.00531)	0.942*** (0.00529)
Obs.	4982	4982	4982	4982
Target ITT	0.057	0.057	0.057	0.057
= Target ITT	0.004	0.016	0.674	0.863
95% CI of ITT	[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 treatment effect and the control mean.

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

5.2 Comparison of model parameters

So far, my results show that the across-policies extrapolation outperforms the across-contexts extrapolation in predicting the treatment effect of the Moroccan CCTs on the school enrollment rates. To understand the sources of the differential predictions, I collapse my comparative analysis to the comparison of the two parameters estimated in my structural model. In particular, I estimate my model separately for each intervention, and compare the extrapolated values of θ and $\Delta R(e; x)$ from the two interventions with the estimated values under the Moroccan CCTs. This analysis allows me to point out which parameter is key for successful predictions.

Table 7 presents the estimates of θ under each intervention. Notably, all of the estimates are positive and statistically significantly different from 0, and the 1st stage F-statistics are greater

²⁹In Appendix E, I investigate whether the differential predictions across the extrapolation methods in the across-contexts extrapolation is because the structural approach implicitly measures the school costs and the cash transfer amount relative to the income, whereas the reduced-form approaches use them in absolute values.

than the conventional threshold. This proves that all of the interventions serve as strong IVs to identify θ . The size of θ varies substantially across the interventions, although it is more similar for the two CCT interventions than the two Moroccan interventions. 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 estimate under the Malawi CCTs than the Moroccan LCTs.

Table 7: Comparison of estimates of θ across interventions

	Malawi	Morocco	
	(1)	(2)	(3)
	CCTs	CCTs	LCTs
θ	1.008*** (0.256)	2.670*** (0.454)	38.90*** (11.30)
Obs.	1479	4981	3016
1st stage F statistics	113.011	3843.510	25.483
CCP estimation	MLE	GMM	GMM
= target θ	0.000		0.001

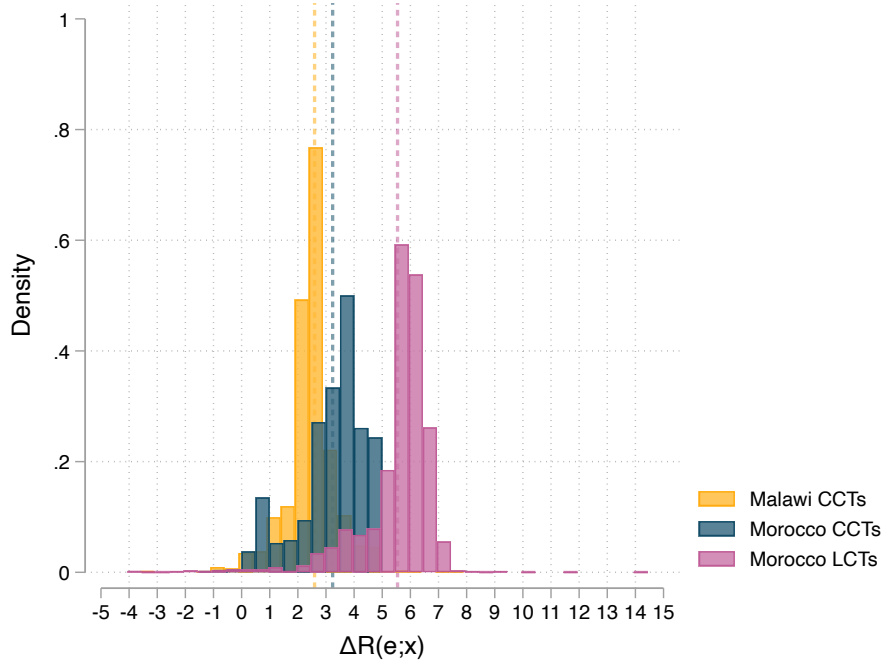
Note: Clustered standard errors (randomization units) in parentheses. Observations are weighted by sampling weights. I report the Kleibergen-Paap F statistics for weak identification.

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

Figure 3 displays the empirical distributions of $\Delta R(e; x)$. Across all of the interventions, the perceived relative returns to schooling are largely positive. Like the estimates of θ , the average size of the relative returns is more comparable within the policy than within the context. This is because the higher school enrollment rate of the treatment group under the CCTs is explained by a larger relative utility of schooling rather than larger relative returns to schooling. In other words, since the LCTs do not increase the relative utility as much as the CCTs do, an increase in the enrollment rate of the treatment group under the LCTs should be driven mainly by the relative returns. Therefore, the average size of the relative returns is greater under the LCTs and different from the two CCTs.

Although the average size of $\Delta R(e; x)$ is similar within the policy, this does not imply that the estimated $\Delta R(e; x)$ under the Moroccan CCTs is also better approximated in the extrapolation from the Malawi CCTs, provided that the across-policies extrapolation outperforms the across-contexts extrapolation. One explanation that reconciles these facts is that the within-intervention variation in $\Delta R(e; x)$ across the treatment status is more similar within the same

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.

context than within the same policy.

Table 8 shows the average relative utility and perceived relative returns of schooling for the treatment and control groups separately for each intervention. The average relative utility in columns (1) and (2) exhibits larger variation between the groups under the CCTs than the LCTs, which is consistent with the reasoning behind the varying size of the estimates of θ . On the other hand, the average relative returns in columns (3) and (4) show similar patterns across the policies. While the average returns are smaller for the treatment group than the control group under the Malawi CCTs, they are, if anything, greater for the treatment group under the two Moroccan interventions. This similarity in the across-groups variation of $\Delta R(e; x)$ reflects an advantage of the across-policies extrapolation when predicting the treatment effect.

To visualize the similarity in $\Delta R(e; x)$ across the policies, I examine the correlation between the extrapolated values of $\Delta R(e; x)$ from the two interventions and the estimated values of $\Delta R(e; x)$ from the Moroccan CCTs.³⁰ Figure 4 depicts these values plotted against years of education of the Moroccan CCTs' sample. The green-square line represents the estimated values and shows a downward-sloping curve over years of education. The blue-empty-circle line

³⁰Because the performance of both extrapolations does not differ by how I extrapolate $\Delta R(e; x)$ as in Table 6, I show subsequent results using the linear projection unless indicated otherwise.

Table 8: Comparison of average relative utility and returns across interventions

	$E[\theta\Delta u]$		$E[\beta\Delta R_i(e)]$	
	(1)	(2)	(3)	(4)
	Control	Treatment	Control	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.

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

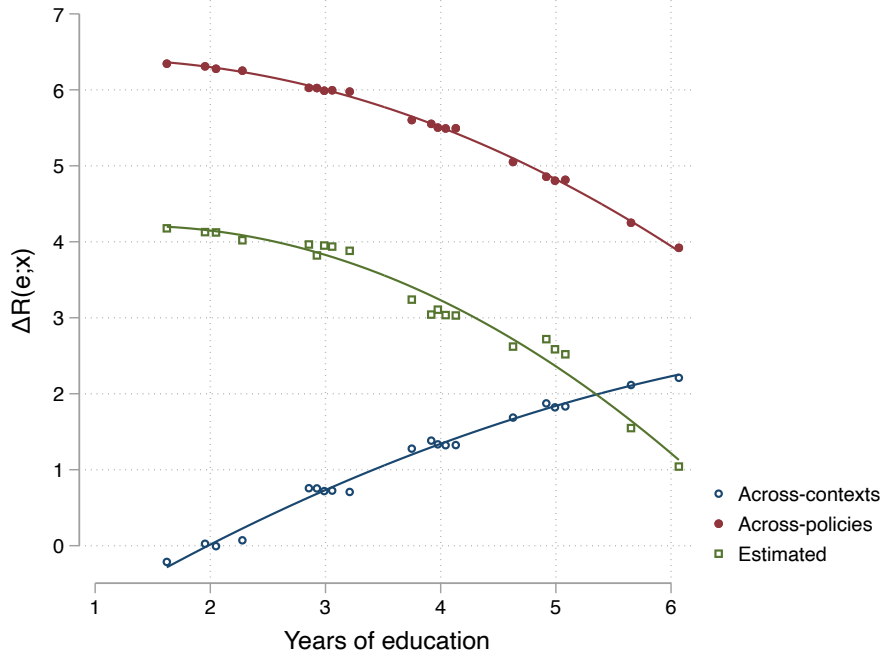
represents the extrapolated values of $\Delta R(e; x)$ from the Malawi CCTs. Unlike the estimated values, the line exhibits an upward-sloping curve, indicating an opposite correlation between $\Delta R(e; x)$ and years of education. In contrast, the red-filled-circle line for the extrapolated values of $\Delta R(e; x)$ from the Moroccan LCTs is parallel to the estimated values. Therefore, although the levels are different, the across-policies extrapolation predicts the distribution of $\Delta R(e; x)$ more accurately than the across-contexts extrapolation.

Finally, to further confirm that the poor performance of the across-contexts extrapolation stems from $\Delta R(e; x)$, I run the across-contexts extrapolation by replacing either θ or $\Delta R(e; x)$ with the estimated values from the Moroccan CCTs. Results in Table 9 show that the performance improves only when I replace $\Delta R(e; x)$. When replacing θ in columns (2) and (3), the treatment effect is overpredicted while the school enrollment rate of the control group is underpredicted similar to the prediction without the replacement. In contrast, when replacing $\Delta R(e; x)$, while the treatment effect is similarly underpredicted as before, the underprediction of the enrollment rate is corrected. As a result, the overall prediction accuracy is improved. Therefore, the analysis suggests that the inaccurate predictions via the across-contexts extrapolation are attributable to the extrapolation of $\Delta R(e; x)$.³¹

One explanation for the opposite correlation pattern in $\Delta R(e; x)$ extrapolated across the

³¹When I do the same exercise for the across-policies extrapolation, I find that the performance becomes worse off when replacing $\Delta R(e; x)$ than θ , which is coherent to the argument in the main text. However, I also find that the across-policies extrapolation underpredicts the treatment effect when replacing θ . This is because if I use the estimated value of θ , then the relative utility of schooling becomes smaller compared to the perceived relative returns to schooling. As a result, the 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 treatment effect becomes smaller than the estimated one. The estimation results are shown in Table G.2 in Appendix G.

Figure 4: Comparison of $\Delta R(e;x)$ 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.

contexts is differences in the education levels of the targeting populations across the experiments. As seen in Figure 4, the estimated values of $\Delta R(e;x)$ for the Moroccan CCTs' sample are 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 perceived returns to these alternative activities increase more rapidly than perceived returns to schooling, then the perceived relative returns to schooling would be decreasing. This implies that the more beneficial outside options are, the more likely children choose to drop out of school.

Figure 5 presents the estimated values of $\Delta R(e;x)$ for the Malawi CCTs' sample against years of education, which exhibit two parts: an increase up to grade 8, which is the final grade in primary school, followed by a decrease. This means the perceived relative returns to schooling are increasing over primary education and decreasing over secondary education for the Malawi CCTs' sample. The increasing relative returns suggest that perceived returns to schooling grow faster than perceived returns to non-schooling activities, which indicates that outside options do not appear attractive compared to schooling during that period. This conjecture is supported by the fact that the Malawi CCTs' sample largely completed primary education, as the experiment

Table 9: Across-contexts extrapolation with replacement of θ or $\Delta R(e; x)$

	Across-contexts			
	Linear	Linear	RF	
	(1)	(2)	(3)	(4)
Treatment	0.0431*** (0.00465)	0.0953*** (0.00486)	0.0909*** (0.00655)	0.0367*** (0.00721)
Control mean	0.702*** (0.00390)	0.688*** (0.00411)	0.663*** (0.00570)	0.901*** (0.00689)
Obs.	4982	4982	4982	4982
Replace θ		✓	✓	
Replace $\Delta R(e; x)$				✓
Target ITT	0.057	0.057	0.057	0.057
= Target ITT	0.004	0.000	0.000	0.006
95% CI of ITT	[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 θ and Replace $\Delta R(e; x)$ mean I use the estimated value of each parameter 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 treatment effect and the control mean.

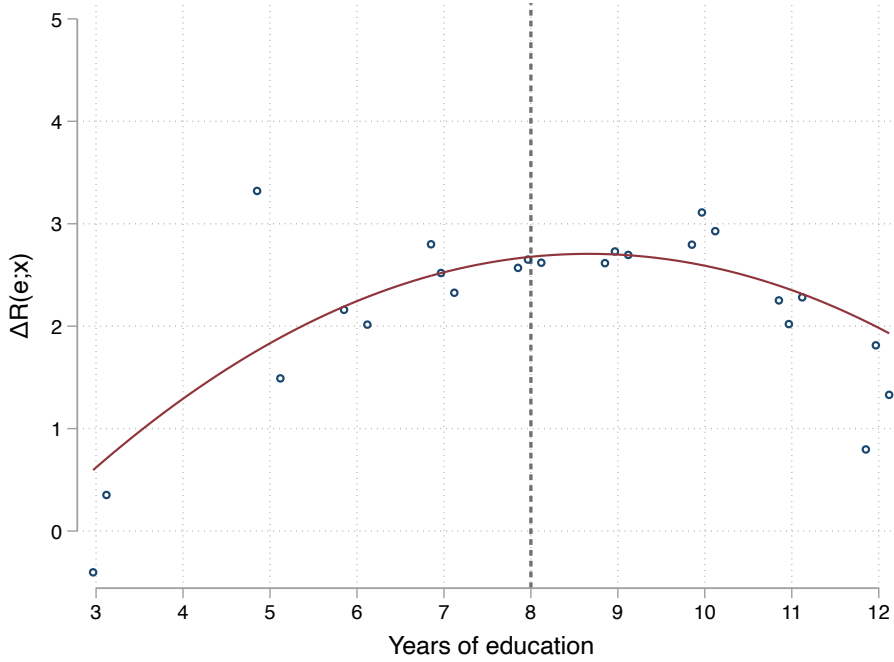
*** p<0.01 ** p<0.05 * p<0.1

targeted girls who were at risk of dropping out of secondary school. Furthermore, the tipping point of the relative returns corresponds to when the Malawi children started choosing dropout. When I extrapolate from the Malawi CCTs, I predict $\Delta R(e; x)$ based on the increasing part, leading to the upward sloping curve of the extrapolated $\Delta R(e; x)$.

Based on the above discussion, it is natural to consider that the poorer prediction of $\Delta R(e; x)$ in the across-contexts extrapolation is simply due to differences in the characteristics of the target populations across the experiments. To examine this possibility, I repeat the across-contexts extrapolation by adjusting three key dimensions in which the target populations differ; years of education, gender, and age.

First, I redo the across-contexts extrapolation with normalizing years of education, motivated by the hypothesis that the poorer prediction of $\Delta R(e; x)$ in the across-contexts extrapolation is simply due to the difference in the timing when children started choosing school dropout, as suggested in Figures 4 and 5. In particular, I first recenter the years of education of the Malawi

Figure 5: $\Delta R(e; x)$ 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.

CCTs' sample by subtracting 7. This modification makes the Malawi children similar to the Moroccan children in terms of the grade at which both could choose to drop out. I also drop the Malawi children in primary education levels when predicting $\Delta R(e; x)$. This sample restriction eliminates the upward-sloping part of $\Delta R(e; x)$ in Figure 5 and thus makes the extrapolated values of $\Delta R(e; x)$ more comparable to the target estimates.

I present the results of the across-contexts extrapolation with the normalization in Table 10. First, I observe that the treatment effect is still underestimated in all of the results, even more so than without the normalization. Second, in contrast, the estimated school enrollment rates of the control group are closer to the target estimate 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 $\Delta R(e; x)$. These results suggest that the normalization improves the overall performance of the across-contexts extrapolation, although it still underestimates the effect of the Moroccan CCTs.

Next, I conduct subsample analyses based on children's gender and age. It is plausible to consider that the across-contexts extrapolation may yield more accurate predictions for a subset of the Moroccan CCTs' sample as the Malawi CCTs' sample is on average older and consists

Table 10: Across-contexts extrapolation with normalized years of education

	Across-contexts			
	Linear		RF	
	(1)	(2)	(3)	(4)
Treatment	0.0150*** (0.000788)	0.0147*** (0.00265)	0.0188*** (0.00215)	0.0189*** (0.00227)
Control mean	0.903*** (0.000674)	0.875*** (0.00221)	0.870*** (0.00186)	0.868*** (0.00195)
Obs.	4982	4982	4982	4982
Normalization	✓	✓	✓	✓
Sample restriction		✓		✓
Target ITT	0.057	0.057	0.057	0.057
= Target ITT	0.000	0.000	0.000	0.000
95% CI of ITT	[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 treatment effect and the control mean.

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

of girls only. Thus, it might be easier to extrapolate the distribution of $\Delta R(e; x)$ across the contexts for subsamples with similar demographics. This is more likely the case if boys and girls have differential returns to education in Morocco. Benhassine et al. (2015) provided suggestive evidence that parental beliefs over children's returns to education in labor markets varied across children's gender.

I show in Table 11 the results 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 substantially improves the overall prediction accuracy with no adverse effects. I do not find improvement differentially across children's gender. For girls, I also examine the predictions separately for older cohorts. Columns (3) and (6) show that as the treatment effect of the Moroccan CCTs for the older girls is larger than for girls of all ages, the predicted effect is also larger for that age group. However, I still underpredict the treatment effect (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 gender in the same age cohorts.

Table 11: Across-contexts extrapolation across age and gender of Moroccan CCTs' sample

	Across-contexts					
	Linear			RF		
	Boys	Girls		Boys	Girls	
	(1)	(2)	(3) Age ≥ 12	(4)	(5)	(6) Age ≥ 12
Treatment	0.0151*** (0.00109)	0.0148*** (0.00129)	0.0221*** (0.00258)	0.0206*** (0.00296)	0.0165*** (0.00343)	0.0346*** (0.00623)
Control mean	0.902*** (0.000913)	0.903*** (0.00112)	0.880*** (0.00222)	0.867*** (0.00250)	0.872*** (0.00297)	0.809*** (0.00552)
Obs.	2666	2313	858	2666	2313	858
Normalization	✓	✓	✓	✓	✓	✓
Target ITT	0.048	0.068	0.139	0.048	0.068	0.139
= Target ITT	0.000	0.000	0.000	0.000	0.000	0.000
95% CI of ITT	[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 treatment effect and the control mean.

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

I also examine the performance of the across-contexts extrapolation across age cohorts of the Malawi CCTs' sample. In particular, when extrapolating $\Delta R(e; x)$, I only use children aged from 13 to 16, which is the highest age of the Moroccan CCTs' sample.³² Table 12 shows results separately for boys and girls in the Moroccan CCTs' sample. I find that the sample restriction on the Malawi CCTs' sample does not further improve the predictions of the across-contexts extrapolation.

Thus, while eliminating differences in years of education narrows the performance gap between the two extrapolations by improving the prediction of the school enrollment rate of the control group, the additional sample restrictions based on children's gender and age do not lead to further improvement in the across-contexts extrapolation predictions. Moreover, neither of these modifications improves the prediction of the treatment effect. Although these results may

³²To 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 Table G.3 and Figure G.2 in Appendix G.

change if adjusting more dimensions in which the two interventions differ, it is unlikely that the prediction accuracy improves substantially, given that the casual forest algorithm with all of the covariates available in my data underpredics the treatment effect as shown in Appendix B. Therefore, observable differences in the target populations across the contexts may not fully explain the disparity in the two extrapolation approaches.

Table 12: Across-contexts extrapolation using young cohorts of Malawi CCTs' sample

	Across-contexts			
	Linear		RF	
	(1) Boys	(2) Girls	(3) Boys	(4) Girls
Treatment	0.00376*** (0.000254)	0.00402*** (0.000318)	0.00664*** (0.000113)	0.00658*** (0.000134)
Control mean	0.980*** (0.000211)	0.980*** (0.000270)	0.963*** (0.0000839)	0.963*** (0.000108)
Obs.	2666	2313	2666	2313
Normalization	✓	✓	✓	✓
Target ITT	0.048	0.068	0.048	0.068
= Target ITT	0.000	0.000	0.000	0.000
95% CI of ITT	[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 treatment effect and the control mean.

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

5.3 What explains variation in perceived relative returns to schooling across contexts?

So far, I have shown that the across-policies extrapolation outperforms the across-contexts extrapolation. This suggests that the two Moroccan interventions function similarly compared to the two CCTs across the contexts. From the decomposition analysis, I have also shown that the discrepancy in the prediction results between the two extrapolation approaches stems from the perceived relative returns to schooling rather than the flow utility cost of schooling. Finally,

the cross-contexts variation in the relative returns cannot be solely attributed to differences in the target populations.

One explanation that aligns with these observations is that the underlying mechanisms of these interventions are reflected in the perceived relative returns to schooling. Intuitively, in my identification argument, the two parameters represent distinct mechanisms regarding how the cash transfers affect schooling decisions. The flow utility cost of schooling is identified with the variation in the probability of schooling explained by the exogenous changes in the utility cost of schooling due to the cash transfers. This identification argument is based on the economic reasoning that the cash transfers induce schooling by relaxing today’s budget constraints. On the other hand, the identification of the relative returns captures any deviations from that mechanism as they are identified with the residual variation. Therefore, the fact that the relative returns are similar for the two Moroccan interventions and play an important role in prediction accuracy suggests that the predictions via the across-policies extrapolation are successful because the effects of the two Moroccan interventions are driven by the same but unconventional mechanisms.

Consistent with this explanation, Benhassine et al. (2015) discussed two possible factors that made the two interventions similar to each other. The first factor is a signaling effect about the importance of education. The authors argued that the endorsement of the program by the Ministry of Education created the signaling effect that could be the main driver of the effects of both interventions. While this additional effect attached to the cash transfers seems to exist in the Malawi experiment according to in-depth interviews in Baird et al. (2011), the size of this effect could vary across the experiments, making it difficult for the across-contexts extrapolation to predict $\Delta R(e; x)$ as accurately as the across-policies extrapolation.³³

Another factor is that the parents of the CCTs’ treated children misunderstood the conditionality attached to the cash transfers. Table 13 shows that 11.1 percent of them in my data correctly understood the conditionality attached to the CCTs. Moreover, only 14.4 percent of them thought that they could receive the transfers by enrolling their children in school. Therefore, the large fraction of the Moroccan CCTs’ sample treated the transfers as unconditional, which could have removed the substitution effects of the CCTs. In contrast, although

³³Baird et al. (2011) implied on page 1720 the existence of the signaling effect: “*The evidence from the in-depth interviews makes it clear that the UCT experiment did not happen in a vacuum. Instead, it took place under a rubric of education that naturally led the beneficiaries to believe that the program aimed to support girls to further their education.*”

my data lack information about how the Malawi CCTs' sample perceived the conditionality, the misunderstanding might have been less prevalent in the Malawi experiment as Baird et al. (2011) showed that the UCTs implemented along with the CCTs did not increase school enrollment. Their in-depth interviews also revealed that girls in the UCT treatment group understood that there was no conditionality attached to the cash transfers. If the income effects of the two Moroccan interventions are comparable, then it might be easier to extrapolate across the policies than the contexts, although the confusion about the conditionality is unlikely to be the primary explanation for the prediction differences, given that the estimates of θ are sufficiently different across the two Moroccan interventions.

Table 13: Knowledge about conditionality in Moroccan experiment

	(1)	(2)
	CCTs	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
Obs.	3704	1740

Note: Observations are weighted by sampling weights. Knowledge about the conditionality is taken from the second round of the Moroccan household survey.

Nevertheless, I examine whether the across-contexts extrapolation could make more accurate predictions if the Moroccan CCTs' sample understood the conditionality correctly. To do this analysis, I compute the treatment effect of the Moroccan CCTs assuming 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, the cash transfers are treated as CCTs and are subtracted from the school costs, and if not, they are treated as unconditional and are added to the per-capita income. After estimating the model, I then simulate schooling decisions assuming that everyone understands the conditionality so that the cash transfers are treated as a subsidy to the school costs for everyone, and compute the treatment effect.³⁴ If the confusion explains the success of the across-policies extrapolation, the predictions of the new treatment effect are more accurate via the across-contexts extrapolation than the across-policies extrapolation.

³⁴To obtain counterfactual perceived 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 14: Prediction of treatment effect on school enrollment rates under perfect understanding
Across-contexts (structural) vs Across-policies (structural)

	Estimation		Across-contexts	Across-policies
	(1) Original	(2) Counterfactual	(3) Linear	(4) Linear
Treatment	0.0567*** (0.0106)	0.122*** (0.00836)	0.0431*** (0.00465)	0.0590*** (0.00545)
Control mean	0.894*** (0.00951)	0.868*** (0.00814)	0.702*** (0.00390)	0.941*** (0.00531)
Obs.	4982	4982	4982	4982
Target ITT			0.122	0.122
= Target ITT			0.000	0.000
95% CI of ITT			[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 treatment effect and the control mean.

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

Column (2) in Table 14 displays the counterfactual treatment effect of the Moroccan CCTs. If the conditionality were perfectly understood, then the Moroccan CCTs would have increased the school enrollment rate of the treatment group by 12.2 percentage points from a base of 86.8 percent, which is more than twice as large as the originally estimated effect size reproduced in column (1). Moreover, it is greater than the effect of the Moroccan LCTs by 5 percentage points. However, columns (3) and (4) show that neither of the extrapolations predicts the counterfactual treatment effect accurately, whereas the predicted values of the across-policies extrapolation are still numerically closer to the target estimates. Therefore, removing the degree of confusion about the conditionality among the Moroccan CCTs' sample does not result in better performance of the across-contexts extrapolation. In other words, the poorer performance of the across-contexts extrapolation cannot be solely explained by the confusion levels regarding the conditionality.

6 Conclusion

Predicting the effects of CCT programs on educational outcomes is crucial for policymakers who want to improve domestic education with limited budgets. Provided that the effect size of CCT programs may vary by contexts, the quantification of the effects of CCT programs helps policymakers understand whether the programs are worth implementation. Existing studies have proposed various approaches for the extrapolation of the estimates of policy effects across contexts. The key assumption they make is that conditional on observables, potential outcomes are independent of context characteristics. However, this assumption may not hold, especially when we have to rely on CCT programs implemented in other contexts that are not necessarily comparable. In such a case, an alternative way of predicting effects is to extrapolate from an adjacent policy in the same context that resembles how CCTs affect educational outcomes. Little is known, however, about the differences between the two extrapolations in the ability to make accurate predictions.

This paper empirically studies this question using the cash transfer experiments in Malawi and Morocco. My analysis reveals that the across-policies extrapolation outperforms the across-contexts extrapolation in accurately predicting the treatment effect of the Moroccan CCTs on school enrollment. By interpreting the prediction differences with the parameters in my structural model, I find that the primary driver of the differential predictions is the perceived returns to schooling measured against outside options. Finally, I investigate the sources of variation in the relative returns across the contexts. I find that differences in the observable characteristics of the target populations cannot fully explain the cross-contexts variation of the relative returns. I then suggest that the remaining variation can be attributed to the differences in how the three cash transfer interventions improve school enrollment.

While this paper explores the two extrapolations in the case of CCTs, the intuition that understanding the underlying mechanisms of policies is important for successful out-of-sample predictions applies to other settings. Compared to extrapolation across contexts, however, the predictive power of extrapolation across policies has been 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.³⁵

Another promising research avenue is how to aggregate evidence from various policies to

³⁵Theoretical work on the predictive power of extrapolation across contexts is emerging. For example, Andrews et al. (2022) proposed how to measure the ability of economic models to transfer evidence across contexts.

make predictions. I show that the utility costs of schooling vary similarly under the same policy while the perceived relative returns to schooling do so under the same context. 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 Table G.2 in Appendix G, I find no improvement of the across-policies extrapolation by using the estimated value of θ 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 parameter values when borrowing them from multiple policies in order to make accurate predictions. However, how to do so and what it means in economics are beyond the scope of this paper and require further investigation.

Finally, this paper has implications for policy design. As Hendren and Sprung-Keyser (2020) suggested, policymakers are willing to pay for precise estimates of policy effects.³⁶ Globally, this suggests potential benefits of coordination in policy designs across countries so that extrapolation across contexts can make accurate predictions. Locally, on the other hand, this suggests that a welfare-maximizing policymaker wants to design policies to be informative about future policies. My comparative analysis of the two extrapolations articulates both points empirically in the case of predicting the treatment effects of CCTs on school enrollment.

³⁶Relatedly, Hjort et al. (2021) showed that Brazilian mayors were 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](https://doi.org/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](https://ssrn.com/abstract=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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.3886/E114579V1).
- Caldés, Natàlia, David Coady, and John A. Maluccio.** 2006. “The Cost of Poverty Alleviation Transfer Programs: A Comparative Analysis of Three Programs in Latin America.” *World Development* 34 (5): 818–837. [10.1016/j.worlddev.2005.10.003](https://doi.org/10.1016/j.worlddev.2005.10.003).
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1257/aer.102.4.1241).

- 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.
- García, Sandra, and Juan E. Saavedra.** 2017. “Educational Impacts and Cost-Effectiveness of Conditional Cash Transfer Programs in Developing Countries: A Meta-Analysis.” *Review of Educational Research* 87 (5): 921–965. [10.3102/0034654317723008](#).
- 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](#).
- Kalouptsi, 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](#).

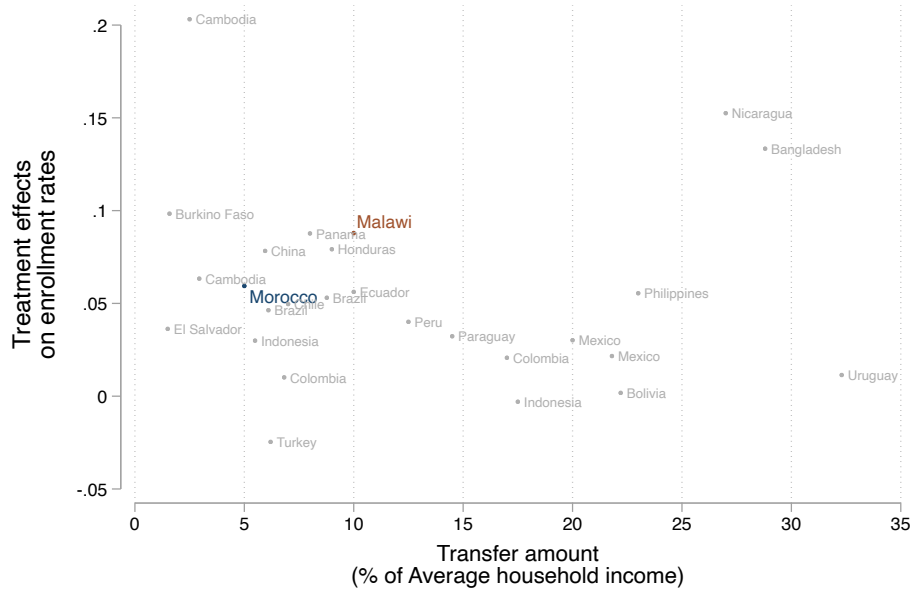
- Rust, John.** 1987. “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher.” *Econometrica* 55 (5): 999. [10.2307/1911259](#).
- Rust, John.** 1994. “Chapter 51 Structural Estimation of Markov Decision Processes.” In *Handbook of Econometrics*, Volume 4. 3081–3143, Elsevier, . [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?”
- Wager, Stefan, and Susan Athey.** 2018. “Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests.” *Journal of the American Statistical Association* 113 (523): 1228–1242. [10.1080/01621459.2017.1319839](#).

Appendix

A 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, which may be more practical for policymakers. Using the information summarized in Baird et al. (2014), I demonstrate the performance of this approach.

Figure A.1: Reported treatment effect estimates in Baird et al. (2014)



Specifically, I make predictions based on a linear relationship between treatment effect estimates and the transfer amount of each program. 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, treatment effects on the odds of schooling, baseline school enrollment rates of control groups, and transfer amount measured as a share of average household income. Then, I recover treatment effects on the school enrollment rates for each program using treatment effects on the odds of schooling and the school enrollment rates of the control groups:

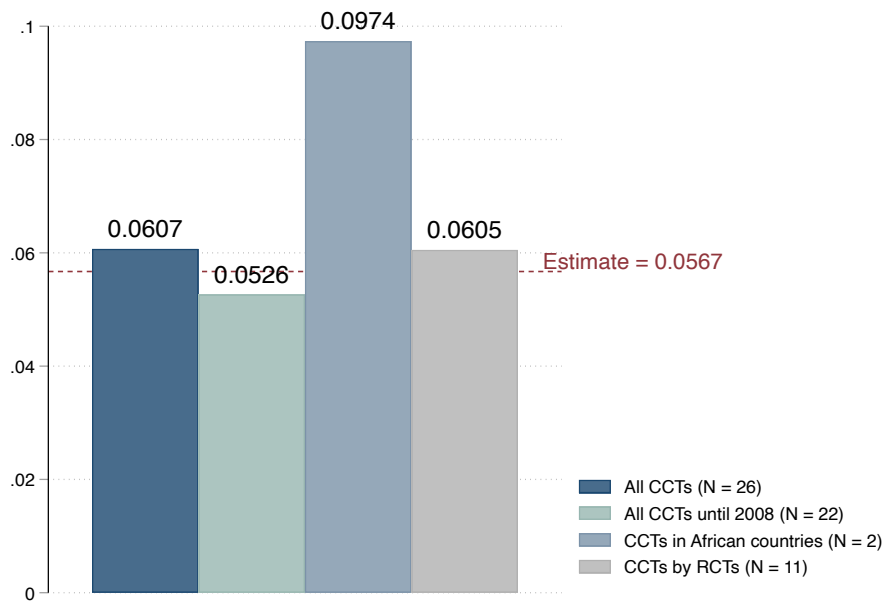
$$TE^{\text{enrollment}} = \frac{TE^{\text{odds}} \times \text{Odds}^{\text{Control}}}{1 + TE^{\text{odds}} \times \text{Odds}^{\text{Control}}} - \text{Enrollment}^{\text{Control}}.$$

Finally, I regress treatment effects on the school enrollment rates on the transfer amount and predict the treatment effect of the Moroccan CCTs with the average ratio of the transfer amount

to household income for the Moroccan CCTs in my data, which is 2.3 percent.

I make several predictions using different sets of CCT programs: (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.^{A.1} Figure A.2 shows the prediction results. While the predictions are numerically close to the estimated effect, they vary across the sample CCTs. To understand what causes the variation and potentially improve the predictions, I focus on a subset of the CCTs that offer more granular data, which motivates my analysis in the main text.

Figure A.2: Simple predictions of effect of Moroccan CCTs on school enrollment rates



^{A.1}The third and fourth predictions are motivated by Vivalt et al. (2022), which showed that policymakers appreciated evidence from similar contexts while researchers valued evidence with internally valid evaluation methods.

B Causal forest for across-contexts extrapolation

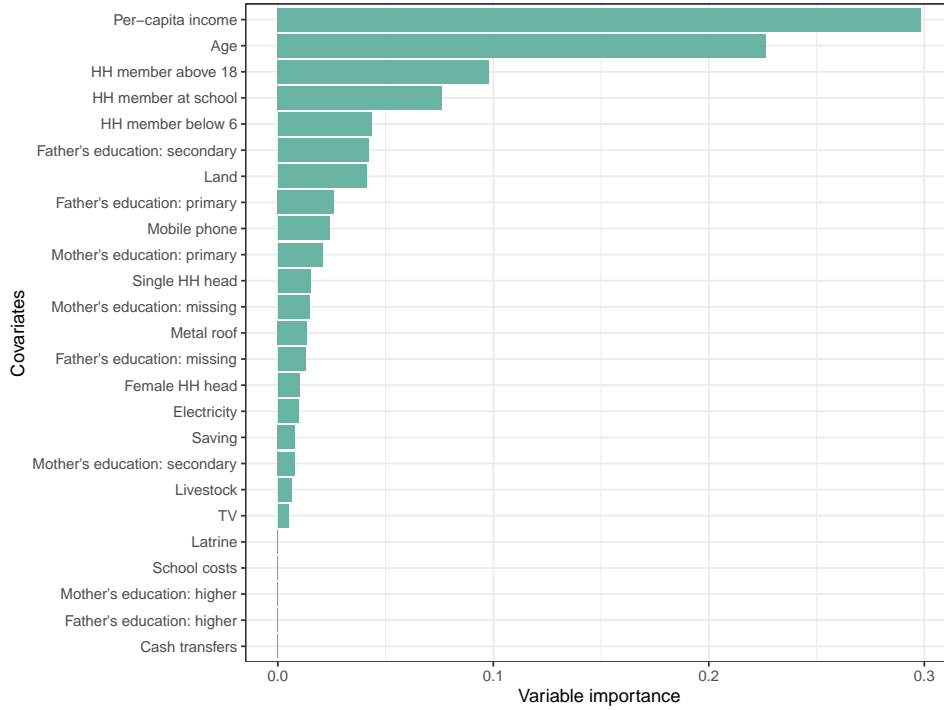
In the reduced-form approaches for the across-contexts extrapolation, I select the set of conditioning variables based on the demographics of the target populations and schooling decisions implied by my structural model. However, a more data-driven way of selecting covariates could improve the performance of the across-contexts extrapolation by identifying hidden moderators of the treatment effect. To check if my results are sensitive to how to select covariates, I predict the treatment effect by using the causal forest algorithm proposed in Wager and Athey (2018). In particular, I first estimate the conditional average treatment effects (CATEs) using the causal forest with the Malawi CCT’s sample, predict the CATEs for the Moroccan CCT’s sample, and aggregate them to obtain the average treatment effect.^{B.1}

To implement the causal forest, I first prepare covariates that are common across the two datasets. In addition to the existing covariates, I include (1) the gender and marital status of household heads, (2) the number of household members younger than 6, older than 18, and at school, parental education levels, (3) whether residence has a metal roof, electricity, and latrines, (4) whether households have a TV, and mobile phone, and (5) whether households have a land, saving technology, and livestock. To identify which covariate moderates the treatment effect, the causal forest splits the sample based on the values of these variables, and estimates the conditional average treatment effects for each subsample to determine whether to grow trees or not. Figure B.1 shows the importance of each covariate, which is computed based on the number of times the causal forest grow trees on that variable.^{B.2}

^{B.1}I use the R package *grf* of version 2.2.1 to implement this. The documentation is available here: <https://grf-labs.github.io/grf/articles/grf.html>.

^{B.2}The number of trees is set at 2000. The distribution of the variable importance remains almost the same when I increase the number of trees to 5000 or 10000.

Figure B.1: Importance of variables for CATEs



After running the causal forest with the Malawi CCT's sample, I predict the treatment effect of the Moroccan CCTs with the predicted CATEs for the Moroccan CCTs' sample. Table B.1 shows that the across-contexts extrapolation predicts the school enrollment rate for the treatment group is statistically significantly higher than the control group by 3.7 percentage points. However, regarding prediction accuracy, the null hypothesis on the equality between the predicted and estimated treatment effects is rejected at a significance level of 0.01, suggesting that the across-contexts extrapolation still mispredicts the target treatment effect. Thus, the conclusion that the across-policies extrapolation outperforms the across-contexts extrapolation remains unchanged.

Table B.1: Across-contexts extrapolation: causal forest

Target: Morocco CCTs	Across-contexts
Treatment	0.037*** (0.000)
Obs.	4982
Target ITT	0.057
= Target ITT	0.000
95% CI of ITT	[0.037, 0.037]

Note: Standard errors in parentheses. I report p-values from F-tests on the null hypothesis that the predicted value is equal to the estimated one.

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

C Choice probabilities estimation

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

The reason why the GMM estimates sometimes can replicate treatment effects when the MLE estimates cannot is because of a trade-off between prioritizing individual choice probabilities versus aggregate shares. GMM directly targets aggregate shares while MLE maximizes the likelihood of individual choices. This trade-off is salient when I estimate the choice probabilities under the Moroccan LCTs because relative changes in income due to the cash transfers are small. Therefore, MLE may not be able to sufficiently differentiate the choice probabilities across the treatment status, leading to imprecise estimates of the treatment effect of the Moroccan LCTs at this stage. This issue can be mitigated by using GMM, although GMM does not necessarily maximize the likelihood of the data. Given this trade-off, I use MLE for the Malawi CCTs and GMM for both of the Moroccan interventions. The replication under the Moroccan CCTs with the GMM estimates is only slightly better than with the MLE estimates. On the other hand, GMM significantly improves the replication under the Moroccan LCTs. The estimates of the two parameters in my structural model estimated either way for all of the interventions are displayed in Table C.2 and Figure C.1.

The flexible logit consists of the second-order polynomials of the state variables, fully interacting with survey round dummy variables. The number of parameters to estimate is 30 for the Malawi intervention and 20 for the two Moroccan interventions. When I estimate the choice probabilities via GMM, I need to have more moment conditions than the number of parameters. Thus I construct as my moments the shares of children choosing schooling across grades, treatment status, and survey rounds. The resulting number of moment conditions is 49 for the Malawi intervention and 24 for the Moroccan interventions.^{C.1} When estimating the choice probabilities 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

^{C.1}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 a small number of children in grades 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).

sample size of each moment as a weighting matrix to prioritize treatment effects for larger subgroups. Finally, I solve the minimization problem using the Nelder-Mead algorithm with the MLE estimates as the starting points. The algorithm always returns the same estimates, although the estimation results are robust to different algorithms. The parameter estimates are presented in Table C.1.

Finally, after estimating the parameters in 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 to 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 P_{it}^1 or $P_{i,t+1}^3$ is numerically 1, then I cannot define the dependent variable and these observations are dropped from my sample of analysis. Since this is more likely to occur among the treated 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 keeping them in the subsequent analysis. Second, without the top-coding, I find $P_{i,t+1}^2$ and $P_{i,t+1}^3$ higher for the treatment group than the control group in all of the interventions while P_{it}^1 is similar across the groups. This is consistent with the treatment effects taking place just after the first round of household surveys. However, I find that $\frac{P_{i,t+1}^2}{1-P_{i,t+1}^3}$ is significantly smaller for the treatment group. This happens because without the top-coding, observations with estimated probabilities extremely close to 1 can have huge values of $\frac{P_{i,t+1}^2}{1-P_{i,t+1}^3}$. For instance, $(P_{i,t+1}^2, P_{i,t+1}^3) = (0.9999, 0.9999)$ has more than 100 times large values than $(P_{i,t+1}^2, P_{i,t+1}^3) = (0.99, 0.99)$, although both can be interpreted as the child almost surely chooses schooling. If these observations are more likely to appear in the control group, then I could have a smaller $\frac{P_{i,t+1}^2}{1-P_{i,t+1}^3}$ for the treatment group. By top-coding the probabilities estimates above 0.99, I do not distort the average of $(P_{it}^1, P_{i,t+1}^2, P_{i,t+1}^3)$ while I have significantly larger values of $\frac{P_{i,t+1}^2}{1-P_{i,t+1}^3}$ for the treatment group.

Table C.1: 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.

Table C.2: Estimates of θ separately for using MLE and GMM estimates

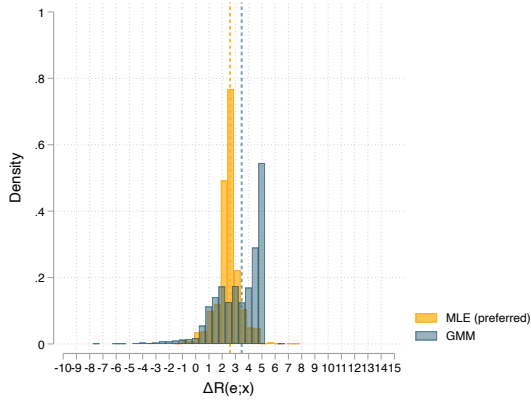
	Malawi		Morocco			
	CCTs		CCTs		LCTs	
	(1)	(2)	(3)	(4)	(5)	(6)
θ	1.008*** (0.256)	0.834*** (0.283)	5.254*** (0.247)	2.670*** (0.454)	7.884 (4.818)	38.90*** (11.30)
Obs.	1479	1479	4981	4981	3016	3016
1st stage F-value	113.011	113.011	3843.510	3843.510	25.483	25.483
CCP estimation	MLE	GMM	MLE	GMM	MLE	GMM

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 Kleibergen-Paap F statistics for weak identification.

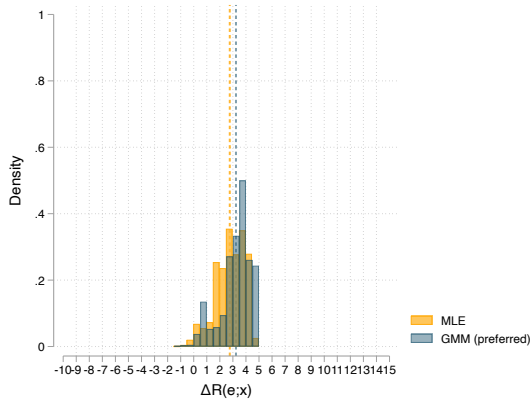
*** p<0.01 ** p<0.05 * p<0.1

Figure C.1: Empirical distribution of $\Delta R(e; x)$ separately for using MLE and GMM estimates

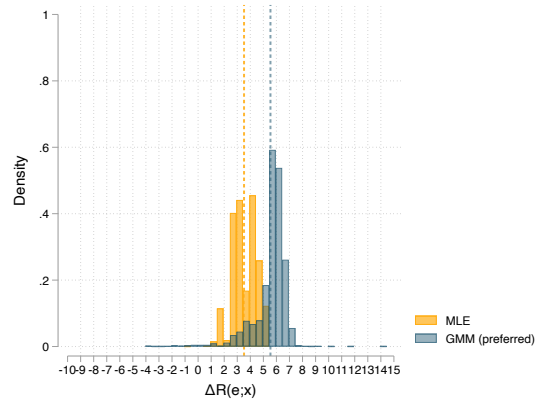
(a) Malawi CCTs



(b) Morocco CCTs



(c) Morocco LCTs



D Fitness of structural model to Moroccan LCTs data

With the parameter estimates in Section 4.2.2, I assess the fitness of my model to the Moroccan LCTs' data based on the simulated treatment effect and the school enrollment rate of the control group. Following the procedures described in Section 3.2.4 but using the data from the Moroccan LCTs, I obtain $\hat{\delta}_1$ and $\hat{\delta}_2$. I then run statistical tests separately:

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

where $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are from the estimation of Equation (5).^{D.1}

Column (3) in Table D.1 presents the predicted values of the treatment effect of the Moroccan LCTs and the school enrollment rate of the control group. The results indicate that the Moroccan LCTs increase the enrollment rate of the treatment group by 5.39 percentage points from a base of 90 percent. The statistical tests show that the null hypotheses about the equality between the predicted and estimated values are not rejected at a significance level of 0.05 for both predictions. These results support the good fitness of my model to the data with 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 with the estimates of the choice probabilities.

I also check how accurately my model can predict individual decisions. Since I choose GMM to estimate the choice probabilities to replicate the treatment effect while sacrificing the maximization of the likelihood of the data, it is possible that the choice probabilities estimates may not be as accurate at the individual level. To empirically investigate this, I simulate individual decisions by drawing preference shocks and compute the share of children with the simulated decisions matched with the actual decisions. Column (3) in Table D.2 shows that my model correctly simulates the individual choices for nearly 90 percent of the Moroccan LCTs' sample. I also check if my model systematically overestimates ($d_{it}^{\text{Data}} = 0, d_{it}^{\text{Model}} = 1$) or underestimates ($d_{it}^{\text{Data}} = 1, d_{it}^{\text{Model}} = 0$) schooling decisions, and find that it makes prediction errors in both directions at similar rates.

For completeness, I also present these two accuracy tests for the models estimated with

^{D.1} Another way of estimating the treatment effects 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 (4). 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.

the Malawi and Moroccan CCTs. Results in Tables D.1 and D.2 suggest that the model is able to replicate the treatment effects as well as individual schooling decisions for those two interventions.

Table D.1: Replication of treatment effect for all interventions

	Malawi	Morocco	
	(1) CCTs	(2) CCTs	(3) LCTs
Treatment	0.0317*** (0.00495)	0.0554*** (0.00777)	0.0539*** (0.00954)
Control mean	0.895*** (0.00241)	0.894*** (0.00747)	0.900*** (0.00889)
Obs.	1490	4982	3018
Target ITT	0.037	0.057	0.073
= Target ITT	0.290	0.869	0.051
95% CI of ITT	[0.022, 0.041]	[0.040, 0.071]	[0.035, 0.073]
Target control mean	0.896	0.894	0.893
= Target control mean	0.721	0.981	0.476
95% CI of control mean	[0.890, 0.899]	[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 treatment effect and the control mean.

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

Table D.2: Model fitness at individual level for all interventions

	Malawi	Morocco	
	(1) CCTs	(2) CCTs	(3) 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 each household survey as the estimation samples. Observations are weighted by sampling weights.

E Reduced-form approaches with normalized variables

As shown in Tables 5 and 6 (also in Table B.1), the predictions of the treatment effect via the across-contexts extrapolation vary by the extrapolation methods, especially for the school enrollment rate of the control group. Specifically, the reduced-form methods predict the treatment effect as 1.56 percentage points from 98.4 percent in the heterogeneous treatment effects approach and null effects from 89.5 percent in the propensity score weighting while the structural methods predict as 4.3 or 4.1 percentage points from 70.2 or 67.6 percent. In Table 10, I also show that the structural model predicts the treatment effect as 1.5 to 1.9 percentage points from 87 percent when adjusting years of education to overlap between the two contexts.

One explanation for the differences between the reduced-form versus structural approaches 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, whereas 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 in the Malawi experiment, ignoring these level differences could lead to divergent predictions. It is thus natural to ask whether the reduced-form approaches produce different predictions with the normalized variables.

Table E.1 shows the performance of the across-contexts extrapolation when normalizing covariates used in the reduced-form methods. Columns (1) and (4) reproduce the results in Table 5. Columns (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 columns (1) and (4), suggesting that the reduced-form and structural approaches yield different predictions not because of defining cash transfer amount and school costs in relative terms. Columns (3) and (6) show results when years of education and children’s age are additionally standardized. 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. What accounts for these differences across the extrapolation methods is, however, beyond the scope of this paper.

Table E.1: Reduced-form extrapolation with normalization of variables

	HTE			PSW		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0156*** (0.000535)	0.0175*** (0.000765)	0.0232*** (0.00327)	0.00660 (0.0184)	-0.0459** (0.0202)	0.0655*** (0.0140)
Control mean	0.984*** (0.000515)	0.979*** (0.000749)	0.899*** (0.00252)	0.895*** (0.0128)	1.000*** (0.00221)	0.920*** (0.0112)
Obs.	4982	4982	4982	1490	1490	1490
Target ITT	0.057	0.057	0.057	0.057	0.057	0.057
= Target ITT	0.000	0.000	0.000	0.007	0.000	0.534
95% CI of ITT	[0.015, 0.017]	[0.016, 0.019]	[0.017, 0.030]	[-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	0.028	0.927	0.000	0.022
95% CI of control mean	[0.983, 0.985]	[0.978, 0.981]	[0.894, 0.904]	[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 the cash transfer amount and the school costs are divided by the 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

F Bias in predictions of school enrollment rates

I discuss how the varying estimates of θ and $\Delta R(e; x)$ across the interventions translate into bias in the predictions of school enrollment rates. I argue that the bias can be attributed to differences in the sizes of the parameter estimates. Specifically, extrapolation with a larger θ than the one under the Moroccan CCTs is 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 always 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, extrapolation with a distribution of $\Delta R(e; x)$ that is on average greater than the one under the Moroccan CCTs is likely to overpredict the perceived relative returns to schooling for both groups. This is because the extrapolation tends to assign larger values for a given year of education than the estimated values. As a result, because the probability of schooling is a function of the relative utility and the relative returns (and is increasing in the sum of the two), the relative size of the parameter values across the interventions can determine the direction of bias in predicting the enrollment rate of the treatment group.

Table F.1 summarizes the above discussion, provided that my parameter estimates reveal the following relationships:

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

and

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

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

In contrast, the direction of the prediction bias for the school enrollment rate of the control group is ambiguous in both extrapolations. Therefore, the discussion based on the average size of the parameter estimates does not determine the direction of prediction bias for the treatment effect of the Moroccan CCTs, which motivates the closer examination of the extrapolated parameter values in Section 5.2.

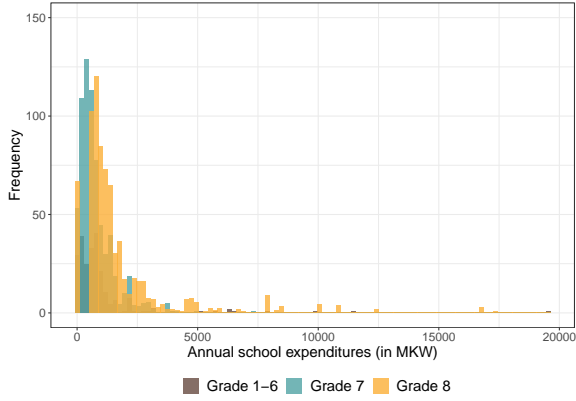
Table F.1: Direction of prediction bias for school enrollment rates

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

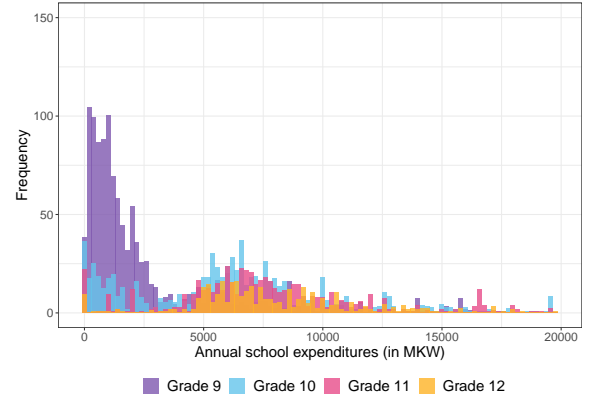
G Supplementary tables and figures

Figure G.1: Distributions of school expenditures

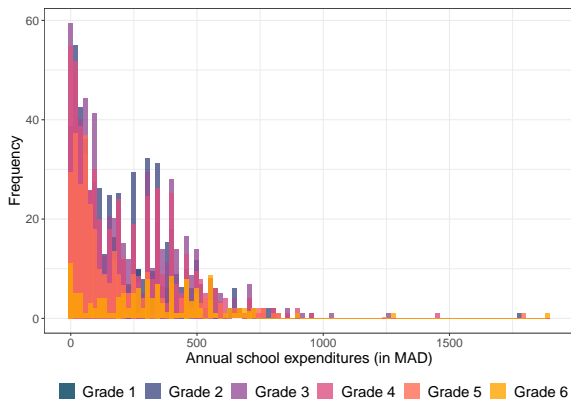
(a) Malawi - Primary school



(b) Malawi - Secondary school



(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 percent). For the Moroccan figure, I drop the outliers with the expenditures higher than 2000 Moroccan Dirham (≈ 0.1 percent).

Table G.1: Robustness of linear extrapolation of $\Delta R(e; x)$

	Across-policies			
	(1) 1st	(2) 2nd (preferred)	(3) 3rd	(4) 4th
Treatment	0.0564*** (0.00550)	0.0590*** (0.00545)	0.0575*** (0.00545)	0.0582*** (0.00547)
Control mean	0.944*** (0.00536)	0.941*** (0.00531)	0.943*** (0.00532)	0.942*** (0.00534)
Obs.	4982	4982	4982	4982
Target ITT	0.057	0.057	0.057	0.057
= Target ITT	0.947	0.674	0.886	0.793
95% CI of ITT	[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
Treatment	0.0132*** (0.000220)	0.0431*** (0.00465)	0.0155*** (0.000972)	0.0365*** (0.0111)
Control mean	0.922*** (0.000177)	0.702*** (0.00390)	0.901*** (0.000855)	0.614*** (0.00927)
Obs.	4982	4982	4982	4982
Target ITT	0.057	0.057	0.057	0.057
= Target ITT	0.000	0.004	0.000	0.070
95% CI of ITT	[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 treatment effect 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

Table G.2: Across-policies extrapolation with replacement of θ and $\Delta R(e; x)$

	Across-policies			
	Linear		RF	
	(1)	(2)	(3)	(4)
Treatment	0.00290*** (0.000175)	0.297*** (0.0190)	0.00301*** (0.000148)	0.297*** (0.0190)
Control mean	0.993*** (0.000157)	0.702*** (0.0184)	0.993*** (0.000134)	0.702*** (0.0184)
Obs.	4982	4982	4982	4982
Replace θ	✓		✓	
Replace $\Delta R(e; x)$		✓		✓
Target ITT	0.057	0.057	0.057	0.057
= Target ITT	0.000	0.000	0.000	0.000
95% CI of ITT	[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 θ and Replace $\Delta R(e; x)$ mean I use the estimated value of each parameter 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 treatment effect and the control mean.

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

Table G.3: Estimate of θ when age is added to state variables under Malawi CCTs

	Malawi CCTs	
	(1)	(2)
θ	1.008*** (0.256)	1.031*** (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 Kleibergen-Paap F statistics for weak identification.

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

Figure G.2: Empirical distribution of $\Delta R(e; x, \text{age})$ under Malawi CCTs

