

Closing the tap: How school enrollment is affected when conditional cash transfer payments break off

Martin Wiegand*

10.03.2018

Abstract

This paper explores how conditional cash transfer programs influence students' schooling decisions when program payments stop in the middle of their school career. To that end, I examine Mexico's PROGRESA, which only covered students until the end of middle school (at age 15) in its early years. The experimental setup permits to study the program's impact on the probability to continue with high school after middle school. However, despite initial randomization, the program itself has likely rendered the respective samples of middle school graduates in the treatment and the control group incomparable. To account for this, I employ a newly developed semiparametric technique that uses a combination of machine learning methods in conjunction with doubly-robust estimation. I find that exposure to PROGRESA during middle school reduced the probability to transfer to high school by 7.5 to 12.5 percentage points. Possible explanations for this effect include parents' loss aversion, motivation crowding, anchoring, and classroom peer effects.

JEL Classification: I22, I25, O15, J24, D04, D91, C52.

Keywords: Education, conditional cash transfer, PROGRESA, machine learning, doubly-robust estimation, loss aversion, motivation crowding, anchoring, classroom peer effects, Mexico.

*Vrije Universiteit Amsterdam, The Netherlands, m.wiegand@vu.nl.

1 Introduction

Around 20 years after their first appearance, conditional cash transfer (CCT) programs for education—initiatives that provide monetary incentives for poor households to send their children to school—have never been more popular. Praised for their potential to increase school enrollment while reducing poverty, they are now widespread in Latin America and are gaining quick traction in Africa and Asia. Even the European Commission has recently launched a CCT program, aiming to encourage around 230,000 refugee children in Turkey to enroll in school (EC Press Release, 16.03.2017). However, a lot of programs do not cover the entirety of a student’s schooldays and instead stop making welfare payments after elementary school or middle school. Even the largest and most well-known CCT programs only started to cover high school students long after their introduction.¹

In this paper, I investigate how CCT programs can affect school enrollment after payments break off, using Mexico’s PROGRESA as a case study. The program is particularly well suited for this research for two reasons. The first is its program design: for the first five years, the program covered students only until the end of middle school (approximately at age 15). The second reason is the richness of the data collected for its evaluation, as well as the experimental setup, which was achieved by the deference of the program in some randomly chosen localities serving as the control group. These features make it possible to estimate the effect of PROGRESA on high school enrollment. I focus on two quantities of interest. The first is the overall treatment effect on the probability of going to high school. It takes all children into account who had finished primary school just before the program started (around the age of 12). The resulting number is easy to interpret, but may largely be driven by the program impact on middle school enrollment. The second quantity of interest is the effect on the probability to transition from middle school to high school. More precisely, I consider the control group students who finished middle school—despite not receiving the program—and compute how their likelihood of continuing school afterwards would differ had they been exposed to the program. This latter term is a more immediate measure of program aftereffects, and thus receives the main

¹This includes the four largest CCTs (in terms of beneficiaries) at the time of writing, namely Brazil’s BOLSA FAMILIA, Mexico’s PROSPERA (formerly known as PROGRESA), the Philippines’ PANTAWID, and Colombia’s FAMILIAS EN ACCIÓN. Further examples of currently running CCTs that do not cover high school are Indonesia’s PROGRAM KELUARGA HARAPAN, Ghana’s LIVELIHOOD EMPOWERMENT AGAINST POVERTY program, Pakistan’s PUNJAB FEMALE SCHOOL STIPEND PROGRAMME, and CCT pilots in Nigeria and Burkina Faso.

focus of this paper.

There are many studies showing that CCT programs increase school enrollment while payments are in place (see e.g. Bastagli et al. 2016), but to my knowledge there is none about the aftereffects on enrollment. The direction of such effects is not obvious: neoclassical theory suggests that past payments free up resources, rendering continued schooling more likely. On the other hand, a number of theories from psychology and behavioral economics, such as loss aversion, motivation crowding, anchoring, and classroom peer effects, could explain why the probability to continue school might actually decrease due to earlier payments. Studying the aftereffects of CCT programs is necessary to understand their full impact on the education distribution, and is highly relevant for the design of future programs: a policy maker with limited funds needs to worry less about early break-offs if CCT programs continue to have a positive effect on enrollment. If, on the other hand, it turns out that CCTs actively discourage students from continued education, this would be a strong case in favor of continuing funding until the end of the school career.

Despite the random selection of PROGRESA communities into treatment and control group, education payments have likely changed the composition of middle school graduates between these groups after two years. Some students in the treatment group may not have finished middle school if it had not been for the payments. So to obtain causal estimates, I control for a large number of baseline characteristics. I employ a newly developed procedure that combines machine learning (ML) methods and treatment effects estimation, based on work by Chernozhukov et al. (2018). The idea is to use ML to learn the relationships of treatment status as well as outcomes with potential confounders. The residuals of these relationships are then used to obtain treatment effects via orthogonal moment conditions. The procedure permits to capture complex functional relationships and to control for large sets of confounders—possibly more than there are observations—without having to know either in advance. This approach relies on far fewer assumptions than conventional propensity score methods, and makes results more credible in comparison.

The estimation results lead to an interesting finding: having been paid for schooling in the past reduces the probability to continue once the payments stop. It appears that paying students up to a point actively discourages them to stay in school afterwards. This effect does not seem to spill over to children from households that were not eligible for the program. For these children, the enrollment effect of the program may even be positive.

The paper proceeds as follows. In section 2, I offer a number of explanations for aftereffects of CCT programs on school enrollment. Section 3 sums up the relevant details about the program and the data used. In section 4, I explain the identification strategy. In section 5, the ML methods used are being explained and the results are presented. The findings are further expanded on and discussed in section 6. Section 7 concludes.

2 Why CCT programs might have aftereffects

The effects of PROGRESA on school enrollment are documented in a number of papers. The studies by Schulz (2004), Behrman et al. (2005), and Behrman et al. (2009) all find that children in the treatment group stay in school with higher probability than those in the control group, and that this effect is particularly pronounced for children in the age for middle school. Todd and Wolpin (2006) and Attanasio et al. (2012) develop and estimate structural models of decision making about child schooling, enabling them to evaluate the costs and benefits of alternative program specifications. Both papers conclude that a shift of program resources from primary school age to middle school age children would have led to a higher increase of total completed years of schooling. Similar to this paper, the studies by Behrman et al. (2005) and Dubois et al. (2012) estimate students' transition probabilities from one grade to another. The study by Behrman et al. (2005) is the only one that considers the probability to go to high school, however without controlling for compositional changes as a result of prior program exposure, and only for small subsets of the data. Dubois et al. (2012) aim to disentangle the effects of PROGRESA on grade repetition and continuation. They circumvent the selection problem by looking only at the first year of implementation, and find among other things that middle school students in the treatment group are more likely to repeat a grade. The authors speculate that this may reflect the incentive to stay in middle school longer due to the limited program coverage. No study so far has concentrated on the schooling impacts of PROGRESA (or any other CCT program) after payments stop, perhaps because it may not be obvious why the decision to continue school should depend on having been paid to go to school before. In the following, I offer some explanations of how this may come about.

One of the goals of CCT programs is helping poor households to finance children's education. If financial constraints are in fact the main driver of educa-

tional underinvestment, then easing these constraints by making cash payments should result in more schooling. For instance, a family may have saved just enough to allow their child to finish middle school. Giving transfer payments until that point may then enable the household to save more, which in turn might allow the child to go to high school. The study by de Janvry et al. (2006) supports the argument that CCT payments can help smoothing out spending on education. It finds that PROGRESA takes a safety net function, in that it protects children from the impacts of shocks on school enrollment. The smoothing of education spending may not only work intertemporally within households, but also between households. Angelucci et al. (2010) show that PROGRESA raises middle school enrollment only for children with large family networks, in which transfers from better-off family members go to children on the margin of enrollment.

While CCT programs allow to save more money for future education and thus might facilitate high school enrollment, a number of insights from psychology and behavioral economics point in the opposite direction. In the following, I highlight four reasons why CCT programs could discourage further schooling after they end: loss aversion, motivation crowding, anchoring, and classroom peer effects.

Loss aversion: Loss aversion is a central feature of prospect theory (Kahneman and Tversky 1979; Tversky and Kahneman 1991). It means that from a psychological point of reference, losses loom larger than gains of equal size. For intertemporal choice problems, this means that people require a larger payment to postpone present consumption than the amount they are willing to pay to speed up consumption (Loewenstein 1988). In the context of CCT programs, the choice between working (more consumption now) and continuing school (more consumption later) may depend on whether going to school is framed as a loss or a forgone gain of current consumption. Since reference points are often derived from past levels of consumption, families who received PROGRESA payments may frame the choice as having less or the same current consumption as before. On the other hand, families who never received PROGRESA payments may perceive the choice as having either the same or higher levels of consumption, making them more likely to choose more education. Weiss et al. (1980) document loss aversion in the domain of schooling (without using this term) in the context of the Seattle-Denver income maintenance experiment. They find that while reducing the relatively low direct costs of schooling led to a large increase in enrollment among young adults, significantly reducing the much higher

opportunity cost of going to school had no such effect.

Motivation crowding: Another possibly relevant theory from behavioral economics is motivation crowding theory (Frey and Jegen 2001; Fehr and Falk 2002). It acknowledges that people’s actions are often motivated by hope for social approval, a desire to be moral, or intrinsic interest. When monetary incentives are added, they can replace those motives. A famous example is given by Gneezy and Rustichini (2000), who show that introducing a fine for parents who are late to fetch their children from kindergarten makes them arrive even later. The explanation is that being late, which used to be the violation of an ethical norm before, is being reframed into a good that can be bought for a reasonable price. Importantly, removing the fine did not make the parents arrive earlier again. In the same way, PROGRESA may put a price tag on the moral obligation to let children go to school. In distinction to the aforementioned experiment, the price of non-conformance is high enough to comply with the program. But nonetheless, sending the children to school will have a price tag, which may reduce the pressure to continue after the payments stop.

Anchoring: If financial constraints were in fact the only reason for educational underinvestment, there would be no reason to make transfer payments *conditional* on school attendance. Instead, an unconditional cash transfer could achieve the same result without the need to monitor compliance, and free from the often raised criticism that CCT programs are paternalistic. One reason for conditionality is that children as well as parents may be poorly informed about the returns to education, or about the natural talent required to complete school (Fiszbein and Schady 2010). For instance, Nguyen (2008) shows that households in Madagascar lack information about returns to education but change decisions rationally when this information is updated. Jensen (2010) shows that eight-graders in the Dominican Republic massively underestimate the rate of return to secondary school. And Dizon-Ross (2017) finds that parents in Malawi hold inaccurate beliefs about their children’s ability, the more so when they have low education themselves, and that they misallocate resources to education accordingly. In this light, making cash transfer programs conditional may be a way to deal with such kinds of misinformation of parents and children, by nudging them into a higher level of educational investment. But in doing so, CCT programs may also convey a signal about the value of education: if the government is willing to pay for it, it must be worth pursuing. Conversely, the drop in payments after middle school may suggest that subsidizing poor students to go to high school is not worth it—be it due to low marginal returns

to schooling at this level, or because students from poor families are deemed unlikely to succeed there. Research in neuroeconomics shows that a higher price is often associated with a higher value (see e.g. Plassmann et al. 2008). By first anchoring the value of schooling to the PROGRESA payouts and then reducing it to zero could make further education less desirable.

Classroom peer effects: If a CCT program works as intended, some students keep attending school who would not have done so in the absence of the program. Presumably, this alters the overall composition of students. For instance, it is conceivable that these additional students are on average less talented, motivated, or are relatively more burdened with work outside of school. A higher rate of bad students might in turn have negative spillovers on the motivation and learning outcomes of the other students and lead to more misbehavior in class, as the literature on classroom peer effects suggests (e.g. Lavy et al. 2012, Carrell and Hoekstra 2011). Thus, by the end of middle school, some students may have lost their ability or motivation to continue with high school.

Last of all, I want to mention the possibility of spillover effects within localities. Most CCT programs only target the poorest households, but they may affect students from other households nonetheless. For the case of PROGRESA, such spillover effects have been documented: Angelucci and De Giorgi (2009) argue that due to inter-household risk-sharing, food consumption increases even for non-eligible households in PROGRESA treatment villages. And Attanasio et al. (2012) find, for their sample of boys between 10 and 16, that school enrollment was substantially higher for the non-eligible children in the treatment group than for those in the control group. If households within a village share program resources, one would expect that spillover effects on high school enrollment take the same direction as for the eligible students. If enrollment rises as a result of increased savings, this effect might well spread across household networks to non-eligible students. And also if the program has aftereffects through a change in the composition of the classroom, or a change of social norms towards going to school, it seems most intuitive to expect that these channels would move all students—eligible or not—in the same direction.

3 Program and data description

PROGRESA is a multi-component antipoverty program and was started in 1997. Its original goal was to improve prior antipoverty programs in Mexico along

a number of dimensions, such as increasing targeting efficiency and reducing administrative costs (Ganther 2009). At first, a limited number of rural localities were selected for inclusion in the program. Localities had to have between 50 and 2,500 inhabitants, access to health and education services, and had to be considered highly deprived based on available census data. Households from selected localities were then classified as poor or not poor based on baseline survey data. Only households classified as poor were eligible.

One declared target was to increase school attendance of children in poor families. The education component of PROGRESA included bimonthly cash transfers to mothers of every child enrolled in grades 3 to 9 who attended at least 85 percent of classes. This includes the last four years of primary school (*primaria*) and all of middle school (*secundaria*) but not high school (*preparatoria*). Payments increased with the age of the child to adjust for the increasing opportunity costs of schooling due to higher child wages. However, according to Schultz (2004), these payments were still lower than the average value of full-time child labor. In 2001, the program was renamed OPORTUNIDADES and extended to urban areas, and in 2002 the schooling grants were extended to include the high school level. It should be noted that the continuation of the program after the presidential election in 2000 was highly uncertain, so that a student's delay of high school in the expectation of later program coverage is unlikely.

For evaluation purposes, eligible localities were randomized into two groups. Payments for the treatment group started in May 1998, while payments for the control group only started in December 1999. The evaluation sample includes 320 treatment localities and 186 control localities. Survey data was collected biannually for all households of the evaluation set from 1997 to 2000. Two surveys were administered before the program started in the treatment group (October 1997 and March 1998), three between the start of the program for the treatment and the control group (October 1998, March 1998, November 1999), and two after the program had started for the control group (March 2000 and November 2000). In 2003, another survey was carried out that included a new comparison group of households from localities which had not been included in the program until that point.

To identify the effect that PROGRESA had on the transition to high school, I consider those students who were expected to start high school in the academic year of 2000/01. By the end of the term in July 2000, the eligible students in the treatment group have benefited from PROGRESA for more than two years. Those

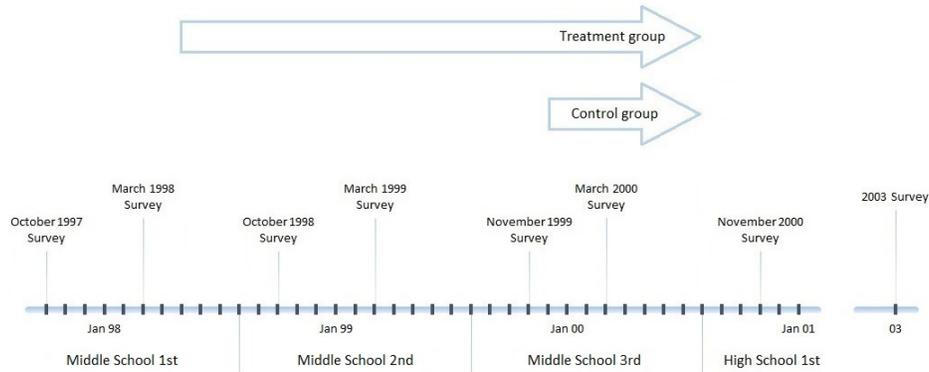


Figure 1: Timeline of relevant events around PROGRESA

in the control group will only have been exposed to the program for the last semester of middle school, when the decision to continue school afterwards has likely been made already.² Figure 1 depicts a timeline with all survey dates. The two arrows indicate how long the students from the cohort under discussion were exposed to the program, for the treatment and the control group, respectively.

I consider three outcome variables related to high school attendance. The first is whether the student went to high school at the time of the November 2000 survey, i.e., right after finishing middle school. The second variable is whether the student had ever been to high school by the time of the 2003 survey. The third variable is whether the student had completed high school by 2003 or was enrolled in the last grade. The last two variables are particularly useful to check the medium-term impact on high school enrollment. After all, it could be that any differences in high school enrollment at the end of middle school fade out after a while. This may happen for instance if parents and students do eventually overcome any potential behavioral anomalies due to the program and start realizing their full education potential. Or it might be that those in the treatment group had actually formed expectations about an inclusion of high school students in the future and thus simply postponed enrollment by a little.

²One could also consider the students who were expected to start high school in the academic year 1999/2000. The eligible students in the treatment group of this older cohort would thus be exposed to the program for one year, and the students in the control group not at all. However, unfortunately, for this cohort it is very hard to unambiguously determine the set of students having finished middle school in 1999, and whether these students continued to high school afterwards. This is due to the fact that some of the relevant questions do not appear in the corresponding survey rounds.

Table 1: Descriptive statistics of outcome variables

	(1)				(2)					
	Obs.	Mean	SE	Diff.	Miss.	Obs.	Mean	SE	Diff.	Miss.
<i>Children from eligible households</i>										
Started high school in 2000	T	797	0.112	0.011	-0.017	185	0.298	0.023	-0.067	0
	C	452	0.128	0.016	[0.381]	73	0.365	0.032	[0.080]	0
Some high school by 2003	T	551	0.324	0.028	-0.077	431	0.321	0.025	-0.067	70
	C	278	0.247	0.018	[0.019]	247	0.388	0.034	[0.111]	32
Completed high school by 2003	T	551	0.178	0.016	-0.049	431	0.257	0.024	-0.052	70
	C	278	0.227	0.025	[0.094]	247	0.308	0.031	[0.192]	32
<i>Children from non-eligible households</i>										
Started high school in 2000	T	417	0.228	0.021	0.111	163	0.384	0.030	0.070	0
	C	273	0.117	0.020	[0.000]	88	0.314	0.033	[0.115]	0
Some high school by 2003	T	326	0.350	0.026	0.087	254	0.436	0.034	0.119	53
	C	194	0.263	0.032	[0.040]	167	0.317	0.036	[0.018]	40
Completed high school by 2003	T	326	0.301	0.025	0.120	254	0.372	0.033	0.128	53
	C	194	0.180	0.028	[0.002]	167	0.244	0.034	[0.008]	40

The columns under (1) include all children of age 11-14 who completed the last year of primary school in 1997. The columns under (2) include those children who were in their last year of middle school in school year 1999/2000. The square brackets contain p -values for a two-sided t -test of the hypothesis that $\text{mean}(\text{treatment group}) - \text{mean}(\text{control group}) = 0$. Missing observations include those inconsistent with respect to gender, age, or highest completed school degree.

Table 1 shows descriptive statistics of the three outcome variables by eligibility and treatment and control group. The columns under (1) summarize the sample of all children who completed primary school in 1997 and who were between 11 and 14 years old at that time. Assuming no middle grade repetitions, these students could have started high school in 2000. I refer to this sample of students as the *unconditional sample*. In the columns under (2), only those children are summarized who finished the last year of middle school in 2000. I refer to this sample of students as the *conditional sample*, since it will be used to compute program treatment effects conditionally on having finished middle school.³ The column labeled *Miss.* reports the number of missing observations for the respective variable and experimental group. These children either do not appear in the later surveys or have inconsistent information between the first two surveys and the November 2000 survey or 2003 survey, respectively.⁴ I split both the unconditional and the conditional sample into two subsamples—eligible and ineligible—and run all estimations separately in order to obtain both the direct program effects as well as spillover effects.

Since it is to be expected that the composition of students who finish middle school differs between treatment and control group, I construct a large number of exogenous characteristics to balance the two groups. They originate from the two pre-treatment surveys, and include such things as demographic and socioeconomic characteristics of the household, parents' level of education, parents' assessment of the student's ability and expectations about future educational outcomes, parents' assessment of teacher and school quality, village characteristics, and travel times to a number of educational institutions. Using high-dimensional econometric techniques allows me to include a large number of potentially relevant characteristics without having to know in advance which

³In fact, this variable is not included as such in any questionnaire and is constructed as follows: I take all the children who reported the last year of middle school as their highest completed grade in the November 2000 survey, and who also reported being enrolled in school in the March 2000 survey. These conditions do not, in principle, rule out the possibility of a student having finished middle school before and then, in school year 1999/00, attempted another grade of further education but did not finish it. So, as an alternative to the second criterion (going to school in March 2000), I consider those children who reported the second year of middle school as their highest completed grade in the November 1999 survey (a question not included in the March 2000 survey). This way, it is not clear whether those who repeated the last grade of middle school are included. In addition, some observations get lost in the process of merging of survey rounds, so that this way of constructing the variable leads to considerably fewer observations to analyze. That is why I stick to the first choice. I did, however, redo the analysis with the alternative construction. The very similar results are available upon request.

⁴Observations with a sex change, age inconsistency, or diminishing highest school degree are set to missing.

of them is actually correlated with treatment status and high school enrollment. A list of all the considered characteristics is included in the appendix.

4 Identification Strategy

For student $i = 1, \dots, n$, let D_i be an indicator variable for living in a treatment locality and Y_i an outcome variable, e.g. an indicator for going to high school after middle school. $Y_i(1)$ and $Y_i(0)$ denote potential outcomes, $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$. X_i is a p -dimensional vector of exogenous control variables.

In the following, I distinguish between unconditional and conditional treatment effects, by which I mean the treatment effects for the correspondent samples, respectively. For the unconditional sample, the statistic of interest is the average treatment effect (ATE),

$$\text{ATE} := \mathbb{E}[Y_i(1) - Y_i(0)]. \quad (1)$$

For the conditional sample, the focus lies on those students who would have finished middle school even without PROGRESA. This is because the program itself has likely added some students to the pool of middle school graduates in the treatment group, who do not have a counterpart in the control group. On the other hand, it is inconceivable that a student who finishes middle school in the absence of payments would not have done so in their presence. To use the parlance of the literature on local average treatment effects: the analysis is concentrated on the *always-takers*, who by virtue of the experimental setup should be fully represented in both groups. It aims to leave out the *compliers*, as their counterfactual is not observed, as well as the *defiers*, who are nonexistent by assumption. To achieve this, I consider the average treatment effect of the non-treated (ATN),

$$\text{ATN} := \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 0]. \quad (2)$$

which is based on the distribution of students in the control group.

If the students under consideration were sampled into treatment and control group at random, the ATE would simply be identified by the difference in sample means between treatment and control group, or average predictive effect (APE),

$$\text{APE} := \mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0]. \quad (3)$$

The APEs of the program are in fact displayed in Table 1, under the columns labeled *Diff*.

There are two reasons why the APE may not be an unbiased estimator of equation (1), despite the initial randomization of households. The first reason is sample selection, a concern mainly for the conditional sample: the children who completed middle school in 2000 may have been influenced in their decision to do so by the program. Therefore, it is likely that at this stage, the treatment group is composed differently than the control group. For instance, it may be that some students in the treatment group only finished middle school because of PROGRESA. These students may be comparatively less likely to continue to high school, and thus create the false impression that the program has a negative effect on high school enrollment.

The second reason for possible bias is attrition. In the unconditional sample, around 21% of the children identified in the two pre-treatment surveys are missing in the November 2000 survey, and around 45% of them are missing in the 2003 survey. In the conditional sample, around 16% of the children are missing in the 2003 survey. Attrition becomes a problem when it does not occur at random. For instance, if those students not going to high school are more likely to leave their original household and disappear from the sample, considering only those who stay would lead to a downward bias.

I address these two concerns separately, starting with sample selection. In order to obtain unbiased treatment effects conditional on finishing middle school, I assume that treatment is independent of outcomes conditionally on control variables X_i ,

$$Y_i(1), Y_i(0) \perp D_i \mid X_i. \quad (4)$$

Under this assumption, Rosenbaum and Rubin (1983) famously showed that it is sufficient to condition on the propensity score instead of the whole vector of controls. There are, however, some limitations commonly associated with this approach. The researcher needs to know exactly which variables to condition on, as well as the functional form of the probability model. Economic intuition may be helpful for model selection up to a point. But despite best efforts, seemingly relevant features may nonetheless lead to overfitted propensity scores, while seemingly unrelated variables may hold a lot of predictive power through correlations with important (possibly unobserved) features. In addition, the established methods require low model complexity for identification—i.e., $p \ll n$ —even in cases where a large number of confounders is

plausible. Consequently, there is little insurance against misspecification of the probability model, which calls the unconfoundedness assumption and the propensity score method into question.

A recent literature in econometrics tries to remedy these shortcomings and to bring treatment effects estimation and machine learning techniques together (see among others Athey et al. 2016, Belloni et al. 2014, 2017, Chernozhukov et al. 2018, Farrell 2015). For this paper, I use the specification and estimation strategy taken in Chernozhukov et al. (2018). To formalize the relationship between D_i , Y_i , and X_i , consider the model

$$Y_i = g_0(D_i, X_i) + U_i, \quad \text{E}[U_i | D_i, X_i] = 0, \quad (5)$$

$$D_i = m_0(X_i) + V_i, \quad \text{E}[V_i | X_i] = 0. \quad (6)$$

This specification is quite general in that it allows for heterogeneous treatment effects and does not require D_i and X_i to be additively separable in the regression function $g_0(D_i, X_i)$. $m_0(X_i)$ is the propensity score, i.e., the conditional probability to be in the treatment group. The ATE is given by

$$\theta := \text{E}[g_0(1, X_i) - g_0(0, X_i)], \quad (7)$$

and the ATN by

$$\gamma := \text{E}[g_0(1, X_i) - g_0(0, X_i) | D_i = 0], \quad (8)$$

Belloni et al. (2014, 2017) point out that in a high-dimensional parameter space, directly estimating equation (5) using sophisticated machine learning methods is ill-advised. While doing so may result in a great fit of Y_i , this approach neglects how treatment assignment is affected by covariates, potentially resulting in a large regularization bias. One way to overcome this bias, as suggested in Belloni et al. (2017), Chernozhukov et al. (2018), and Farrell (2015), is *double machine learning*. The idea is to estimate the nuisance functions $g_0(D_i, X_i)$ and $m_0(X_i)$ separately using ML methods. The parameters of interest are then identified using moment conditions based on scores that are Neyman-orthogonal, which means that they are insensitive to the small deviations that come about by replacing the nuisance functions with their estimates. Chernozhukov et al. (2018) show that a Neyman-orthogonal score function for

the ATE is

$$\begin{aligned} \varphi(W_i; \theta, \eta) = & g(1, X_i) - g(0, X_i) + \frac{D_i(Y_i - g(1, X_i))}{m(X_i)} \\ & - \frac{(1 - D_i)(Y_i - g(0, X_i))}{1 - m(X_i)} - \theta, \end{aligned} \quad (9)$$

with $W_i = (Y_i, D_i, X_i)$ and nuisance functions $\eta(X_i) = (g(0, X_i), g(1, X_i), m(X_i))$. A Neyman-orthogonal score function for the ATN is

$$\begin{aligned} \varphi(W_i; \gamma, \eta) = & \frac{D_i(1 - m(X_i))(Y_i - g(1, X_i))}{m(X_i)(1 - p_D)} - \frac{(1 - D_i)(Y_i - g(0, X_i))}{1 - p_D} \\ & + \frac{(1 - D_i)(g(1, X_i) - g(0, X_i))}{1 - p_D} - \gamma \frac{1 - D_i}{1 - p_D}, \end{aligned} \quad (10)$$

with nuisance functions $\eta = (g(0, X_i), g(1, X_i), m(X_i), p_D)$, where $p_D := \mathbb{E}[D_i]$. A condition for \sqrt{n} -consistency of these estimators is that all nuisance functions are estimable at a rate of at least $o(n^{-1/4})$. While this condition is not trivially satisfied, the possibility to aggregate or choose the best out of multiple ML methods guarantees estimability for a wide range of structures.⁵

Another crucial part of this method is cross-fitting to prevent bias induced by overfitting. For a fixed integer K , the sample is randomly split into folds I_1, \dots, I_K of equal size. For each $k \in \{1, \dots, K\}$, the nuisance functions are estimated using only the observations outside of I_k . The resulting functional estimates are then used to predict $g_0(D_i, X_i)$ and $m_0(X_i)$ in fold I_k , which are in turn used to obtain point estimates and variances of θ and γ from each fold. These are then averaged to obtain the estimators $\hat{\theta}$ and $\hat{\gamma}$, which are approximately unbiased and asymptotically normally distributed. Nonetheless, the sample-splitting procedure introduces additional uncertainty. Therefore, the above procedure is repeated a number of times B with different random splits. Chernozhukov et al. (2018) recommend to take the median of estimates for each split as the final estimator:

$$\hat{\theta}^{\text{median}} = \text{median} \left\{ \hat{\theta}_b \right\}_{b=1}^B, \quad (11)$$

with a conservative variance estimator which takes into account the variation

⁵An alternative approach to deal with regularization bias is discussed in Athey et al. (2016). It does not require estimability of the propensity score, but in turn limits the complexity of the regression function by assuming strong sparsity.

introduced by sample splitting,

$$\hat{\sigma}^{2,\text{median}} = \text{median} \left\{ \hat{\sigma}_b^2 + \left(\hat{\theta}_b - \hat{\theta}^{\text{median}} \right)^2 \right\}_{b=1}^B. \quad (12)$$

Having discussed how the problem of nonrandom sample selection is approached, I now turn to nonrandom attrition. Let R_i be an indicator variable for remaining in the sample, thus taking the value 1 if Y_i is non-missing and 0 otherwise. I make an assumption similar to unconfoundedness, namely that attrition is independent of outcomes conditional on treatment status D_i and control variables X_i ,

$$Y_i(1), Y_i(0) \perp R_i \mid (D_i, X_i). \quad (13)$$

So, while attrition on its own may be predictive of outcomes, this predictive power comes entirely from observable variables. The approach is similar to the one taken in Behrmann et al. (2009) on the medium-term effects of PROGRESA, where attrition from the 2003 survey is also assumed to be random conditionally on a (small) number of observables and treatment status. I propose the following extension of the model above to accommodate this assumption:

$$R_i = r_0(D_i, X_i) + W_i, \quad \text{E}[W_i \mid D_i, X_i] = 0. \quad (14)$$

$r_0(D_i, X_i)$ is the conditional probability that student i 's outcome is observed. R_i does not enter the regression function g_0 , but $r_0(D_i, X_i)$ is needed for reweighting in the moment conditions. The corresponding score functions are for the ATE

$$\begin{aligned} \varphi(W_i; \theta, \eta) = & g(1, X_i) - g(0, X_i) + \frac{R_i D_i (Y_i - g(1, X_i))}{r(1, X_i) m(X_i)} \\ & - \frac{R_i (1 - D_i) (Y_i - g(0, X_i))}{r(0, X_i) (1 - m(X_i))} - \theta, \end{aligned} \quad (15)$$

with nuisance functions $\eta(X_i) = (g(0, X_i), g(1, X_i), m(X_i), r(0, X_i), r(1, X_i))$, and for the ATN

$$\begin{aligned} \varphi(W_i; \gamma, \eta) = & \frac{R_i D_i (1 - m(X_i)) (Y_i - g(1, X_i))}{r(1, X_i) m(X_i) (1 - p_D)} - \frac{R_i (1 - D_i) (Y_i - g(0, X_i))}{r(0, X_i) (1 - p_D)} \\ & + \frac{(1 - D_i) (g(1, X_i) - g(0, X_i))}{1 - p_D} - \gamma \frac{1 - D_i}{1 - p_D}, \end{aligned} \quad (16)$$

with nuisance functions $\eta(X_i) = (g(0, X_i), g(1, X_i), m(X_i), r(0, X_i), r(1, X_i), p_D)$. These score functions both are zero in expectation for the true values of nuisance functions and parameters, and they both fulfill the Neyman-orthogonality property.

5 Estimation

I estimate the ATE and the ATN using 10-fold cross-fitting with 100 repetitions. For the separate estimation of the nuisance functions $g_0(0, X_i)$, $g_0(1, X_i)$, $m_0(X_i)$, $r(0, X_i)$, and $r(1, X_i)$, I use six different ML methods, as well as an ensemble method that combines all of them, and a technique that combines each of the previously mentioned methods that works best for each nuisance function. The first three ML techniques are regularized logistic regression techniques, namely the Lasso (with ℓ_1 penalty), Ridge (with ℓ_2 penalty), and elastic net (with both ℓ_1 and ℓ_2 penalty). Furthermore, I use two tree-based techniques—namely random forest and gradient boosting with logistic regression trees—as well as support vector machines (SVM).⁶ The optimal tuning parameters for each of these methods are obtained via 10-fold cross validation, repeated 10 times. For the ensemble technique, first all the previously mentioned techniques are used to predict the nuisance functions. To avoid overfitting, for each step k in the original cross-fitting procedure, predictions are made using another—nested—cross-fit with 10 folds on the observations outside I_k .⁷ These predictions are then again used in a linear regression model to obtain the best combination of all methods, with the constraint that coefficients are non-negative and sum to one. For the last technique, I choose for each target function the ML method that produces the smallest out-of-sample mean squared error.

For the random forest, boosted trees, and SVM, the dictionary of considered controls encompasses all the variables listed in the appendix, with categorical variables expanded into dummy variables. For the regularized regression tech-

⁶I also tried out neural networks with one hidden layer, but chose to leave them out eventually. They either showed very poor predictive performance or—for an appropriate amount of hidden nodes—turned out too computationally expensive.

⁷Note that due to this nested cross-fitting, the observations that are available to tune the ML methods for the ensemble technique are slightly fewer than those for the other techniques: on average, 81% instead of 90% of all observations are being used at a time. Therefore, to find the optimal tuning parameters for the ensemble method, I use a repeated 5-fold cross validation on the whole dataset (thus using 80% of all observations each time) to match the observations at hand.

Table 2: Estimates of ATE and ATN on high school education

Dependent variable	Effect	eligible		non-eligible	
		uncond.	cond.	uncond.	cond.
Started high school in 2000	ATE	-0.013 (0.022)	-0.099** (0.050)	0.112*** (0.037)	0.036 (0.055)
	ATN		-0.125** (0.060)		0.043 (0.061)
Some high school by 2003	ATE	-0.105** (0.044)	-0.105* (0.055)	0.095* (0.050)	0.087 (0.065)
	ATN		-0.100** (0.048)		0.069 (0.069)
Completed high school by 2003	ATE	-0.055 (0.035)	-0.090* (0.046)	0.119** (0.047)	0.111** (0.055)
	ATN		-0.075* (0.043)		0.092* (0.056)
Observations		1,507	646	941	475

*: significant at the 10% level. **: significant at the 5% level. ***: significant at the 1% level. Average treatment effect (ATE) and average treatment effect of the non-treated (ATN) as well as standard errors (in parentheses, obtained via wild bootstrap) are estimated for the unconditional sample (all children having completed primary school in 1997) and for the conditional sample (all children who completed middle school in 2000), and separately for children from eligible and non-eligible households. They are estimated for three outcome variables: whether the student continued with high school in school year 2000/01, whether the student had attended some high school by the end of 2003, and whether the student had finished or was about to finish high school by the end of 2003. All estimates are obtained via the respective orthogonal score functions that take into account attrition, (15) and (16), using the median method. Results are based on 10-fold estimation using 100 sample splits. Only results from combining the best-fitting ML methods for each nuisance function are reported. Outcomes for all the ML methods are reported in the appendix.

niques, I use an extended set of candidate variables. Next to the variables already mentioned, it includes squared and cubed terms of all numerical variables, and cubic B-splines with five interior knots of three continuous variables—two household poverty indices and a village-level poverty index. Last of all, it includes interactions of all the previously mentioned variables with a subset of 28 variables that I deem particularly relevant; these include student characteristics, household demographics, parents’ education and expectations, local wages paid to children, and poverty levels. Missing values are treated as follows: for categorical variables, a new *missing* category is created. For numerical variables, missing entries are assigned the average of all non-missing entries, and an additional *missing* dummy is created. Gradient boosting is the only method which does not require missing value imputation. After dropping duplicates and perfectly collinear variables, the basic dictionary of variables for the eligible students in the conditional sample includes 239 variables, whereas the extended dictionary includes 12,508 variables. The sets of variables for the non-eligible students, as well as for eligible and non-eligible students in the unconditional sample, are very similar in magnitude.

To exclude extreme values for the propensity score and to guarantee overlap between treatment and control group, I apply a trimming procedure that is solely based on the propensity score (for details see Crump et al. 2009, Imbens and Rubin 2015). It produces an interval $(\alpha, 1 - \alpha)$ such that all observations with propensity scores outside this interval are discarded. α turns out to be between 0.05 and 0.1 in all cases.

For the unconditional sample, I compute the ATE for the three high school indicators discussed above as outcomes. For the conditional sample, I compute both the ATE and ATN. Following Belloni et al. (2017), I use Mammen’s (1993) wild bootstrap method to obtain standard errors, using 10,000 bootstrap replications. The results from combining the nuisance function estimates of the respective best-fitting ML methods (i.e., my preferred specification) are reported in Table 2. More detailed results, with estimates for each of the ML methods used, are depicted in Tables 6, 7, 8, and 9 in the appendix.

6 Discussion

I start by discussing the eligible students, since they are the main focus of this paper. Looking at the first column of Table 2, it appears that the program

did not have a statistically or economically significant effect on timely high school enrollment for the overall student population. However, it seems that by 2003, the program made it less likely by about 10 percentage points for students to have enrolled in high school at some point. Looking at the third outcome, it is unclear whether this translated into lower high school completion rates for treated students. It is a somewhat puzzling result that students who had just completed primary school when the program started would not have higher eventual high school continuation rates. This is especially so since it is known that PROGRESA had a positive effect on middle school enrollment and grade progression. Therefore, it must be that the program had a negative effect on the continuation decisions of middle school graduates. This hypothesis is confirmed when looking at the second column. The ATN for high school in 2000 is -12.5 percentage points. This effect shrinks to -7.5 percentage points when high school completion by 2003 is considered, but it stays (marginally) statistically significant.

For the non-eligible students, the third column of Table 2 shows that being in the treatment group increased high school enrollment and graduation by 10 to 12 percentage points. Knowing about the positive effects on middle school enrollment and the spillover effects on the non-eligible students, this result is not particularly surprising. One would, however, not expect that it is driven by the program's direct effect on middle school graduates. Looking at the fourth column, this expectation is partly fulfilled in that the program effect on immediate high school continuation is insignificant and, while still positive, close to zero. But the ATN on high school completion by 2003 is 9 percentage points and statistically significant. This result is unexpected, since it implies a sort of reverse spillover effect: those who saw their peers getting paid became more likely to continue schooling, while those actually getting paid lost interest. There are nonetheless ways to rationalize this result. One explanation may be peer effects: an increase in eligible low ability students in middle school as a result of the program may reduce learning outcomes for other low ability students, while it may strengthen the relative position of high ability students (of which many may be non-eligible) and thus heighten their self-esteem and motivation. An influx of poor students in middle school may also increase the need of non-poor students and their parents to distinguish themselves through further education. It is also conceivable that seeing their peers getting paid triggers ineligible students' will to demonstrate their capability despite being in a relative disadvantage.

Table 3: Comparing parents' assessment of student's ability

	Obs.	Mean	Diff.	<i>p</i> -value
<i>Children from eligible households</i>				
Student is good at school	507	0.617	0.004	0.933
Student can make it at least to high school	635	0.310	0.065	0.092
Student can make it to university	635	0.099	0.052	0.037
<i>Children from non-eligible households</i>				
Student is good at school	377	0.634	-0.005	0.915
Student can make it at least to high school	464	0.360	-0.029	0.521
Student can make it to university	464	0.159	0.011	0.757

Based on the conditional sample. The column *Diff.* displays the difference between means in the treatment group and the control group.

The precise channels for the observed treatment effects are impossible to learn with certainty from the data at hand. Nonetheless, I examine two of the aforementioned possible causes, namely classroom peer effects and loss aversion, and check whether they constitute credible explanations for the observed effects. Starting with classroom peer effects: if it is true that PROGRESA worsens the pool of middle school students, one would expect this to show in measures of performance. Unfortunately, such measures are not available until the 2003 survey. However, parents' assessments are available from the pre-treatment survey of March 1998. I consider three binary variables: whether (according to the parents) the student is good at school, whether the student is apt enough to go to high school or further, and whether the student is apt enough to go to university. While these assessments are certainly very noisy signals of a student's ability, there is no obvious reason to believe they should not at least convey some information thereon. Table 3 shows the differences in averages of the assessment variables between the treatment and control group, both for eligible and non-eligible students. Judging by this table, it does not become obvious that the students who finished middle school in the treatment group should be less apt than the middle school graduates in the control group. If anything, the numbers suggest the opposite. Thus, the theory of negative classroom peer effects is not supported by the data.

Next, in order to examine loss aversion as a possible explanation, I check whether financial concerns are responsible for the differences in high school en-

rollment between treatment and control group. To that end, I look at a question asking for the reasons why students did not go to school in the November 2000 survey. Of the possible answers to this question, three point to financial constraints, namely: (1) there is not enough money to send the student to school, (2) the student is needed for work, and (3) the student is needed at home. I lump these together in an indicator variable that is 1 if one of these three reasons were given and 0 otherwise. In the same fashion, I create another indicator that lumps together all the other reasons why a student might not go to high school. The most frequent reasons here include: (1) the student does not like school, (2) the student attends a vocational school, (3) the student is already grown up, and (4) the school is too far away. Yet another variable related to financial constraints may be actual household expenditures. Therefore, I also construct a variable of monthly per capita expenditure, using prices and quantities of goods as indicated in the section on consumption and expenditure from the November 2000 survey.

I use these three variables—student does not go to high school for money reasons, student does not go to high school for reasons other than money, and monthly per capita expenditure—as outcomes and apply the same model as before on the conditional sample. This way it should be possible to see whether or not the program effects can be explained through intermediate effects on household finances. In particular, if loss aversion is the main driver of the negative treatment effect for the eligible students, this might show in a higher share of students not going to school for money reasons and higher per capita expenditures in the treatment group. Table 4 sums up the results, with more detailed outputs in Tables 10 and 11 in the appendix.

The results show that not going to high school for money reasons as well as per capita expenditure seem to be nearly unaffected by the program. On the other hand, the program does seem to disincline eligible students from high school for other reasons than money, whereas it has the opposite effect on non-eligible students. This result indicates that loss aversion may not be the main explanation for the negative treatment effect on eligible students. Instead, it seems more likely that the program impacts enrollment through other factors such as students' motivation or social norms.

Table 4: Estimates of ATE and ATN on further outcomes

Dependent variable	Effect	eligible	non-eligible
Not going to high school for money reasons	ATE	-0.019 (0.041)	0.050 (0.058)
	ATN	-0.007 (0.048)	0.018 (0.060)
Not going to high school for reasons other than money	ATE	0.118*** (0.046)	-0.087** (0.043)
	ATN	0.118** (0.053)	-0.083* (0.049)
Monthly per capita expenditure (in Mexican pesos)	ATE	-2.712 (12.362)	2.688 (24.539)
	ATN	-0.572 (11.148)	12.935 (32.423)
Observations		646	475

*: significant at the 10% level. **: significant at the 5% level. ***: significant at the 1% level. Average treatment effect (ATE) and average treatment effect of the non-treated (ATN) as well as standard errors (in parentheses, obtained via wild bootstrap) are estimated separately for children from eligible and non-eligible households. They are estimated for three outcome variables: whether the student did not continue with high school in school year 2000/01 for money-related reasons, whether the student did not continue with high school in school year 2000/01 for reasons other than money, and monthly per capita expenditure in the student's household. All estimates are obtained via the respective orthogonal score functions (9) and (10), using the median method. Results are based on 10-fold estimation using 100 sample splits. Only results from combining the best-fitting ML methods for each nuisance function are reported. Outcomes for all the ML methods are reported in the appendix.

7 Conclusion

The positive effects of CCT programs like PROGRESA on school enrollment have been demonstrated in numerous studies. However, surprisingly, their aftereffects have not been explored so far. With this paper, I try to fill this gap by estimating PROGRESA's impact on the probability to continue school after program payments stop. The main finding is that for the eligible students, the program has large and significant negative effects. There are a number of possible explanations. Monetary incentives may crowd out the social norm of sending children to school or reduce the intrinsic motivation to attend school regularly once they stop being in place. Establishing program payments and then reducing them to zero again may convey the false signal that education is not worth it at the later levels. Payments may shift parents' income reference point such that the sudden drop needs to be compensated by the child's wage income. And a change in the composition of students induced by the program may lead to negative classroom peer effects. Though conclusive evidence in favor of one over the other explanations is lacking, it seems that loss aversion and classroom peer effects are not much supported by the data, leaving motivation crowding and anchoring as the remaining candidates.

The paper also looks at possible spillovers to the students who were not eligible to the program but lived in treatment villages. Curiously, it seems that—if anything—these students are more likely to finish high school as a result of their peers getting paid. This could be explained by increased self-esteem or heightened desire to separate themselves in the face of a change in classroom composition. But the inconsistency across different outcome variables suggests that this finding might as well be a fluke.

Of course, the findings of this paper are confined to the relatively short time the program had been in effect by the year 2000. For younger cohorts who start middle school with the program already in place, one may expect positive unconditional treatment effects, via the intermediate positive effect the program has on middle school enrollment. On the other hand, the conditional treatment effects might be even more extreme due to a longer exposure to the program.

The main result is striking, since it points to a case where monetary incentives have very costly unintended side-effects. Moreover, the finding should also be considered in the program design of future CCT programs. Even in cases where coverage on all school levels is not feasible due to budgetary constraints, there may be ways to counter the adverse program effect. This could

for instance be done by systematically informing students and parents about the marginal rate of return to continued education. Another way may be to let go of the conditionality of payments altogether, particularly to counter possible crowding-out effects and anchoring. Further research that systematically explores these channels can help to find ways to reduce the effect.

References

- Angelucci, M. and G. De Giorgi (2009). Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? *American Economic Review* 99(1), 486–508.
- Athey, S., G. W. Imbens, and S. Wager (2016). Approximate residual balancing: De-biased inference of average treatment effects in high dimensions.
- Attanasio, O. P. and K. M. Kaufmann (2014). Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics* 109, 203 – 216.
- Attanasio, O. P., C. Meghir, and A. Santiago (2012). Education choices in mexico: Using a structural model and a randomized experiment to evaluate *progesa*. *The Review of Economic Studies* 79(1), 37–66.
- Bastagli, F., J. Hagen-Zanker, L. Harman, V. Barca, G. Sturge, T. Schmidt, and L. Pellerano (2016). Cash transfers: What does the evidence say? a rigorous review of programme impact and of the role of design and implementation features.
- Behrman, J. R., S. W. Parker, and P. E. Todd (2009). 7 medium-term impacts of the oportunidades conditional cash transfer program on rural youth in mexico. *Poverty, Inequality, and Policy in Latin America*, 219.
- Behrman, J. R., P. Sengupta, and P. Todd (2005). Progressing through *progesa*: An impact assessment of a school subsidy experiment in rural mexico. *Economic Development and Cultural Change* 54(1), 237–275.
- Belloni, A., V. Chernozhukov, I. Fernandez-Val, and C. Hansen (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica* 85(1), 233–298.

- Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies* 81(2), 608–650.
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21(1), C1–C68.
- Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* 96(1), 187–199.
- de Janvry, A., F. Finan, E. Sadoulet, and R. Vakis (2006). Can conditional cash transfer programs serve as safety nets in keeping children at school and from working when exposed to shocks? *Journal of Development Economics* 79(2), 349 – 373. Special Issue in honor of Pranab Bardhan.
- Dubois, P., A. de Janvry, and E. Sadoulet (2012). Effects on school enrollment and performance of a conditional cash transfer program in Mexico. *Journal of Labor Economics* 30(3), 555–589.
- Farrell, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics* 189(1), 1 – 23.
- Fehr, E. and A. Falk (2002). Psychological Foundations of Incentives. IEW - Working Papers 095, Institute for Empirical Research in Economics - University of Zurich.
- Fiszbein, A. and N. R. Schady (2010). *Conditional Cash Transfers*. The World Bank.
- Frey, B. S. and R. Jegen (2001). Motivation crowding theory. *Journal of Economic Surveys* 15(5), 589–611.
- Gantner, L., S. E. Frandsen, A. Kuyvenhoven, and J. von Braun (2009). *PROGRESA: An Integrated Approach to Poverty Alleviation in Mexico (5-1)*, pp. 211–220. Cornell University Press.
- Gneezy, U. and A. Rustichini (2000). A fine is a price. *The Journal of Legal Studies* 29(1), 1–17.

- Imbens, G. W. and D. B. Rubin (2015). *Trimming to Improve Balance in Covariate Distributions*, pp. 359–374. Cambridge University Press.
- Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics* 125(2), 515–548.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives* 5(1), 193–206.
- Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47(2), 263–291.
- Loewenstein, G. F. (1988). Frames of mind in intertemporal choice. *Manage. Sci.* 34(2), 200–214.
- Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics* 21(1), 255–285.
- Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from madagascar. *Unpublished manuscript* 6.
- Plassmann, H., J. O’Doherty, B. Shiv, and A. Rangel (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences* 105(3), 1050–1054.
- Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41–55.
- Schultz, T. P. (2004). School subsidies for the poor: evaluating the mexican progresá poverty program. *Journal of Development Economics* 74(1), 199 – 250. *New Research on Education in Developing Economies*.
- Todd, P. E. and K. 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. *The American Economic Review* 96(5), 1384–1417.
- Tversky, A. and D. Kahneman (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics* 106(4), 1039–1061.
- Weiss, Y., A. Hall, and F. Dong (1980). The effect of price and income on investment in schooling. *Journal of Human Resources* 15(4), 611–640.

Appendix

Appendix Table 1: List of potential control variables

Variable description	Type
<i>Student and household characteristics</i>	
Student is female	binary
Age of student in 1997	continuous
Student attended school in March 1998	binary
Degree of poverty index in 1997 (by 1997 criteria)	continuous
Degree of poverty index in 1997 (by 2003 criteria)	continuous
Number of household members below age 15	count
Father lives in the household	binary
Mother lives in the household	binary
Father is literate	binary
Mother is literate	binary
Father went to school	binary
Mother went to school	binary
Father finished at least primary school	binary
Mother finished at least primary school	binary
Father finished at least middle school	binary
Mother finished at least middle school	binary
<i>Parents' assessments and opinions</i>	
Parents say student is good at school in 1998	binary
Parents say student is able to finish middle school	binary
Parents say student is able to continue after middle school	binary
Parents say student can finish high school	binary
Parents say student is able to finish university	binary
Desired level of schooling for girls is at least middle school	binary
Desired level of schooling for girls is more than middle school	binary
Desired level of schooling for girls is at least high school	binary
Desired level of schooling for girls is university	binary
Desired level of schooling for boys is at least middle school	binary
Desired level of schooling for boys is more than middle school	binary

Variable description	Type
Desired level of schooling for boys is at least high school	binary
Desired level of schooling for boys is university	binary
Children eat breakfast before school	binary
Reason why children don't eat breakfast before school	categorical
Parent talked to teacher this year	binary
Reason for talk with teacher	categorical
Parent participates in parent / guardian association of school	binary
Parent participates in school work	binary
In school, there are problems with lack of discipline	binary
In school, there are problems with lack of interest of the teachers	binary
In school, there are problems with poor communication between teachers and parents	binary
In school, there are problems with poor teacher attendance	binary
The teacher is usually prepared	binary
The teacher is usually fulfilled	binary
The teacher is usually on time	binary
The teacher is usually patient with the children	binary
Age from which girls can help younger siblings	continuous
Age from which boys can help younger siblings	continuous
Age from which girls can help with work	continuous
Age from which boys can help with work	continuous
Age from which girls can work to earn money	continuous
Age from which boys can work to earn money	continuous
<i>Household expenditures</i>	
Weekly expenditures for public transport to school	continuous
Weekly expenditures for public transport for other trips	continuous
Weekly expenditures for cigarettes and tobacco	continuous
Weekly expenditures for alcoholic beverages	continuous
Weekly expenditures for nonalcoholic beverages	continuous
Monthly expenditures for hygiene items	continuous
Monthly expenditures for medicine	continuous
Monthly expenditures for medical consultations	continuous
Biannual expenditures for household articles	continuous
Biannual expenditures for toys	continuous

Variable description	Type
Biannual expenditures for girls' clothes	continuous
Biannual expenditures for boys' clothes	continuous
Biannual expenditures for women's clothes	continuous
Biannual expenditures for men's clothes	continuous
Biannual expenditures for girls' shoes	continuous
Biannual expenditures for boys' shoes	continuous
Biannual expenditures for women's shoes	continuous
Biannual expenditures for men's shoes	continuous
Biannual expenditures for school supplies	continuous
Biannual expenditures for school contributions	continuous
If family had more money, they would spend it on food	rank
If family had more money, they would spend it on housing repairs	rank
If family had more money, they would spend it on clothing or shoes	rank
If family had more money, they would spend it on debt settlement	rank
If family had more money, they would spend it on animals	rank
If family had more money, they would spend it on seeds or plants	rank
If family had more money, they would spend it on work tools	rank
If family had more money, they would spend it on medicine	rank
If family had more money, they would spend it on school supplies	rank
If family had more money, they would save it	rank
<i>Location characteristics</i>	
Marginality index	continuous
Degree of marginality very high (1) or high (0) in 1997	binary
Village is indigenous	binary
Village has a municipal delegate	binary
Village has a municipal subdelegate	binary
Village has a commissioner of agricultural land	binary
Village has a commissioner of communal goods	binary
Village has a municipal development committee	binary
Village has a health committee	binary
Village has a education committee	binary
Village has a agricultural committee	binary
Village has a DICONSA store officer	binary
Village has a production cooperative	binary

Variable description	Type
Village has religious organizations	binary
Village has political organizations	binary
Village has a school parent association	binary
Village has community assemblies	binary
Village has NGOs	binary
Village has a communal work system (tequio)	binary
Source of water	categorical
Type of garbage disposal	categorical
Electricity available everywhere	binary
Electricity at least partly available	binary
Public drainage at least partly available	binary
Public phone available	binary
Number of preschools	count
Number of primary schools	count
Number of distance middle schools	count
Most important sector in this village	categorical
Second most important sector in this village	categorical
Third most important sector in this village	categorical
Child labor takes place in this village	binary
Average daily salary paid to children	continuous
Travel time (minutes) to nearest private middle school	continuous
Travel time (minutes) to nearest public middle school	continuous
Travel time (minutes) to nearest distance middle school	continuous
Travel time (minutes) to nearest middle school	continuous
Travel time (minutes) to nearest private high school	continuous
Travel time (minutes) to nearest public high school	continuous
Travel time (minutes) to nearest high school	continuous
Travel time (minutes) to nearest national college of technical professional education	continuous
Travel time (minutes) to nearest agricultural technological center	continuous
Travel time (minutes) to nearest industrial technology and services center	continuous
Travel time (minutes) to nearest agricultural college	continuous
Travel time (minutes) to nearest industrial and services college	continuous

Table 6: Detailed estimates of conditional ATE and ATN on high school education for children from eligible households

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Ensemble	Best Method
<i>(1) Dependent variable: going to high school in November 2000.</i>								
ATE	-0.082* (0.047)	-0.066 (0.046)	-0.080* (0.045)	-0.085** (0.036)	-0.079 (0.050)	-0.084* (0.050)	-0.103** (0.052)	-0.099** (0.050)
ATN	-0.090* (0.053)	-0.073 (0.050)	-0.083 (0.051)	-0.093** (0.039)	-0.086 (0.060)	-0.100* (0.056)	-0.121** (0.061)	-0.125** (0.060)
<i>(2) Dependent variable: some high school by winter 2003.</i>								
ATE	-0.089* (0.053)	-0.069 (0.059)	-0.086* (0.050)	-0.113** (0.052)	-0.130 (0.083)	-0.082 (0.057)	-0.108** (0.055)	-0.105* (0.055)
ATN	-0.103** (0.050)	-0.073 (0.058)	-0.092* (0.048)	-0.084 (0.054)	-0.137 (0.089)	-0.085 (0.052)	-0.111** (0.050)	-0.100** (0.048)
<i>(3) Dependent variable: completed high school by winter 2003.</i>								
ATE	-0.077 (0.050)	-0.058 (0.054)	-0.069 (0.048)	-0.099* (0.055)	-0.132 (0.082)	-0.064 (0.053)	-0.088* (0.050)	-0.090* (0.046)
ATN	-0.077* (0.044)	-0.068 (0.051)	-0.074* (0.044)	-0.062 (0.052)	-0.066 (0.073)	-0.070 (0.047)	-0.069 (0.043)	-0.075* (0.043)

*: significant at the 10% level. **: significant at the 5% level. ***: significant at the 1% level. Average treatment effect (ATE) and average treatment effect of the non-treated (ATN) as well as standard errors (in parentheses, obtained via wild bootstrap) are estimated for children from eligible households. They are estimated for three outcome variables: whether the student continued with high school in school year 2000/01, whether the student had attended some high school by the end of 2003, and whether the student had finished or was about to finish high school by the end of 2003. The estimates for the first outcome are obtained via the orthogonal estimating equations (9) and (10), while their respective modifications for attrition (15) and (16) are employed for the other two outcomes. All estimates use the median method. The Results are based on 10-fold estimation using 100 sample splits. Column labels denote the method used to estimate nuisance functions.

Table 7: Detailed estimates of conditional ATE and ATN on high school education for children from non-eligible households

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Ensemble	Best Method
<i>(1) Dependent variable: going to high school in November 2000.</i>								
ATE	0.075* (0.044)	0.075 (0.048)	0.058 (0.052)	0.054 (0.042)	0.045 (0.056)	0.076 (0.065)	0.033 (0.056)	0.036 (0.055)
ATN	0.091* (0.047)	0.089 (0.056)	0.073 (0.063)	0.055 (0.046)	0.075 (0.062)	0.082 (0.072)	0.039 (0.063)	0.043 (0.061)
<i>(1) Dependent variable: some high school by winter 2003.</i>								
ATE	0.138** (0.060)	0.128** (0.060)	0.137** (0.059)	0.091* (0.055)	0.061 (0.080)	0.114* (0.060)	0.099* (0.060)	0.087 (0.065)
ATN	0.142** (0.060)	0.120* (0.070)	0.137** (0.066)	0.110* (0.060)	-0.004 (0.117)	0.097 (0.064)	0.061 (0.076)	0.069 (0.069)
<i>(3) Dependent variable: completed high school by winter 2003.</i>								
ATE	0.159*** (0.052)	0.140** (0.055)	0.157*** (0.057)	0.094 (0.059)	0.112 (0.074)	0.127** (0.056)	0.112** (0.054)	0.111** (0.055)
ATN	0.166** (0.068)	0.134** (0.062)	0.169** (0.076)	0.113** (0.055)	0.097 (0.085)	0.107* (0.058)	0.089* (0.051)	0.092* (0.056)

*: significant at the 10% level. **: significant at the 5% level. ***: significant at the 1% level. Average treatment effect (ATE) and average treatment effect of the non-treated (ATN) as well as standard errors (in parentheses, obtained via wild bootstrap) are estimated for children from non-eligible households. They are estimated for three outcome variables: whether the student continued with high school in school year 2000/01, whether the student had attended some high school by the end of 2003, and whether the student had finished or was about to finish high school by the end of 2003. The estimates for the first outcome are obtained via the orthogonal estimating equations (9) and (10), while their respective modifications for attrition (15) and (16) are employed for the other two outcomes. All estimates use the median method. The Results are based on 10-fold estimation using 100 sample splits. Column labels denote the method used to estimate nuisance functions.

Table 8: Detailed estimates of unconditional ATE on high school education for children from eligible households

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best Method
<i>(1) Dependent variable: going to high school in November 2000.</i>							
ATE	-0.024 (0.024)	-0.012 (0.022)	-0.023 (0.021)	-0.021 (0.018)	-0.059 (0.048)	-0.006 (0.026)	-0.013 (0.022)
<i>(2) Dependent variable: some high school by winter 2003.</i>							
ATE	-0.133** (0.060)	-0.078* (0.046)	-0.128** (0.057)	-0.088*** (0.033)	-0.082* (0.045)	-0.096** (0.041)	-0.105** (0.044)
<i>(3) Dependent variable: completed high school by winter 2003.</i>							
ATE	-0.054 (0.036)	-0.051* (0.031)	-0.053* (0.029)	-0.054* (0.030)	-0.047 (0.039)	-0.045 (0.036)	-0.055 (0.035)

*: significant at the 10% level. **: significant at the 5% level. ***: significant at the 1% level.
Average treatment effect (ATE) as well as standard errors (in parentheses, obtained via wild bootstrap) are estimated for children from eligible households. They are estimated for three outcome variables: whether the student continued with high school in school year 2000/01, whether the student had attended some high school by the end of 2003, and whether the student had finished or was about to finish high school by the end of 2003. All estimates are obtained via the orthogonal score function that takes into account attrition (15), using the median method. Results are based on 10-fold estimation using 100 sample splits. Column labels denote the method used to estimate nuisance functions.

Table 9: Detailed estimates of unconditional ATE on high school education for children from non-eligible households

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best Method
<i>(1) Dependent variable: going to high school in November 2000.</i>							
ATE	0.127*** (0.028)	0.125*** (0.030)	0.124*** (0.028)	0.124*** (0.024)	0.121*** (0.044)	0.117*** (0.031)	0.112*** (0.037)
<i>(2) Dependent variable: some high school by winter 2003.</i>							
ATE	0.115** (0.048)	0.092** (0.045)	0.101** (0.044)	0.098** (0.041)	0.042 (0.076)	0.095** (0.047)	0.095* (0.050)
<i>(3) Dependent variable: completed high school by winter 2003.</i>							
ATE	0.124*** (0.040)	0.124*** (0.041)	0.114*** (0.040)	0.116*** (0.039)	0.088* (0.053)	0.119*** (0.042)	0.119** (0.047)

*: significant at the 10% level. **: significant at the 5% level. ***: significant at the 1% level.
Average treatment effect (ATE) as well as standard errors (in parentheses, obtained via wild bootstrap) are estimated for children from non-eligible households. They are estimated for three outcome variables: whether the student continued with high school in school year 2000/01, whether the student had attended some high school by the end of 2003, and whether the student had finished or was about to finish high school by the end of 2003. All estimates are obtained via the orthogonal score function that takes into account attrition (15), using the median method. Results are based on 10-fold estimation using 100 sample splits. Column labels denote the method used to estimate nuisance functions.

Table 10: Detailed estimates of conditional ATE and ATN on further outcomes for children from eligible households

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Ensemble	Best Method
<i>(1) Dependent variable: not going to high school for money reasons.</i>								
ATE	-0.011 (0.050)	-0.014 (0.043)	-0.009 (0.051)	-0.019 (0.038)	-0.020 (0.044)	-0.029 (0.055)	-0.019 (0.044)	-0.019 (0.041)
ATN	0.016 (0.059)	-0.003 (0.045)	0.013 (0.058)	0.007 (0.042)	0.011 (0.050)	-0.022 (0.067)	0.003 (0.047)	-0.007 (0.048)
<i>(2) Dependent variable: not going to high school for reasons other than money.</i>								
ATE	0.090** (0.043)	0.095*** (0.036)	0.091** (0.041)	0.098*** (0.033)	0.098** (0.048)	0.107** (0.044)	0.116*** (0.045)	0.118*** (0.046)
ATN	0.082 (0.060)	0.119*** (0.045)	0.081 (0.053)	0.092** (0.041)	0.065 (0.072)	0.109* (0.066)	0.123** (0.058)	0.118** (0.053)
<i>(3) Dependent variable: monthly per capita expenditure.</i>								
ATE	-8.069 (15.018)	38821.820 (105879.300)	-8.743 (2836.579)	-5.153 (11.265)	-14.690 (15.241)	-21.415 (22.892)	1.252 (15.176)	-2.712 (12.362)
ATN	-6.047 (14.370)	-6720.530 (34812.970)	-5.991 (15.170)	-4.378 (12.892)	-4.439 (12.292)	-8.765 (16.073)	-0.884 (11.172)	-0.572 (11.148)

*: significant at the 10% level. **: significant at the 5% level. ***: significant at the 1% level. Average treatment effect (ATE) and average treatment effect of the non-treated (ATN) as well as standard errors (in parentheses, obtained via wild bootstrap) are estimated for children from eligible households. They are estimated for three outcome variables: whether the student continued with high school in school year 2000/01, whether the student had attended some high school by the end of 2003, and whether the student had finished or was about to finish high school by the end of 2003. The estimates for the first outcome are obtained via the orthogonal estimating equations (9) and (10), while their respective modifications for attrition (15) and (16) are employed for the other two outcomes. All estimates use the median method. The Results are based on 10-fold estimation using 100 sample splits. Column labels denote the method used to estimate nuisance functions.

Table 11: Detailed estimates of conditional ATE and ATN on further outcomes for children from non-eligible households

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Ensemble	Best Method
<i>(1) Dependent variable: not going to high school for money reasons. N = 475.</i>								
ATE	0.014 (0.045)	0.002 (0.052)	0.016 (0.054)	0.025 (0.044)	0.014 (0.055)	0.020 (0.064)	0.043 (0.058)	0.050 (0.058)
ATN	0.008 (0.048)	0.001 (0.058)	-0.011 (0.069)	0.024 (0.049)	-0.042 (0.076)	0.027 (0.073)	0.011 (0.062)	0.018 (0.060)
<i>(2) Dependent variable: not going to high school for reasons other than money.</i>								
ATE	-0.084** (0.039)	-0.073* (0.042)	-0.072 (0.049)	-0.083** (0.039)	-0.094* (0.052)	-0.099* (0.056)	-0.082* (0.045)	-0.087** (0.043)
ATN	-0.091** (0.043)	-0.079 (0.049)	-0.078 (0.055)	-0.060 (0.046)	-0.067 (0.055)	-0.111* (0.064)	-0.068 (0.048)	-0.083* (0.049)
<i>(3) Dependent variable: monthly per capita expenditure.</i>								
ATE	-41.061 (50.675)	-137480.900 (257217.100)	-11.915 (31.621)	-5.578 (16.565)	-0.357 (23.650)	1.407 (22.494)	-7.602 (33.618)	2.688 (24.539)
ATN	-25.992 (32.127)	-16017.500 (121356.300)	-15.940 (26.332)	8.775 (21.542)	1.627 (32.516)	-6.253 (29.432)	-3.929 (26.891)	12.935 (32.423)

*: significant at the 10% level. **: significant at the 5% level. ***: significant at the 1% level. Average treatment effect (ATE) and average treatment effect of the non-treated (ATN) as well as standard errors (in parentheses, obtained via wild bootstrap) are estimated for children from non-eligible households. They are estimated for three outcome variables: whether the student continued with high school in school year 2000/01, whether the student had attended some high school by the end of 2003, and whether the student had finished or was about to finish high school by the end of 2003. The estimates for the first outcome are obtained via the orthogonal estimating equations (9) and (10), while their respective modifications for attrition (15) and (16) are employed for the other two outcomes. All estimates use the median method. The Results are based on 10-fold estimation using 100 sample splits. Column labels denote the method used to estimate nuisance functions.