The Marginal Returns of Distance Education: Evidence from Mexico's Telesecondaries*

Emilio Borghesan Gabrielle Vasey

November 2020

Abstract

This paper analyzes a large-scale and long-running distance education program in Mexico. We use an empirical framework that combines value-added modeling with a sample selection model to estimate Marginal Treatment Effects (MTEs) for learning in Telesecondary schools relative to traditional Mexican secondary schools. The estimated MTEs reveal that school choice is not random, but that the effect of Telesecondary attendance is positive for nearly everyone. Using performance on nationally standardized exams as a measure of knowledge, we find that the average student experiences a 0.34 standard deviation improvement in math and a 0.21 standard deviation improvement in Spanish after one year of attendance in Telesecondary schools. We conclude by estimating the effects of counterfactual policies that expand Telesecondary availability and find that they generate improvements in academic performance.

*We thank Jere Behrman, Susan Parker, and Petra Todd for their help and guidance, as well as seminar participants at the University of Pennsylvania for helpful comments. We are also deeply thankful to Mira Potter-Schwartz and Elizaveta Brover for providing excellent research assistance. The Mexican Secretariat of Public Education (SEP) has generously provided data for this analysis. Borghesan: Department of Economics, University of Pennsylvania, 133 S 36th St, Philadelphia, PA 19104, USA (email: borghesa@sas.upenn.edu). Vasey: Department of Economics, University of Pennsylvania, 133 S 36th St, Philadelphia, PA 19104, USA (email: gvasey@sas.upenn.edu).

1 Introduction

The rise of the internet and video-conferencing platforms has made distance learning an attractive option for students across much of the world. For students in remote locations, distance learning offers access to great instructors at a fraction of the cost of traditional brick-and-mortar schools.[1] However, lectures delivered through a screen may be less effective than the more hands-on approach employed in traditional schools.

Any analysis of distance learning depends on the specific alternative to which it is compared. In this paper, we study the effectiveness of distance education on student achievement in Mexico relative to traditional Mexican schools. Since 1968, Mexico has undertaken an ambitious effort to provide distance education to secondary and post-secondary school students.[2] Starting in the seventh grade, students in Mexico have the option of attending so-called Telesecondary schools, brick-and-mortar establishments where lectures that have been pre-recorded by high quality instructors in Mexico City are transmitted through television broadcast. Students embarking upon a secondary school education therefore decide between a traditional school with subject-specific teachers and in-class instruction and a Telesecondary school that provides televised lectures and standardized assignments all under the supervision of a single adult monitor.

This paper combines a value-added model with a semiparametric sample selection model to evaluate the effect of Telesecondary attendance on student learning. We measure learning in two subjects, math and Spanish, by value added: the intertemporal difference in test scores in nationally standardized exams administered immediately before and one year after the start of secondary school. We define the relative effectiveness of Telesecondaries as the difference in value added between the two school types. Focusing on value-added allows us to quantify the degree of learning that is attributable to the school the student attends, while

[1] The cost per student of providing high-quality distance education for secondary school students in Mexico, the setting of this paper, is approximately one-half that of providing instruction in traditional schools (Martinez Rizo 2005).

[2] Secondary schools in Mexico enroll students in grades seven through nine and are akin to middle schools in the United States, while post-secondary schools enroll students in grades ten through twelve. Throughout the paper, we use the words "secondary" and "post-secondary" to be consistent with the Mexican educational system.

the use of a selection model allows students to choose schools on the basis of characteristics that are unobserved by the econometrician.

We use data on national standardized tests in math and Spanish, the Evaluación National de Logro Académico en Centros Escolares (ENLACE), and augment it with surveys of students, parents, and principals from a random sample of schools as well as geocode data identifying the location of each school. The surveys provide a rich set of observable characteristics of both the child and her family, while the ENLACE and geocode data provide us with information on the school choice of each student, including the location and type of school attended, as well as test scores in math and Spanish at the end of the last year of primary school (sixth grade) and the first year of secondary school (seventh grade). Controlling for sixth grade test scores lets us isolate how much student knowledge in seventh grade is due to secondary school attendance alone.

We use a semiparametric sample selection model to allow for correlation between a student's choice of school and unobserved determinants of academic outcomes at each school. Any sample selection model requires an instrumental variable that affects the decision of which school to attend but does not directly affect outcomes at each school. The instrument we use in this paper is a measure of relative distance. For each student, we have two distance measures. The first is the distance between their primary school and the nearest Telesecondary school. The second is the distance between their primary school and the nearest traditional secondary school. Our instrument is the different between these two measures.[3]

This instrumental variable is highly predictive of attendance in Telesecondary schools: A one kilometer reduction causes the student's probability of attending a Telesecondary school to increase by 3.3% on average. Cameron and Taber (2004) and Carneiro and Heckman (2002) have raised concerns that distance to secondary school is correlated with student ability in the United States. We discuss why endogeneity of this sort is less likely in the Mexican context than in the United States. As a robustness check, we include the distance students actually

[3] A large body of empirical work uses a measure of distance to school as an instrument. See Card (1995), Kane and Rouse (1995), Kling (2001), Currie and Moretti (2003), Cameron and Taber (2004), Carneiro, Heckman, and Vytlacil (2011), and Carneiro, Lokshin, and Umapathi (2017).

travel to secondary school in our outcome equations and find that our results are unchanged.

We find that Telesecondary schools are highly beneficial: The average treatment effect (ATE) of Telesecondary attendance relative to attendance in traditional schools is a 0.342 standard deviation increase in math scores and a 0.218 standard deviation increase in Spanish scores after just one year of attendance. These ATEs conceal considerable heterogeneity in who benefits from Telesecondaries: some students see gains of over 0.5 standard deviations while others experience no benefit.

Our analysis uncovers a pattern of nonmonotonic selection into Telesecondary schools. Students who are the most likely to attend benefit less than students who are slightly less likely to attend. These students in turn benefit more than students who are relatively unlikely to attend. We estimate Marginal Treatment Effect (MTE) curves for math and Spanish that reject the hypothesis of no selection on unobservables.

We then investigate the reasons behind the nonmonotonic pattern of selection. We decompose the MTEs into a component that depends only on the match between students and Telesecondary schools and another component that depends only on the match between students and traditional schools. This decomposition reveals that nonmonotonicity in the MTEs for math and Spanish stem from considerable heterogeneity in the quality of the match between students and traditional secondary schools among students who are likely to choose Telesecondary schools.

We use our estimated MTEs to evaluate counterfactual policies that expand the availability of Telesecondary schools. The first policy we consider is a dramatic increase in Telesecondary availability that reduces the distance to Telesecondary schools for everyone in the sample by 5 km. The second policy under consideration is a school-construction program that builds a Telesecondary school adjacent to the 18% of Mexican primary schools without a Telesecondary school within 5 km. We find that the first policy raises math (Spanish) scores by 0.360 (0.242) standard deviations, while the second raises scores by 0.223 (0.164) standard deviations. The effects of the counterfactual policies differ and neither correspond to the estimates obtained by Two-Stage Least Squares which use distance as an instrument (0.300 and 0.173 standard deviations for math and Spanish, respectively), highlighting the importance of adopting a framework allowing for heterogeneous treatment effects and self-selection as we do in this paper.

Our results are similar to those in Bianchi, Lu, and Song (2020), who study the effect of distance education in rural China on academic and labor market outcomes. Exploiting the differential rollout of a distance education initiative across space and time in a difference-in-differences design, they find that exposure to computer-aided learning raises math skills by 0.18 standard deviations and Chinese skills by 0.23 standard deviations. Relative to their paper, we examine the effect on educational outcomes after a year of attendance rather than seven to ten years later, we allow for the choice of school to be nonrandom, and we demonstrate the extent of heterogeneity in educational outcomes for students with various probabilities of enrolling in distance education.

Our study has several main contributions. First, we believe that this is the first empirical setting in the education literature in which the Marginal Treatment Effect can be nonparametrically identified over its entire support. We are able to precisely compute all treatment parameters and the effects of counterfactual policies using our semiparametric estimates of the MTE. The reasons for nonparametric identification stem from the both the large size of the sample (over $120,000$ observations) and the strong instrument, which induces significant variation in the probability of attending a Telesecondary school. Nonparametric identification turns out to be important, as the MTE is nonmonotonic and parametric approaches fail to identify this feature. In addition, our paper is the first to consider the impact of Telesecondaries on academic achievement. With our unique data, we are able to identify the schooling options available to each student and associate potential academic outcomes with attendance in each type of school, thereby allowing us to estimate the effect of Telesecondary schools on academic achievement.

1.1 Related Literature

A burgeoning literature evaluates the effects of schooling using sample selection models similar to the one in this paper. Carneiro, Heckman, and Vytlacil (2011) evaluate the decision to attend college in the United States. They find evidence of considerable heterogeneity in the pecuniary returns to college attendance and a pattern of positive sorting on gains: When considering policies that expand college attendance, they find that the returns to the marginal student induced to attend college are signficantly lower than the returns to students already attending college. Carneiro, Lokshin, and Umapathi (2017) use similar methods to analyze the pecuniary returns to secondary school attendance in Indonesia. Using distance to the nearest secondary school as an instrument that influences the probability of attendance but not outcomes directly, they also find positive sorting on gains and considerable heterogeneity, whereby the students with the lowest likelihood of secondary school attendance actually have negative returns. Cornelissen et al. (2018) analyze the decision of parents to enroll their children in day care in Germany, and, contrary to the two previous papers, find a pattern of reverse sorting on gains. Students who are not enrolled in Germany's universal child care program would experience increases in their readiness for primary school had they attended child care, while those currently attending experience little benefit. The authors conclude that universal child care disproportionately subsidizes families that are well-off and is not sufficiently accessible for minority households.

A large literature examines the interaction of technology and education and is surveyed in three excellent recent papers by Bulman and Fairlie (2016), Escueta et al. (2017), and Rodriguez-Segura (2020). Papers that examine the effectiveness of distance learning in developing countries are much fewer in number. Beg et al. (2019) use data from two randomized controlled trials (RCTs) in Pakistan and find that the combination of high quality videos with in-person teaching raises student performance on standardized tests. In another RCT, Johnston and Ksoll (2017) find that a similar program in Ghana, which broadcasts live instruction from experts into rural schools, had a positive impact on test scores. Given the positive findings on distance education, it is perhaps surprising that this low-cost alternative is not prevalent in developing countries and rural areas around the world. Our paper validates these findings from RCTs by conducting analysis of a longstanding distance education policy using observational data and a framework that allows for nonrandom school choice.

Although Telesecondaries are widespread and have existed for over 50 years, little research on their effectiveness exists. Recently, two papers have used difference-in-differences designs to estimate the effect of proximity to Telescondaries on educational attainment and labor market outcomes (Fabregas (2019), Navarro-Sola (2019)). Both papers use data from the Mexican Census and thus lack information on student test scores and the type of school attended. Even without this information, both papers find that Telesecondaries raise educational attainment and future income, although the estimates in Navarro-Sola (2019) are substantially larger than those in Fabregas (2019). Behrman, Parker, and Todd (2020) look at the impact of conditional cash transfers on schooling trajectories and find that cash transfers primarily raise schooling through increasing enrollment in Telesecondary schools.

1.2 Secondary Schooling in Mexico

Throughout the twentieth century, Mexico struggled to attract qualified teachers to rural areas to instruct the millions of school-age children living there. Introduced in 1968, *Telesecundarias*, or Telesecondary schools, were seen as a solution to this problem. Telesecondary schools are physical structures with four key scholastic components. The first is the television program. Every subject begins with students watching 15 minutes of a pre-recorded televised lecture. Lectures for each subject are recorded by high-quality instructors, so-called *Telemaestros*, who are selected for their communication skills.

Following the video lectures, a single teacher leads students in a 35-minute lesson on the same subject. The teacher does not specialize in a particular subject: Students have the same teacher for all courses. This teacher follows a teaching guide designed for the Telesecondary schools filled with teaching suggestions for each subject. The 35 minutes are spent in myriad ways, with the teacher leading question and answer sessions, engaging students in group activities, and giving assignments for students to do on their own.

The third educational resource is an encyclopedia-like book that contains the essential information in each subject taught during that year. These books are similar to textbooks in traditional secondary schools and are used by students as references while doing their assignments. The final component of Telesecondary education is the learning guide. Like a workbook, learning guides are filled with questions that students can answer individually as well as suggestions for group activities that reinforce learning. Class time is frequently devoted to doing assignments in the learning guides.

The four main educational components – televised instruction, in-person teaching, reference texts, and learning guides – are designed to be complementary. Students evidently see them that way: Ethnographic research indicates that students see each component as reinforcing the knowledge acquired through the televised lectures (Estrada 2003).

Telesecondaries were first introduced in rural areas and predominate in Mexico's poorer South. While they have expanded into suburban and urban area, students from the South and from rural areas are still over-represented (see Table 3). A reform in 1993 mandated schooling through grade nine and resulted in increases in the construction of both new Telesecondary schools and Telesecondary enrollment. We study a single cohort, those students who were in grade six in 2007/08, after secondary schooling became compulsory.

2 Model

2.1 Student Achievement

We apply the potential outcomes framework of Rubin (1974) to a value-added model of learning. Students can either attend a Telesecondary or a traditional school.[4] We define the random variable D where $D = 1$ denotes attendance in a Telesecondary school and $D = 0$ denotes attendance in a traditional school. We study the effects of telesecondary attendance on two outcomes: math and Spanish test scores in seventh grade. For each course $C \in \{\text{Math, Spanish}\}$, the potential outcomes Y_0^C and Y_1^C correspond to the test score a student would achieve had she enrolled in a traditional or Telesecondary school, respectively. For ease of notation, we will omit the C superscripts. The same causal model will be used for each of the two outcomes.

We model test score outcomes and school choice according to the selection model in Heckman and Vytlacil (2005):

$$Y_1 = \beta_1 X + U_1 \ , \tag{1}$$

$$Y_0 = \beta_0 X + U_0 \ , \tag{2}$$

$$D = \mathbb{1}(\mathbf{Z}\gamma > V) \ , \tag{3}$$

where X is a vector of observable characteristics influencing outcomes, Z is a vector of observable characteristics influencing the choice of secondary school, and (U_1, U_0, V) are unobserved by the econometrician. Students are assumed to know (U_1, U_0, V) and may act upon them. Equations (1) and (2) are value-added equations: X contains the previous year's test scores in both math and Spanish. As demonstrated in Todd and Wolpin (2003), such a specification is consistent with an educational production function in which the effects of time-varying investments on test scores decline geometrically with the time between when the investment was made and when the test was taken.

The effect of attending a Telesecondary school relative to a traditional school on

[4] Mexico has three secondary school types: General, Technical, and Telesecondary. We consider the choice between a traditional school (General/Technical) and a Telesecondary school. Table 2 reveals that General and Technical school have similar distributions of observable household characteristics and student student test scores, so we feel that it is reasonable to consider them as a single alternative for the purposes of evaluating learning in math and Spanish between the sixth and seventh grades. However, the ensuing analysis goes through without modification if they are treated as separate alternatives as long as the results are re-interpreted as the causal effect of Telesecondary education relative to the next best alternative. A small fraction of students in Mexico attend Private schools. We exclude them from the analysis. Students who drop out are also omitted.

the test scores for an individual is given by $Y_1 - Y_0$, and the average effect for individuals with a specific set of observable characteristics is $ATE(X) = \mathbb{E}[Y_1 - Y_0|X = x]$. The fundamental challenge in estimating any sort of treatment effect is that the econometrician only observes one of the two potential outcomes, $Y = Y_0 + D(Y_1 - Y_0)$.

The Marginal Treatment Effect (MTE), introduced by Björklund and Moffitt (1987) and extended in Heckman and Vytlacil (1999, 2001b, 2005, 2007), is the average treatment effect for an individual at a particular margin of "resistance to treatment." V, in equation (3), represents this resistance to treatment. An individual with a higher V is, on the basis of unobservables, less likely to attend a Telesecondary school. We apply the following useful transformation to equation (3) to obtain $D = \mathbb{1}(Z\gamma > V) = \mathbb{1}(F_V(Z\gamma) > F_V(V)) = \mathbb{1}(P(Z) > U_D)$, where P(Z) is the propensity score and $U_D \sim U[0,1]$. Following the transformation, the MTE can be written as

$$MTE(x, u_D) = \mathbb{E}[Y_1 - Y_0|X = x, U_D = u_D]\,. \tag{4}$$

The Marginal Treatment Effect gives the average difference in outcomes at Telesecondary schools relative to traditional schools for individuals with observable characteristics x and latent resistance to treatment u_D.

2.2 Identification

The Marginal Treatment Effect is identified under the following assumptions, stated in Heckman and Vytlacil (2005) and modified slightly to fit the notation presented here:

(A-1) Z is a nondegenerate random variable conditional on X.

(A-2) $(U_0, U_1, U_D) \perp\!\!\!\perp Z|X$.

(A-3) U_D is absolutely continuous with respect to the Lebesque measure.

(A-4) $\mathbb{E}|Y_1|$ and $\mathbb{E}|Y_0|$ are finite.

(A-5) $1 > \mathbb{P}(D = 1|X) > 0$

Assumption (A-1) ensures that the instrument influences attendance in Telesecondary schools conditional on covariates, X, while (A-2) assumes that the instrument is exogenous in the sense that it is independent of unobservable variables in the selection and outcome equations conditional on X. Assumptions (A-3) - (A-5) are technical assumptions that are satisfied in our setting. Under these assumptions, $MTE(X, U_D)$ is nonparametrically identified by the Local Instrumental Variables estimand (Heckman and Vytlacil 2001a), $\frac{\partial \mathbb{E}[Y|X=x,P(z)=p]}{\partial p}$:

$$\frac{\partial \mathbb{E}[Y|X=x, P(z)=p]}{\partial p} = MTE(X,p) \ .$$

Thus, the Marginal Treatment Effect at each value of the latent resistance to treatment, U_D, is identified by individuals who are indifferent between being treated and not, because when $p = U_D$, the individual is on the knife edge between participating and not participating.

2.3 Parameters of Interest

A large class of parameters corresponding to the effect of Telesecondary attendance on schooling outcomes can be written as weighted averages of the Marginal Treatment Effect. In this paper we are interested in $ATE(X)$, as well as the average effect of treatment on the treated, $TT(X) = \mathbb{E}[Y_1 - Y_0 \mid X = x, D = 1]$, the average effect of treatment on the untreated, $TUT(X) = \mathbb{E}[Y_1 - Y_0 \mid X = x, D = 0]$, and a range of treatment effects corresponding to the effects of never-before implemented policies. These Policy-Relevant Treatment Effects, first defined in Heckman and Vytlacil (2001b), are defined for a shift from a pre-existing policy a to a new policy a' and provide a normalized effect of the policy change:

$$PRTE_{a',a}(X) \frac{\mathbb{E}[Y_{a'} - Y_a \mid X = x]}{\mathbb{P}(D'_a = 1 \mid X = x) - \mathbb{P}(D_a = 1 \mid X = x)} \ . \tag{5}$$

Any treatment parameter, including ATE, TT, TUT, and $PRTE_{a',a}$ for policies a and a', can be computed by integrating the MTE with respect to the distribution of U_D induced by the treatment parameter under consideration:

$$ATE = \int MTE(X, U_D) dF(X, U_D) ,$$

$$TT = \int MTE(X, U_D) dF_{U_D, X|D=1}(x, u_D \mid D = 1) , \tag{6}$$

$$TUT = \int MTE(X, U_D) dF_{U_D, X|D=0}(x, u_D \mid D = 0) , \tag{7}$$

$$PRTE_{a',a} = \int MTE(X, U_D) dF_{U_D, X|D_a=0, D_{a'}=1}(x, u_D \mid D_a = 0, D_{a'} = 1) . \tag{8}$$

Section 4 discusses the methods we use to integrate the MTE to obtain these treatment parameters.

3 Data

We examine a single cohort, students who were in the sixth grade in 2007/08, and combine data on them from three different sources. The first is an administrative data set with student scores on nationally-standardized tests in math and Spanish. These exams, the ENLACE, were administered at the end of each school year in 2007/08 and 2008/09. In addition to test scores, this data set contains information on the age, gender, conditional cash transfer status, school attendance, school ID, and school type of each student.

We link the test score data with information on student, parent, and school characteristics from a random survey of schools administered during the years the ENLACE exam was administered. These surveys provide detailed information on parental education, monthly family income, home infrastructure, number of siblings, and other household characteristics.

Finally, we collect the latitude and longitude of each primary and secondary school and calculate the distance (in kilometers) between the primary school each student attends and the nearest secondary school of each type. We subtract the distance to the nearest traditional school from the distance to the nearest Telesecondary school to obtain a measure of relative distance. We use this relative distance measure as an exclusion restriction which affects the choice of school, D, but does not affect outcomes, Y_1 and Y_0, directly. While it might be preferable to measure distance from the student's actual home (rather than primary school) to each secondary school, this relies on data that is not available in any of our sources. In section 3.1, we show that the instrument that we have constructed is highly predictive of attendance in Telesecondary schools.

Figure 1 shows the distribution of the instrument by school attendance. A negative value on the x-axis indicates that a Telesecondary school is closer, while a positive value indicates that a traditional school is closer. The figure reveals that students mostly attend the school that is closer, but when the two schools are equally close (distance = 0), many more students attend traditional schools.

We omit from our analysis students whose relative distance measure lies outside the middle 99% of the distribution and students who attend a secondary school more than 15 km from their primary school. This is done for two reasons. First, we want to consider students who have a choice set consisting of two feasible alternatives, and so we drop students with only one nearby school. Second, we want to omit students who move to a different school district. Such movements could cause correlation between the instrument and the outcomes, Y_1 and Y_0. We also omit students who drop out. We end up with a sample of 126,590 students. In Section 5, we discuss how omitting dropouts influences the interpretation of our counterfactual analysis.

Table 1 lists all the outcome variables (Y), covariates (X), and the instrumental variable ($Z \setminus X$) that we use in our empirical analysis. Apart from the outcome variables, sixth grade test scores, and the instrument, all variables are categorical.

We present summary statistics for these variables in Table 3 by school type. The table reveals that students who attend Telesecondary schools are disadvataged according to a wide range of metrics relative to students who attend traditional schools. They are disproportionately beneficiaries of the conditional cash transfer program Prospera, they come from poorer household with less educated moth-

Variable	Definition
Y	Math score in 7th grade, Spanish score in 7th grade
X	**Parent:** Mother's education, family income, Prospera status, rural residence, residence in Northern state, number of books in the home, whether the family has access to a computer. **Child:** Math score in 6th grade, Spanish score in 6th grade, age, sex, number of siblings.
$Z \backslash X$	Relative distance between the nearest Telescondary and the nearest traditional school.

Table 1: Variables used in estimation

ers, and they fare worse academically in the year prior to secondary school.[5] Nevertheless, they make up nearly half of the gap in Spanish and nearly the entire gap in Math relative to their peers at traditional secondary schools after just a year of Telesecondary attendance.

3.1 Is Relative Distance a Valid Instrument?

Identification of the Marginal Treatment Effect requires that the instrument satisfy (A-1) and (A-2)′. The first assumption, that relative distance predict attendance in Telesecondary schools conditional on observable covariates, X, is easily verified. Table 4 displays the average marginal effects of each variable in Z on the probability of attending Telesecondary school (estimated via Probit). The average marginal effect of relative distance on the probability of attending Telesecondary school is 3.3% per kilometer and highly significant.

The second assumption, that the instrument be independent of unobservable variables in the outcome and selection equations (U_1, U_0, V), is untestable. In what follows, we discuss potential threats to instrumental exogeneity.

Since our specification controls for lagged test scores in addition to family and

[5]The conditional cash transfer program in Mexico began in 1997 and has been has been called Progresa, Oportunidades, and Prospera. The main educational component of the program is that families receive a cash transfer if their child is enrolled in school. Parker and Todd (2017) provide a review of the literature on the effects of conditional cash transfer in Mexico and conclude that it has been effective in increasing school enrollment, reducing grade retention, and increasing educational attainment.

child characteristics, any threats to instrumental exogeneity must be caused by a correlation between distance and unobserved *time-varying* determinants of student outcomes that occur between the sixth and the seventh grade. Such a correlation could occur if parents knew their child's realizations of U_1 and U_0 and moved to be closer to the school with the higher unobserved outcome. As discussed in section 3, we drop from the sample any children who attend a secondary school more than 15 km from their primary school in an effort to eliminate this threat to exogeneity.

Another threat to identification could be that the distance traveled to school directly causes worse academic performance. This could occur if fatigue caused by walking long distances to school lowered a student's ability to concentrate. To alleviate this concern, we conduct a robustness check that controls for the distance actually traveled to secondary school in equations (1) and (2). Reassuringly, our estimates of the MTE and treatment parameters are unaffected by its inclusion.

4 Estimation

Estimating $MTE(X, U_D)$ under assumptions (A-1) - (A-5) require that we estimate $MTE(X, U_D)$ separately for each X. When X is high-dimensional, as in our setting, $MTE(X, U_D)$ is only identified by the support of $P(Z)$ given X. Even if the unconditional support of $P(Z)$ is the entire unit interval, $supp(P(Z)) \mid X = x)$ may consist of only a few points. Since this will be too few to estimate $MTE(X, U_D)$ with any degree of precision, we strengthen assumption (A-2) to (A-2)′:

(A-2)′ $(X, Z) \perp\!\!\!\perp (U_1, U_0, U_D)$

Assumption (A-2)′ is standard in the literature estimating selection models. It has two consequences. Under this assumption, the MTE is additively separable in X and U_D so that

$$MTE(X, U_D) = X(\beta_1 - \beta_0) + K(U_D) \,.$$

An additional consequence of assumption (A-2)′ is that $MTE(X, U_D)$ can now be identified over the unconditional support of $P(Z)$ rather than $supp(P(Z) \mid X)$. The cost of the assumption is that it restricts the pattern of selection on unobservables – given by the shape of $MTE(X, \cdot)$ – to be the same across individuals with different observable characteristics, X. It rules out the possibility that $MTE(X, \cdot)$ has a different slope depending on the value of X (level shifts can be accommodated). In Appendix A, we consider the validity of this restriction, by binning X according to observable characteristics and estimating $MTE(X, U_D)$ separately on the subsamples defined by these bins. Reassuringly, the shape of selection does not vary much across the different subsamples. Only statistically significant level shifts in $MTE(X, U_D)$ are apparent across the different groups.

We estimate the MTE using two methods: a fully parametric approach that specifies the distribution of unobservables and a semiparametric approach that leaves the joint distribution of unobservables unspecified and estimates $\frac{\partial \mathbb{E}[Y|X=x, P(z)=p]}{\partial p}$ using Local Polynomial Modeling.

The fully parametric approach specifies the unobservables in the selection and outcome equations as jointly normally distributed:

$$\begin{pmatrix} U_1 \\ U_0 \\ V \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{10} & \sigma_{1V} \\ & \sigma_0^2 & \sigma_{0V} \\ & & 1 \end{pmatrix} \right] . \tag{9}$$

Under these assumptions, the Marginal Treatment Effect has the following simple functional form:

$$MTE(X, u_D) = X(\beta_1 - \beta_0) + (\sigma_{1V} - \sigma_{0V})\Phi^{-1}(U_D) .$$

We estimate the parameters $\beta_1, \beta_0, \sigma_{1V}, \sigma_{0V}$ via a two-step method that first estimates the propensity score via Probit and then includes control functions in the outcome equations as follows:

$$\mathbb{E}(Y_1 \mid D = 1, X, Z) = X\beta_1 + \mathbb{E}(U_1 \mid D = 1) ,$$

$$
\begin{aligned}
&= X\beta_1 + \sigma_{1V}\left(-\frac{\phi(\Phi^{-1}(P(Z)))}{P(Z)}\right) , \\
\mathbb{E}(Y_0 \mid D=0, X, Z) &= X\beta_0 + \mathbb{E}(U_0 \mid D=0) , \\
&= X\beta_0 + \sigma_{0V}\left(\frac{\phi(\Phi^{-1}(P(Z)))}{1-P(Z)}\right) .
\end{aligned}
$$

The second approach is the semiparametric Local Instrumental Variables Estimator developed in Heckman and Vytlacil (2001a). This approach differs from the Generalized Roy Model in that it does not make any assumption regarding the joint distribution of (U_1, U_0, V).[6] We estimate $\frac{\partial \mathbb{E}[Y|X=x,P(z)=p]}{\partial p}$ using the partially linear model estimator of Robinson (1988). To understand the approach note that, assumptions (A-1), (A-2)′, (A-3) - (A-5), together with the assumption that the outcome models in (1) and (2) are linear, yields a conditional expectation function that is linear in X and XP and nonlinear in the propensity score, P, as follows:

$$
\begin{aligned}
\mathbb{E}[Y|X=x, P(z)=p] &= \mathbb{E}[Y_0 + D(Y_1 - Y_0)|X=x, P(z)=p] \\
&= X\beta_0 + \mathbb{E}[DX(\beta_1 - \beta_0)|X=x, P(z)=p] + \\
&\mathbb{E}[U_0 + D(U_1 - U_0)|X=x, P(z)=p] \\
&= X\beta_0 + PX(\beta_1 - \beta_0) + K(P) .
\end{aligned}
$$

Because of this form for the conditional expectation, the Marginal Treatment Effect evaluated at $U_D = P$ is given by

$$
MTE(X, P) = X(\beta_1 - \beta_0) + \frac{\partial K(P)}{\partial P} . \tag{10}
$$

The semiparametric estimator of (10) entails two steps. First, the estimated propensity score, P, is partialed out of the other variables by running nonparametric regressions of Y, X, and PX on P. Then the residualized Y is regressed linearly on the residualized X and PX to obtain estimates of β_0 and $\beta_1 - \beta_0$. In the second

[6]We estimate the propensity score model using a Probit which assumes normality of the marginal distribution of V, but makes no assumption regarding its joint distribution with (U_1, U_0).

step, the derivative of the conditional expectation of $\tilde{Y} \equiv Y - X\hat{\beta}_0 - XP(\hat{\beta}_1 - \hat{\beta}_0)$ with respect of P is estimated nonparametrically to obtain an estimate of $\frac{\partial K(P)}{\partial P}$.

All nonparametric regressions are estimated using local polynomial regression. Following the recommendations in Fan and Gijbels (1996) we use local linear regression to estimate the conditional expectations in the first stage and local quadratic regression to estimate $\frac{\partial K(P)}{\partial P}$ in the second stage. A single bandwidth is used for all nonparametric regressions for a particular outcome (math or Spanish scores). We choose bandwidths using the plug-in estimator of Fan and Gijbels (1996), which aims to minimize the Integrated Mean Square Error (IMSE) in the final nonparametric regression. The IMSE-minimizing bandwidth depends negatively on the function's curvature (second derivative) and on the density of the data, and positively on the conditional variance of the outcome variable. The plug-in method selects a bandwidth of 0.28 for both math and Spanish.

We compute treatment parameters for both the parametric and semiparametric approaches. Formulas exist for the parametric approach:

$$
\begin{aligned}
ATE &= \bar{X}(\beta_1 - \beta_0) \ , \\
TT &= \frac{1}{N_T} \sum_{i=1}^{N_T} D_i \{ X(\beta_1 - \beta_0) + \mathbb{E}[U_1 - U_0 | D_i = 1] \} \ , \\
TUT &= \frac{1}{N_G} \sum_{i=1}^{N_G} (1 - D_i) \{ X(\beta_1 - \beta_0) + \mathbb{E}[U_1 - U_0 | D_i = 0] \} \ ,
\end{aligned}
$$

where $D_i = 1$ if an individual attends a Telesecondary school and 0 otherwise, N_T denotes the number of students attending Telesecondary schools, and N_G denotes the number of students attending traditional schools. As a result of the assumption that (U_1, U_0, V) are jointly normally distributed, $\mathbb{E}[U_1 - U_0 | D_i = 1] = (\sigma_{1V} - \sigma_{0V}) \left(-\frac{\phi(\Phi^{-1}(P(Z)))}{P(Z)} \right)$ and $\mathbb{E}[U_1 - U_0 | D_i = 0] = (\sigma_{1V} - \sigma_{0V}) \left(\frac{\phi(\Phi^{-1}(P(Z)))}{1 - P(Z)} \right)$

For the semiparametric approach, we integrate $MTE(X, U_D)$ with respect to the appropriate distributions in equations (6) - (8) using the simulation method introduced in Carneiro, Lokshin, and Umapathi (2017). The simulation approach, which is only valid under assumption (A-2)', involves creating an equally-spaced grid for U_D for each individual and averaging $MTE(X, U_D)$ for the values of

U_D on the grid that are less than that individual's propensity score $P(Z)$ for TT, greater than that individual's propensity score for TUT, and between $P(Z_a)$ and $P(Z_{a'})$ for PRTE. Figure 5 displays the densities used to compute ATE, TT, and TUT plotted as a function of U_D. ATE uniformly samples individuals with all levels of U_D while TT oversamples individuals with low U_D and TUT oversamples individuals with high U_D.

4.1 Empirical Results

Figure 3 presents the MTEs for seventh grade math scores evaluated at mean values of X. The parametric and semiparametric MTEs are plotted side-by-side with 90% confidence bands in grey. Figure 4 repeats the analysis with Spanish as the outcome variable. The horizontal axis measures the latent variable U_D, while the vertical axis measures the expected benefit to attending a Telesecondary school relative to a general school for students with that level of U_D, $\mathbb{E}[Y_1 - Y_0 | X = \overline{x}, U_D = u_D]$. The MTEs in both figures are precisely estimated.

The semiparametric figures reveal a pattern of non-monotonic selection on gains. Students who, on the basis of unobservables, are most likely to attend Telesecondaries (low U_D) have value added that is indistinguishable from zero for both math and Spanish. As U_D increases, students are less likely to attend Telesecondaries but their benefits from attendance increase rapidly to a peak at about $U_D = 0.35$ for both Math and Spanish. The average benefit to attending Telesecondary schools is positive and statistically significant for students with this level of U_D. From this point onward, as U_D increases, average benefits decrease in both math and Spanish, although there are large (but noisy) gains to Telesecondary attendance for students with the largest values of U_D.

While the MTEs for both math and Spanish have similar shapes, the pattern of nonmonotonic selection is more pronounced for Spanish than for math. The Spanish MTE curve also displays greater variability, ranging from 0.04 at $U_D = 0$ to 0.96 at $U_D = 1$, while math ranges from -0.06 at $U_D = 0$ to 0.76 at $U_D = 1$.

Tables 5 and 6 present estimates of standard treatment parameters for math and Spanish, respectively. All treatment parameters are positive, underscoring the findings from the MTE curve that Telesecondary attendance is beneficial for a large majority of students. Standard errors reveal that the treatment parameters are precisely estimated and are significant at conventional levels of significance. The semiparametric estimate of the ATE for math indicates that a randomly selected student would be expected to perform 0.342 standard deviations better in mathematics in the seventh grade had she attended a Telesecondary school instead of a traditional school. The estimates of TT (0.279 standard deviations) are smaller than those of TUT (0.356), highlighting that while selection is nonmononotic, it is mostly negative selection (an upward-sloping MTE) rather than positive selection (downward-sloping MTE). The semiparametric estimates of ATE, TT, and TUT for Spanish are 0.218, 0.168, and 0.229 standard deviations, respectively, smaller than for math, but still revealing evidence of negative selection.

The parametric estimates of treatment parameters for math are noticeably higher than the semiparametric estimates. The parametric approach forces the MTE curve to be monotonic, and so does a bad job of estimating the MTE for both Math and Spanish, but the misspecification is worse for math than for Spanish. Parametric estimates for Spanish are quite similar to the semiparametric estimates, owing to the fact that Spanish's parametric MTE curve is well-centered between the maxima and minima of the semiparametric MTE curve.

4.2 Evidence of Selection on Unobservables

A Marginal Treatment Effect that is nonconstant in U_D is evidence of a pattern of selection on unobservables. We test formally for evidence of selection using methods developed in Heckman, Schmierer, and Urzua (2010). As explained in Heckman and Vytlacil (2005), the Local Average Treatment Effect (LATE) introduced by Imbens and Angrist (1994) is simply the integral of the MTE over a region of the domain of U_D. One way of testing for evidence of selection is to test whether LATEs defined by integrating the MTE over different intervals of $supp(U_D)$ are equivalent. Tables 7 and 8 display the results of these tests for math and Spanish, respectively. To perform these tests, we partition the support of U_D into 25 intervals of width 0.04 and test whether the integrated MTEs on adjacent (but not overlapping) intervals differ. We then test that all LATEs are jointly equal to each other. The tests are conducted by using 50 bootstrapped data sets and computing simulated p-values under the null hypothesis of equivalent LATEs.

The Table reveals that many adjacent LATEs differ from one another and that, jointly, the LATEs are not equal at the 10% significance level for either math or Spanish. For example, the entry in column 1 of Table 7 indicates that

$$\mathbb{E}(Y_1 - Y_0|X = \overline{x}, 0.04 < U_D \leq 0.08) - \mathbb{E}(Y_1 - Y_0|X = \overline{x}, 0 \leq U_D \leq 0.04) = 0.196 \ ,$$

and that the p-value that this difference is different than 0 is $p = 0.000$. The joint p-value for the test of the hypothesis that all LATEs are equivalent is smaller for Spanish than for math, $p = 0.000$ versus $p = 0.080$, and is driven by the greater curviness in the estimated MTE for Spanish. As a result of these tests, we reject the hypothesis of no selection on unobservables. Attendance in Telesecondary schools is correlated with unobserved determinants of student achievement.

4.3 Explaining the Pattern of Selection

The MTEs for both math and Spanish are nonmonotonic and display evidence of reverse selection on gains. In this section we discuss the source of this nonmonotonicy and what it implies for the unobserved outcomes at both Telesecondary and traditional schools.

Following methods outlined in Brinch, Mogstad, and Wiswall (2017) we can rewrite the Marginal Treatment Effect as

$$MTE(X, U_D) = X(\beta_1 - \beta_0) + k_1(U_D) - k_0(U_D) \ , \tag{11}$$

where $k_j(U_D) = \mathbb{E}[U_j|U_D]$ for $j = 1, 2$. $k_1(U_D)$ can be thought of as the average unobserved match quality between student and Telesecondary school for students with a particular resistance to attending Telesecondary schools (given by U_D), while $k_0(U_D)$ represents the unobserved match quality between students and traditional schools.

In this section, we estimate $k_1(U_D)$ and $k_0(U_D)$ individually to determine whether the shape of the MTE curve is determined primarily between variability in match qualities between students and Telesecondary schools or between students and traditional schools. As shown in Heckman and Vytlacil (2007) and Brinch, Mogstad, and Wiswall (2017), $k_1(U_D)$ and $k_0(U_D)$ can be estimated by using a control function approach on each of the $D = 1$ and $D = 0$ subsamples. Under Assumption (A-2)′,

$$\mathbb{E}[Y_j|X = x, P(Z) = p, D = j] = X\beta_j + K_j(p) \ ,$$

for $j = 0, 1$, where

$$K_1(p) = E\left(U_1 \mid U_D \leq p\right) \ ,$$

and

$$K_0(p) = E\left(U_0 \mid U_D > p\right) \ .$$

We can obtain k_1 and k_0 from K_1 and K_0 using the following identities in Brinch, Mogstad, and Wiswall (2017):

$$\begin{aligned} k_1(p) &= p\frac{\partial K_1(p)}{\partial p} + K_1(p) \\ k_0(p) &= -(1-p)\frac{\partial K_0(p)}{\partial p} + K_0(p) \ . \end{aligned} \tag{12}$$

Figure 6 shows estimates of k_1 and k_0 for Math as an outcome variable. Each estimated curve, k_j for $j = 0, 1$, is obtained via two semiparametric regressions on the subsample with $D = j$. We compute the conditional expectation of $Y - X\beta_j$ given the estimated propensity score, p, using Local Linear Regression to obtain $K_j(p)$ and the derivative of the conditional expectation of $Y - X\beta_j$ given p via Local Quadratic Regression to obtain $\frac{\partial K_j(p)}{\partial p}$. $k_j(p)$ is then obtained using the identities in 12. The bandwidths used are the same as in the estimation of the MTE function in Section 4: 0.28 for both math and Spanish.

Figure 6 demonstrates that variation in mean unobserved outcomes at Telesecondaries, $k_1(U_D)$, is dwarfed by variation in mean unobserved outcomes at traditional schools, $k_0(U_D)$ for students with low U_D, namely those that are most likely to attend Telesecondary schools. Essentially, Telesecondary and traditional schools are equally good for students who are most likely to attend Telesecondaries. But the quality of the match at traditional schools rapidly declines as students become less likely to attend Telesecondary schools (U_D increases). As U_D increases still further, the match quality between students and traditional schools is remarkably constant. At the same time, the match quality between students and Telesecondaries improves as U_D increases.

That $k_0(U_D)$ is flat for nearly all students suggests little heterogeneity in the quality of the match between students and traditional schools. There is, however, significant variation in learning at Telesecondaries, reflected in the nonzero slope of $k_1(U_D)$. It is puzzling that students who have the highest benefit from attending Telesecondary schools, given by $k_1(U_D)$ for U_D close to 1, are among the least likely to attend. It suggests that either students do not decide which school to attend based on expected test scores at each school (a simple model with $D = \mathbb{1}(Y_1 - Y_0 > 0)$ would generate a pattern of positive selection), or that some different actor who is not altruistic chooses the school for the student.

4.4 Sensitivity Analysis

In this section we consider alternative specifications, including those designed to investigate the validity of Assumption (A-2)′.

To alleviate concerns that the instrument is directly correlated with academic outcomes, we augment equations (1) and (2) for academic outcomes with the distance actually traveled to secondary school as an explanatory variable. If traveling long distances causes students' academic performance to suffer, or if student's distance from the secondary school they attend were correlated with unobserved determinants of academic outcomes, then its inclusion would change our estimated treatment parameters in a significant way.

Estimates of the MTEs and treatment parameters in specifications that control for distance traveled to secondary school are presented in Appendix B. Reassuringly, we find that including distance traveled to secondary school makes little difference. Figures B-1 and B-2 display the estimated MTE curves for these specifications. The pattern of selection, given by the shape of the MTE curves for math and Spanish, as well as the point estimates of treatment parameters presented in Tables B-1 and B-2 are little changed from our main specification. We estimate an ATE of 0.324 standard deviations for math scores and 0.218 for Spanish. These estimates lie within one standard deviation of the estimates generated by our main specification. The other treatment parameters are equally close.

We also investigate whether our results may be driven by the inclusion of a large number of students in Mexico City. Over 10% of our sample attends school in Mexico City, a region which is unique for a variety of reasons including its wealth and high population density. We present the estimated treatment parameters from this investigation in Table B-3 and B-4 in Appendix B. While the smaller sample size causes the treatment parameters to be more noisily estimated, the tables reveal that our results are robust to the exclusion of Mexico City from the analysis. We estimate an ATE of 0.328 standard deviations for math value-added and 0.225 for Spanish, which lie within one standard deviation of the estimates generated by our main specification. The other treatment parameters also do not differ by more than one standard deviation.

5 Counterfactuals

The effects of counterfactual policies can be evaluated by integrating the MTE with respect to a probability distribution induced by the proposed policy. A baseline policy, a, is characterized by a particular distribution of the instrument, Z^a. The move from policy a to a new policy a' corresponds to a shift in the distribution of the instrument from F_{Z^a} to $F_{Z^{a'}}$. This shift induces some students to attend Telesecondaries who would not otherwise attend. The treatment effect for students induced to switch attendance from traditional to Telesecondary schools as a result of the policy is given $PRTE_{a,a'}$, which is positive if these students learn more in Telesecondaries than traditional schools.

We consider a class of policies that expand access to Telesecondary schools so that $Z^{a'} \leq Z^a$ for all students. As we omit individuals who drop out between the sixth and seventh grades, we will not be able to say anything about the distribution of test scores for students who are induced to attend Telesecondary schools instead of dropping out as a result of the counterfactual policy.[7] Our PRTE estimates apply only to the population of students who were already attending secondary school under the baseline policy, which is the actual policy in 2008.

We consider two counterfactual policies. The first is a hypothetical policy that reduces the relative distance to Telesecondary schools by 5 km for every student. It is not a feasible policy, as it would entail moving general schools farther away for students whose nearest Telesecondary is under 5 km. The policy counterfactual is merely analyzed as an example of the gains to a policy that can drastically raise Telesecondary attendance.

The second counterfactual is a feasible school-building policy that constructs a Telesecondary school directly adjacent to the eighteen percent of primary schools that have no Telesecondary within a 5 km radius. This has the effect of reducing the distance between primary and Telesecondary schools to zero for all students who formerly had only a distant Telesecondary .

Figure 7 displays the probability distributions of U_D corresponding to these two treatment parameters. The first counterfactual induces a probability distribution unlike any of the other treatment parameters: the distribution is considerably less skewed than TT and has the most mass around $U_D = 0.30$. As a result, it oversamples individuals with some of the largest values of $\mathbb{E}[Y_1 - Y_0 \mid U_D]$ and produces a large positive value for PRTE. The distribution corresponding to the second counterfactual is closer to the distribution for TT. Relative to TT, it oversamples individuals with low U_D and undersamples those with high U_D. The figure also shows the weights corresponding to a Two-Stage Least Squares regression that uses relative distance as an instrument for Telesecondary attendance. The IV weights correspond to neither of the PRTE weights (nor do they correspond to the weights for ATE, TT, or TUT), highlighting the importance of identifying the Marginal Treatment Effect for the purposes of conducting counterfactual analysis.

[7] A revealed preference argument demonstrates that no students will transition from dropping out to attendance in traditional schools as a result of the policy. A similar argument can be made to show that no students will be induced to drop out under the proposed policy if they were attending secondary school under the baseline policy.

Table 9 presents the estimates of the PRTEs for both math and Spanish alongside the IV estimate resulting from running Two-Stage Least Squares with relative distance to Telesecondaries as the excluded instrument. All parameters are precisely estimated and statistically significant at conventional levels of significance. We find that the first policy causes a 0.360 standard deviation increase in math scores and a 0.242 standard deviation increases in Spanish scores in the seventh grade. The second policy also causes improvements, but they are smaller, 0.223 standard deviations for math and 0.164 standard deviations for Spanish, owing to the less dramatic nature of the policy. The IV estimates, 0.300 for math and 0.173 for Spanish, lie in the middle of the treatment effects of the two counterfactual policies.

6 Conclusion

In this paper we show that distance learning, as it has been conducted in Mexico over the last several decades, is highly effective in raising academic achievement relative to traditional Mexican secondary schools. We find evidence of considerable heterogeneity in value-added in math and Spanish at Telesecondary schools, but that nearly all students benefit from Telesecondary attendance. The gains are large: over a 0.3 standard deviation increase in math scores and a 0.2 standard deviation increase in Spanish scores over the course of a single year. Our counterfactual simulations suggest that further expansions would yield positive academic dividends, as well. The size of the treatment effects corresponding to these policies varies with the degree to which the policy makes Telesecondary schools more accessible.

Telesecondary schools have long been seen as inferior to traditional schools in Mexican media and popular discourse, in part due to the disadvantaged population they serve. We hope that the information provided in this paper highlights the unique role Telesecondary schools play in improving student learning and can change this perception. We predict that future expansions of Telesecondary schools would raise academic achievement in Mexico. As secondary school is an input to subsequent levels of schooling and because there are such large academic gains to a single year of Telesecondary schooling, there are likely to be significant long-term gains to a policy that expands access to Telesecondaries.

References

Beg, Sabrin A, Adrienne M Lucas, Waqas Halim, and Umar Saif. 2019. “Beyond the Basics: Improving Post-primary Content Delivery through Classroom Technology.” Technical Report, National Bureau of Economic Research.

Behrman, Jere R, Susan W Parker, and Petra Todd. 2020. “Impacts of Prospera on Enrollment, School Trajectories and Learning.” *Working Paper*.

Bianchi, Nicola, Yi Lu, and Hong Song. 2020. “The Effect of Computer-Assisted Learning on Students’ Long-Term Development.” *Available at SSRN 3309169*.

Björklund, Anders, and Robert Moffitt. 1987. “The Estimation of Wage Gains and Welfare Gains in Self-Selection Models.” *The Review of Economics and Statistics*, pp. 42–49.

Brinch, Christian N, Magne Mogstad, and Matthew Wiswall. 2017. “Beyond LATE with a Discrete Instrument.” *Journal of Political Economy* 125 (4): 985–1039.

Bulman, George, and Robert W Fairlie. 2016. “Technology and Education: Computers, Software, and the Internet.” In *Handbook of the Economics of Education*, Volume 5, 239–280. Elsevier.

Cameron, Stephen V, and Christopher Taber. 2004. “Estimation of Educational Borrowing Constraints using Returns to Schooling.” *Journal of Political Economy* 112 (1): 132–182.

Card, David. 1995. “Using Geographical Variation in College Proximity to Estimate the Return to Schooling.” *Aspects of Labor Market Behavior: Essays in Honor of John Vanderkamp*, pp. 201–222.

Carneiro, Pedro, and James J Heckman. 2002. “The Evidence on Credit Constraints in Post-Secondary Schooling.” *The Economic Journal* 112 (482): 705–734.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil. 2011. “Estimating Marginal Returns to Education.” *American Economic Review* 101 (6): 2754–2781.

Carneiro, Pedro, Michael Lokshin, and Nithin Umapathi. 2017. "Average and Marginal Returns to Upper Secondary Schooling in Indonesia." *Journal of Applied Econometrics* 32 (1): 16–36.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. 2018. "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance." *Journal of Political Economy* 126 (6): 2356–2409.

Currie, Janet, and Enrico Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Vapital: Evidence from College Openings." *The Quarterly journal of economics* 118 (4): 1495–1532.

Escueta, Maya, Vincent Quan, Andre Joshua Nickow, and Philip Oreopoulos. 2017. "Education Technology: An Evidence-based Review." Technical Report, National Bureau of Economic Research.

Estrada, Rafael Quiroz. 2003. "Telesecundaria: Los Estudiantes y los Sentidos Que Atribuyen a Algunos Elementos del Modelo Pedagógico." *Revista Mexicana de Investigación Educativa* 8, no. 17.

Fabregas, Raissa. 2019. "Broadcasting Education: The Long-term Effects of Mexico's Telesecundarias." Technical Report, Working Paper.

Fan, Jianqing, and Irene Gijbels. 1996. *Local Polynomial Modelling and its Applications: Monographs on Statistics and Applied Probability*. Volume 66. CRC Press.

Heckman, James J, Daniel Schmierer, and Sergio Urzua. 2010. "Testing the Correlated Random Coefficient Model." *Journal of econometrics* 158 (2): 177–203.

Heckman, James J, and Edward J Vytlacil. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences* 96 (8): 4730–4734.

———. 2001a. "Local Instrumental Variables." In *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, edited by K. Morimune C. Hsiao and J. L. Powell, 1–46. New York: Cambridge University Press.

———. 2001b. "Policy-relevant Treatment Effects." *American Economic Review* 91 (2): 107–111.

———. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica* 73 (3): 669–738.

———. 2007. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments." *Handbook of econometrics* 6:4875–5143.

Imbens, Guido W, and Joshua D Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica: Journal of the Econometric Society*, pp. 467–475.

Johnston, Jamie, and Christopher Ksoll. 2017. "Effectiveness of Interactive Satellite-Transmitted Instruction: Experimental Evidence from Ghanaian Primary Schools. CEPA Working Paper No. 17-08." *Stanford Center for Education Policy Analysis*.

Kane, Thomas J, and Cecilia Elena Rouse. 1995. "Labor-Market Returns to Two- and Four-year College." *The American Economic Review* 85 (3): 600–614.

Kling, Jeffrey R. 2001. "Interpreting Instrumental Variables Estimates of the Returns to Schooling." *Journal of Business & Economic Statistics* 19 (3): 358–364.

Martinez Rizo, F. 2005. "La Telesecundaria Mexicana. Desarrollo y Problematica Actual." *Coleccion de Cuadernos de Investigacion*, no. 16.

Navarro-Sola, Laia. 2019. "Secondary School Expansion through Televised Lessons: The Labor Market Returns of the Mexican Telesecundaria." Technical Report, Working paper.

Parker, Susan W, and Petra E Todd. 2017. "Conditional Cash Transfers: The Case of Progresa/Oportunidades." *Journal of Economic Literature* 55 (3): 866–915.

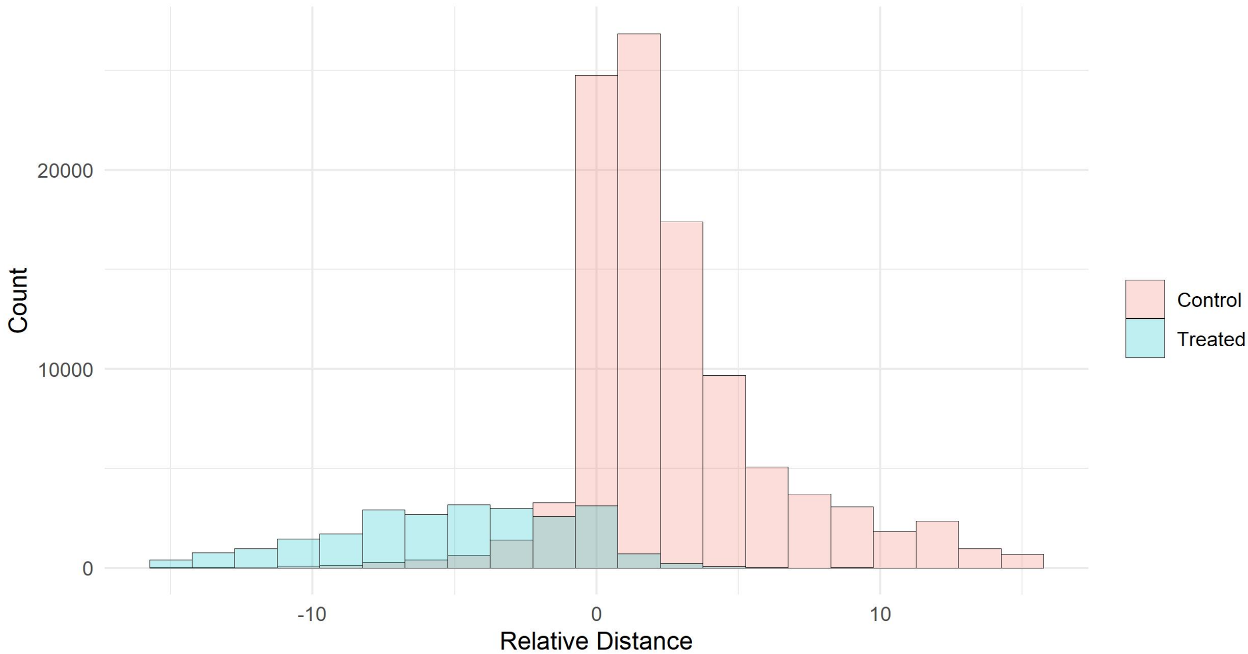
Robinson, Peter M. 1988. "Root-N-consistent Semiparametric Regression." *Econometrica: Journal of the Econometric Society*, pp. 931–954.

Rodriguez-Segura, Daniel. 2020. "Educational Technology in Developing Countries: A Systematic Review."

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of educational Psychology* 66 (5): 688.

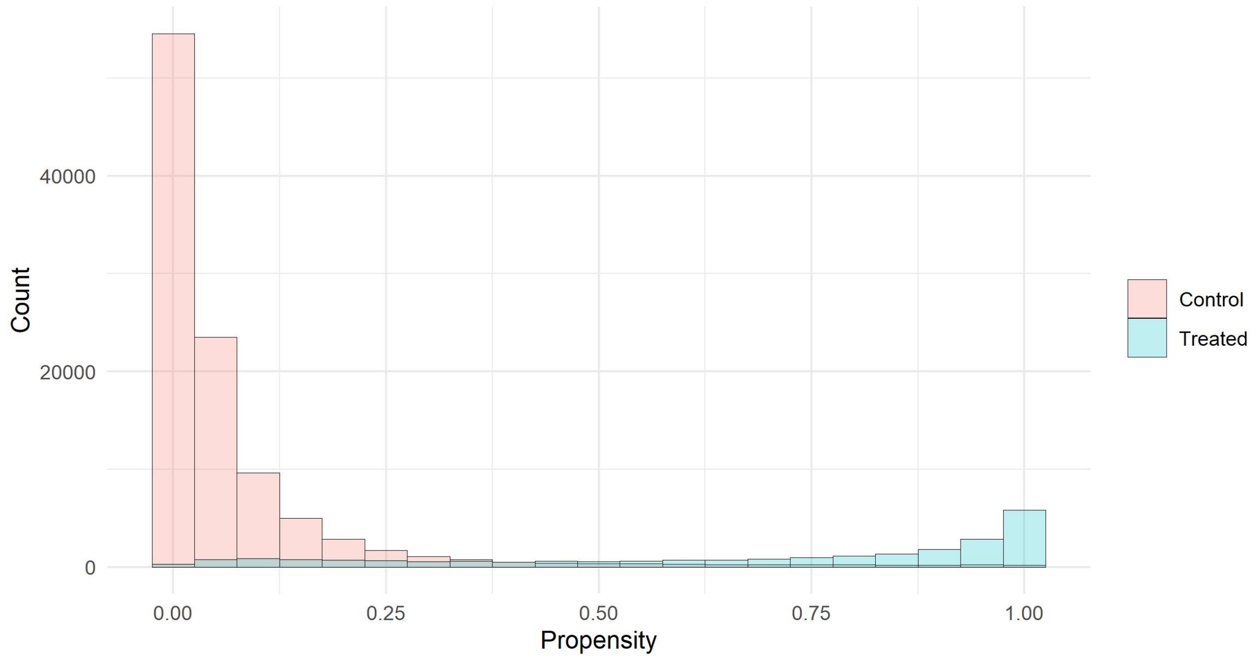
Todd, Petra E, and Kenneth I Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *The Economic Journal* 113 (485): F3–F33.

Figure 1: Relative Distance to Telesecondary School



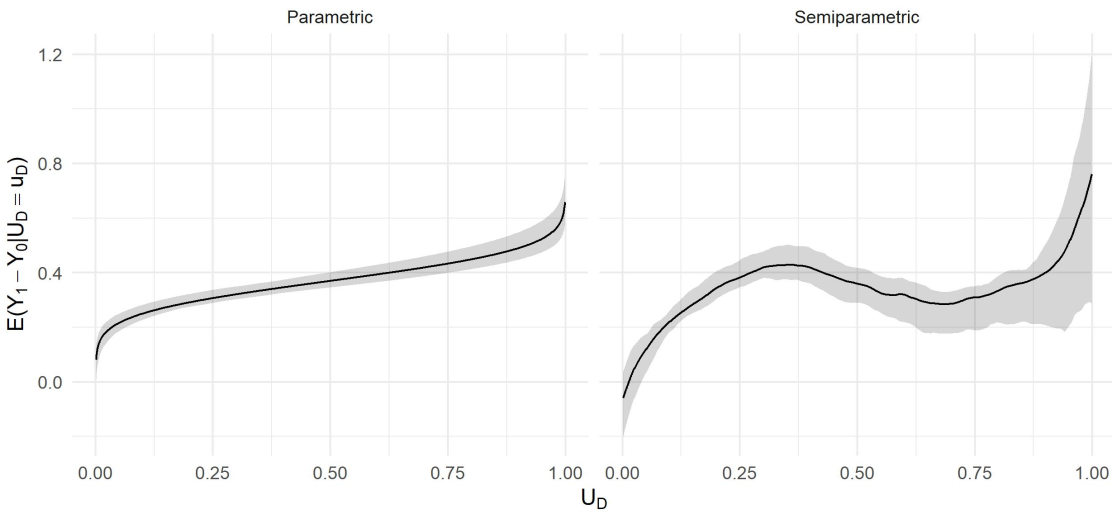
The figure plots histograms of the instrumental variable by treatment status. Control units refer to students in traditional schools, while treated units refer to students in Telesecondary schools. The instrument is the difference between two measures of distance. The first is the distance from the student's primary school to the nearest Telesecondary school, while the second is the distance from the student's primary school to the nearest traditional school. Relative distance is negative whenever Telesecondary schools are closer.

Figure 2: Estimated Propensity Score by treatment status



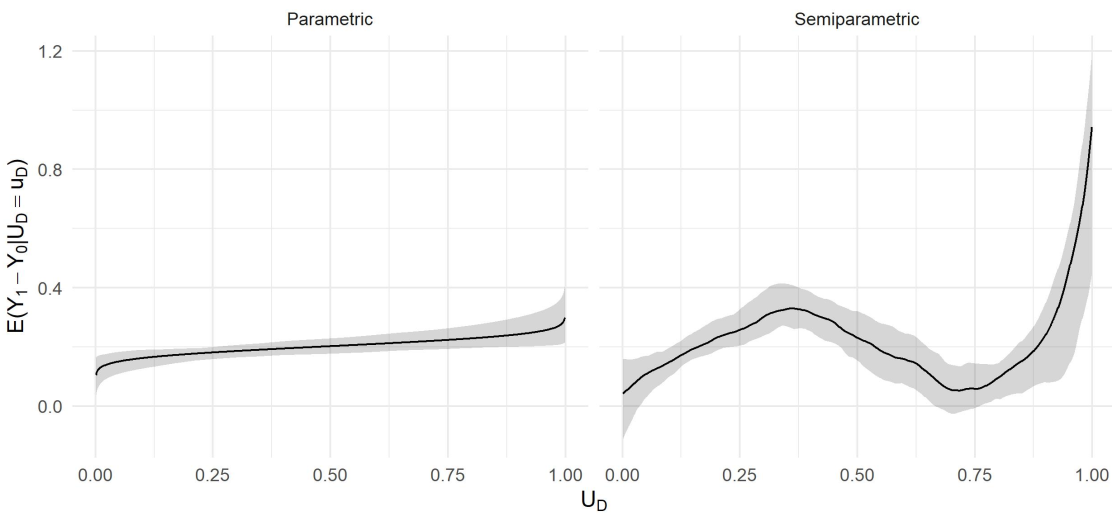
The figure plots histograms of estimated propensity scores by treatment status. Control units refer to students in traditional schools, while treated units refer to students in Telesecondary schools. Propensity scores model Telesecondary attendance as a function the child's sixth grade math and Spanish scores, age, sex, number of siblings, the mother's education, family income, number of books in the home, family access to a computer, Prospera status, rural residence, residence in a Northern state, and the relative distance between the nearest Telesecondary and nearest traditional school. Relative distance is the difference between two measures of distance. The first is the distance from the student's primary school to the nearest Telesecondary school, while the second is the distance from the student's primary school to the nearest traditional school. The propensity score model is estimated via Probit. The figure shows that there is common support: the distribution of estimated propensity scores for both treated and control units is the entire $[0, 1]$ interval.

Figure 3: Marginal Treatment Effect: Math



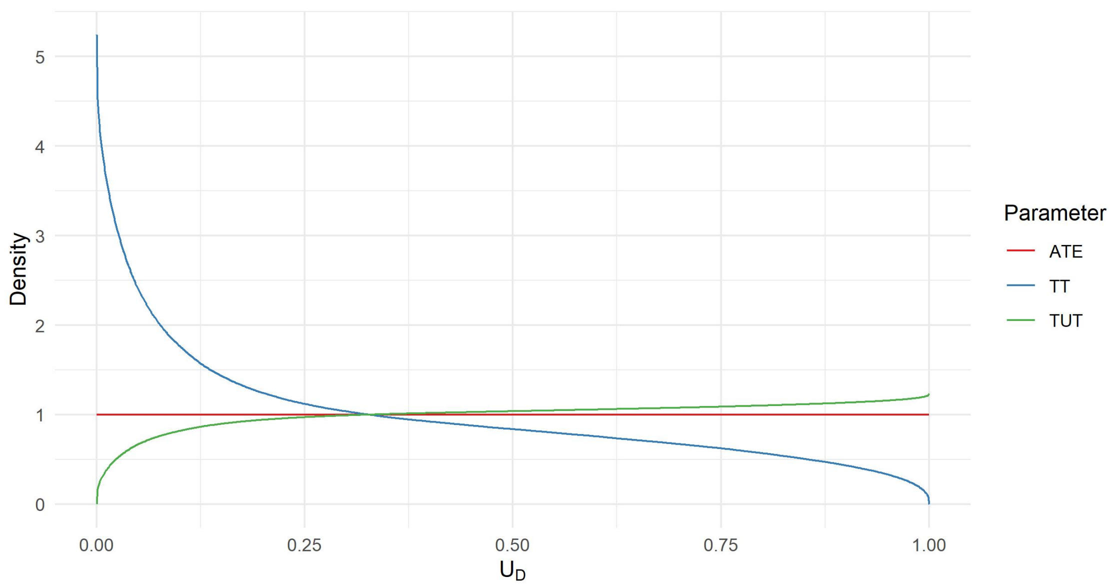
The dependent variable in the outcome equation is the score on the seventh grade nationally-standardized (ENLACE) math exam. The outcome variable has been standardized and is measured in standard deviations from the mean. The outcome equations include controls for sixth grade math and Spanish scores, the age, gender, and number of siblings of the child, family income, mother's education, the number of books in the home, whether the family has access to a computer at home, and dummies for rural residence and residence in a Northern state. The school choice model includes the same controls and also includes the relative distance between the nearest Telesecondary and nearest general secondary school as an exclusion restriction. The school choice model is estimated using via Probit. The parametric MTE is estimated using a two-step Least Squares method that controls for selection on unobservables. The semiparametric MTE is estimated using Local Quadratic Regression and an Epanechnikov kernel with a bandwidth of 0.28. The bandwidth is chosen to minimize the Integrated Mean Square Error in the final stage of estimation. Both MTEs are evaluated at the mean value of the covariates, $X = \overline{x}$. Confidence intervals are computed from boostrapping using 50 draws.

Figure 4: Marginal Treatment Effect: Spanish



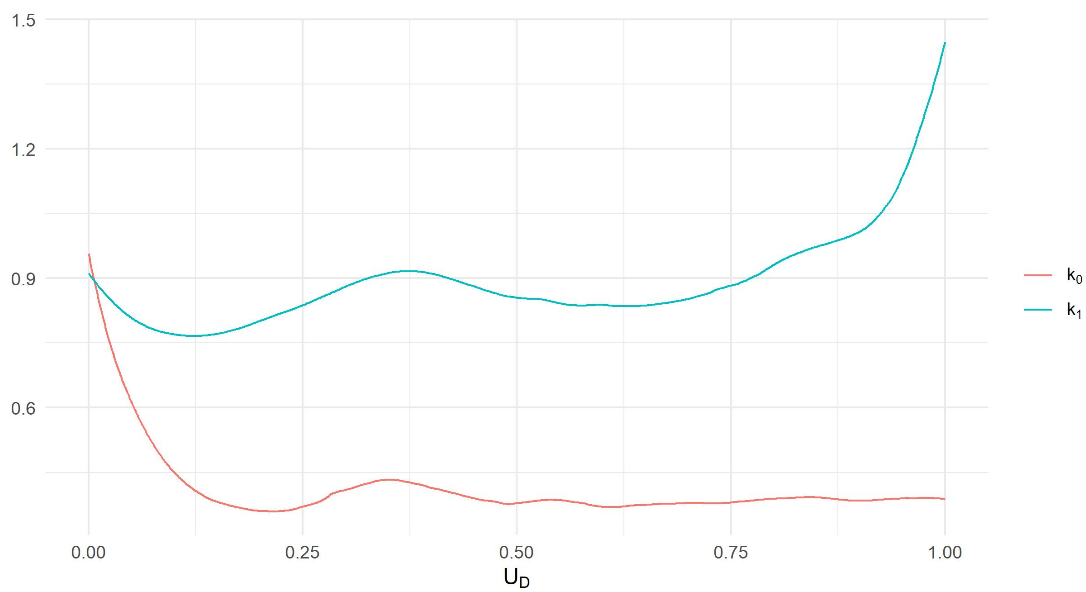
The dependent variable in the outcome equation is the score on the seventh grade nationally-standardized (ENLACE) Spanish exam. The outcome variable has been standardized and is measured in standard deviations from the mean. The outcome equations include controls for sixth grade Math and Spanish scores, the age, gender, and number of siblings of the child, family income, mother's education, the number of books in the home, whether the family has access to a computer at home, and dummies for rural residence and residence in a Northern state. The school choice model includes the same controls and also includes the relative distance between the nearest Telesecondary and nearest general secondary school as an exclusion restriction. The school choice model is estimated via Probit. The parametric MTE is estimated using a two-step Least Squares method that controls for selection on unobservables. The semiparametric MTE is estimated using Local Quadratic Regression and an Epanechnikov kernel with a bandwidth of 0.28. The bandwidth is chosen to minimize the Integrated Mean Square Error in the final stage of estimation. Both MTEs are evaluated at the mean value of the covariates, $X = \overline{x}$. Confidence intervals are computed from boostrapping using 50 draws.

Figure 5: Treatment Parameter Weights



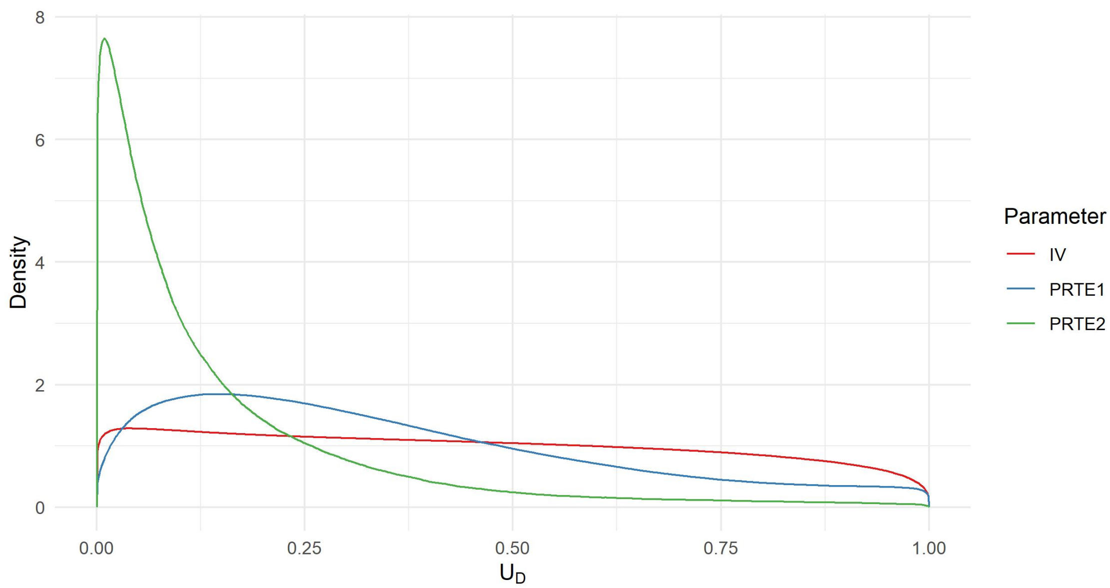
The figure shows the distribution of weighting functions used to construct three standard treatment parameters. The average treatment effect (ATE) integrates the MTE with respect to the unit uniform distribution. The average effect of treatment on the treated (TT) integrates the MTE with respect to the distribution of U_D conditional on attendance in Telesecondaries, $f_{U_D,X|D=1}(x, u_D \mid D = 1)$, while the average effect of treatment on the untreated (TUT) integrates the MTE with respect to the distribution of U_D conditional on attendance in traditional schools, $f_{U_D,X|D=0}(x, u_D \mid D = 0)$.

Figure 6: The Source of Reverse Selection



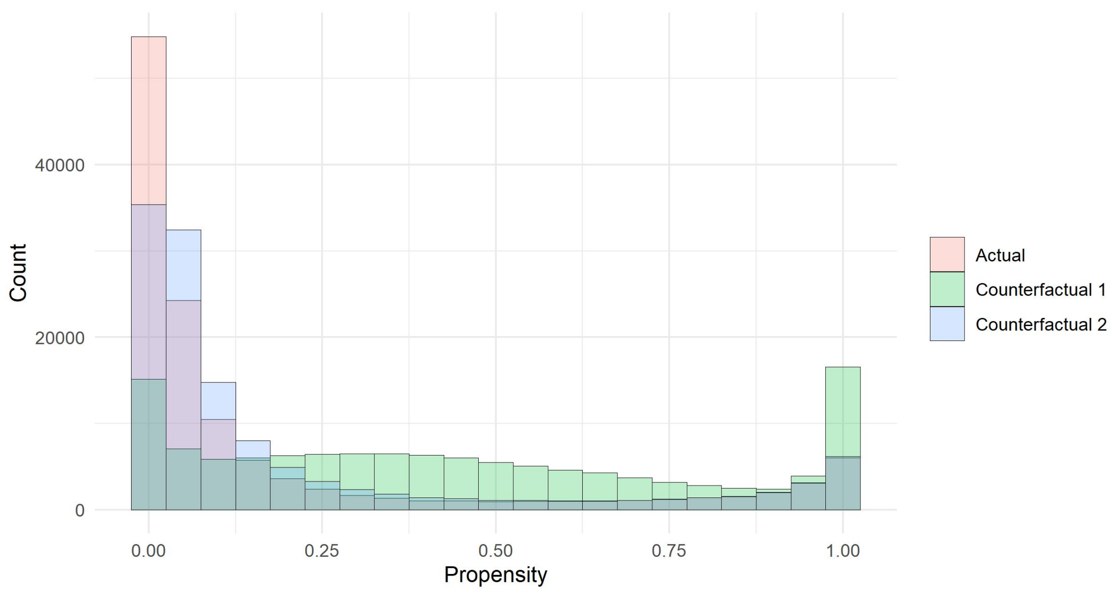
The figure plots $k_1(U_D) = \mathbb{E}[U_1 \mid U_D]$ and $k_0(U_D) = \mathbb{E}[U_0 \mid U_D]$ evaluated at $X = \overline{x}$ on the vertical axis against U_D on the horizontal axis. U_1 is the child's unobserved outcome in the equation for math value-added in Telesecondary schools. U_0 is the child's unobserved outcome in the equation for math value-added in traditional schools. Details on the estimation of $k_1(\cdot)$ and $k_0(\cdot)$ are provided in section 4.3.

Figure 7: Counterfactual Treatment Parameter Weights



The figure shows the distribution of weighting functions used to construct estimates of Policy-Relevant Treatment Effects (PRTEs) for two policies discussed in section 5 as well as the weights induced by Two-Stage Least Squares which uses relative distance as an instrument for Telesecondary attendance. PRTE1 corresponds to the weights induced by a counterfactual policy which reduces relative distance between Telesecondaries and traditional secondary schools by 5 km. PRTE2 corresponds to the weights induced by a counterfactual policy that constructs Telesecondary schools adjacent to all primary schools that do not have a Telesecondary within a 5 km radius.

Figure 8: Histogram of Propensity Scores under Counterfactual Policies



The figure plots histograms of the probability of attending a Telesecondary school under the current policy (2008) as well as two counterfactual policies. Counterfactual 1 reduces relative distance between Telesecondaries and general secondary schools by 5 km. Counterfactual 2 is a school-building policy that constructs Telesecondary schools adjacent to all primary schools that do not have a Telesecondary within a 5 km radius. The counterfactual policies are discussed in greater detail in section 5.

Table 2: Summary Statistics by Middle School Type

	General	Technical	Telesecondary	Dropped
Cohort Size	672,349	397,050	276,395	180,853
Proportion of Cohort	0.44	0.26	0.18	0.12
Mean Math Score (7th Grade)	497	498	492	-
Mean Spanish Score (7th Grade)	500	501	483	-
Mean Math Score (6th Grade)	528	532	483	457
Mean Spanish Score (6th Grade)	524	527	474	454
Fraction Female	0.50	0.50	0.50	0.47
Mean Age (2008)	11.9	11.9	12.2	12.8
Fraction Prospera	0.16	0.18	0.66	0.33

The table displays characteristics of students who attend each of three secondary school types – General, Technical, and Telesecondary – as well as students who drop out. Based on information presented in this table, we consider General and Technical schools as a single alternative for the purposes of estimating value-added between the sixth and seventh grades.

Table 3: Summary Statistics

	Telesecondary		Traditional	
	Mean	S.D.	Mean	S.D.
Math Score (7th grade)	-0.042	(0.984)	0.021	(0.980)
Spanish Score (7th grade)	-0.147	(0.976)	0.032	(0.970)
Math Score (6th grade)	-0.238	(0.915)	0.154	(0.972)
Spanish Score (6th grade)	-0.290	(0.879)	0.182	(0.941)
Relative Distance	-5.078	(4.128)	2.668	(3.495)
Age	12.177	(0.815)	11.920	(0.608)
Siblings	3.697	(2.407)	2.407	(1.672)
Prospera	0.652	(0.476)	0.161	(0.368)
Female	0.506	(0.500)	0.508	(0.500)
Computer at Home	0.126	(0.332)	0.417	(0.493)
Rural Residence	0.674	(0.469)	0.090	(0.287)
Northern State	0.051	(0.219)	0.296	(0.457)
Books in the Home : ≤ 10	0.707	(0.455)	0.471	(0.499)
Books in the Home : 20	0.178	(0.383)	0.259	(0.438)
Books in the Home : 50	0.067	(0.250)	0.156	(0.363)
Books in the Home : ≥ 100	0.047	(0.212)	0.114	(0.318)
Mother's Education : Primary	0.760	(0.427)	0.414	(0.493)
Mother's Education : Middle	0.187	(0.390)	0.282	(0.450)
Mother's Education : Secondary	0.040	(0.196)	0.233	(0.423)
Mother's Education : Postsecondary	0.013	(0.113)	0.071	(0.257)
Income (Pesos/mo) : ≤ 2500	0.579	(0.494)	0.230	(0.421)
Income (Pesos/mo) : 2500-2999	0.263	(0.440)	0.301	(0.459)
Income (Pesos/mo) : 3000-7499	0.116	(0.320)	0.315	(0.465)
Income (Pesos/mo) : ≥ 7500	0.041	(0.198)	0.153	(0.360)

The table displays summmary statistics on outcome variables, covariates, and the instrument – relative distance – for students in the sample. The statistics are broken down by the type of secondary school attended.

Table 4: Propensity Score Model

	Average Derivative	Standard Error
Relative Distance	-0.033	0.000
Math Score (6th Grade)	-0.006	0.001
Spanish Score (6th Grade)	-0.011	0.001
Age	0.016	0.001
Siblings	0.004	0.000
Female	0.003	0.001
Prospera	0.033	0.002
Family Income : Low	-0.013	0.002
Family Income : Medium	-0.020	0.002
Family Income : High	-0.023	0.003
Mother's Education : Middle	-0.018	0.002
Mother's Education : Secondary	-0.033	0.003
Mother's Education : Post-Secondary	-0.012	0.004
Books in the Home : 20	-0.010	0.002
Books in the Home : 50	-0.017	0.002
Books in the Home : $\geq$100	-0.009	0.003
Computer	-0.024	0.002
Rural Residence	0.021	0.002
Northern State	-0.050	0.003

The table shows the average marginal effects of each variable in the propensity score model for Telesecondary attendance. Relative distance is the instrument, and it is computed as the difference between two distance measures. The first is the distance from the student's primary school to the nearest Telesecondary school, while the second is the distance from the student's primary school to the nearest traditional school. Relative distance is negative whenever Telesecondary schools are closer. All other variables are included in the outcome models for seventh grade test scores. The omitted category in each of Family Income, Mother's Education, and Books in the Home is the lowest one. Computer is a binary variable that equals one if the student has access to a computer at home. Standard errors are calculated via 250 bootstrap replications.

Table 5: Estimated Treatment Effects: Math

	Parametric		**Semiparametric**	
	Estimate	Standard Error	Estimate	Standard Error
Average Treatment Effect	0.37	(0.0177)	0.342	(0.0227)
Treatment on the Treated	0.317	(0.0142)	0.279	(0.0147)
Treatment on the Untreated	0.383	(0.0191)	0.356	(0.0267)

The table displays three treatment parameters corresponding to the effect of Telesecondary attendance on seventh grade Math scores, measured in standard deviations. The three treatment parameters are obtained by integrating the MTE with respect to the densities displayed in Figure 5. The simulation method of Carneiro, Lokshin, and Umapathi (2017) is used to integrate the semiparametric MTE. Standard errors are obtained through 50 bootstrap replications.

Table 6: Estimated Treatment Effects: Spanish

	Parametric		**Semiparametric**	
	Estimate	Standard Error	Estimate	Standard Error
Average Treatment Effect	0.202	(0.0159)	0.218	(0.0206)
Treatment on the Treated	0.185	(0.0129)	0.168	(0.0194)
Treatment on the Untreated	0.207	(0.0172)	0.229	(0.0239)

The table displays three treatment parameters corresponding to the effect of Telesecondary attendance on seventh grade Spanish scores, measured in standard deviations. The three treatment parameters are obtained by integrating the MTE with respect to the densities displayed in Figure 5. The simulation method of Carneiro, Lokshin, and Umapathi (2017) is used to integrate the semiparametric MTE. Standard errors are obtained through 50 bootstrap replications.

Table 7: Tests for Selection on Unobservables: Math

Range of U_D for $LATE^j$	(0,0.04)	(0.08,0.12)	(0.16,0.2)	(0.24,0.28)	(0.32,0.36)	(0.4,0.44)
Range of U_D for $LATE^{j+1}$	(0.08,0.12)	(0.16,0.2)	(0.24,0.28)	(0.32,0.36)	(0.4,0.44)	(0.48,0.52)
Difference in LATEs	0.196	0.101	0.0712	0.0373	-0.0214	-0.0456
p-value	0.000	0.000	0.000	0.200	0.500	0.120
Range of U_D for $LATE^j$	(0.48,0.52)	(0.56,0.6)	(0.64,0.68)	(0.72,0.76)	(0.8,0.84)	(0.88,0.92)
Range of U_D for $LATE^{j+1}$	(0.56,0.6)	(0.64,0.68)	(0.72,0.76)	(0.8,0.84)	(0.88,0.92)	(0.96,1)
Difference in LATEs	-0.0411	-0.0305	0.0165	0.043	0.0529	0.249
p-value	0.200	0.420	0.640	0.440	0.520	0.120
Joint p-value	0.080					

The table shows the results of tests for equality of LATEs for math value-added defined by adjacent and non-overlapping regions of the domain of U_D. Given an interval $[L_j, H_j]$, the LATE for that interval is given by $LATE^j = \mathbb{E}[Y_1 - Y_0 | X = \bar{X}, L_j \leq U_D < H_j]$, which is simply the average of the MTE between L_j and H_j evaluated at $X = \bar{x}$. p-values test the hytpothesis that the difference between adjacent LATEs is equal to zero. p-values are obtained through 50 bootstrap replications.

Table 8: Tests for Selection on Unobservables: Spanish

Range of U_D for $LATE^j$	(0,0.04)	(0.08,0.12)	(0.16,0.2)	(0.24,0.28)	(0.32,0.36)	(0.4,0.44)
Range of U_D for $LATE^{j+1}$	(0.08,0.12)	(0.16,0.2)	(0.24,0.28)	(0.32,0.36)	(0.4,0.44)	(0.48,0.52)
Difference in LATEs	0.0804	0.0653	0.0504	0.0595	-0.0205	-0.0709
p-value	0.140	0.020	0.120	0.020	0.560	0.060
Range of U_D for $LATE^j$	(0.48,0.52)	(0.56,0.6)	(0.64,0.68)	(0.72,0.76)	(0.8,0.84)	(0.88,0.92)
Range of U_D for $LATE^{j+1}$	(0.56,0.6)	(0.64,0.68)	(0.72,0.76)	(0.8,0.84)	(0.88,0.92)	(0.96,1)
Difference in LATEs	-0.0651	-0.0673	-0.0426	0.0612	0.123	0.462
p-value	0.100	0.020	0.200	0.300	0.10	0.00
Joint p-value	0.000					

The table shows the results of tests for equality of LATEs for Spanish value-added defined by adjacent and non-overlapping regions of the domain of U_D. Given an interval $[L_j, H_j]$, the LATE for that interval is given by $LATE^j = \mathbb{E}[Y_1 - Y_0 | X = \bar{X}, L_j \leq U_D < H_j]$, which is simply the average of the MTE between L_j and H_j evaluated at $X = \bar{x}$. p-values test the hytpothesis that the difference between adjacent LATEs is equal to zero. p-values are obtained through 50 bootstrap replications.

Table 9: Counterfactual Treatment Effects

	PRTE1	PRTE2	IV
Math	0.360	0.223	0.300
	(0.0177)	(0.0325)	(0.0145)
Spanish	0.242	0.164	0.173
	(0.0179)	(0.0326)	(0.0147)

The table displays treatment parameters corresponding to two counterfactual policies discussed in section 5. PRTE1 corresponds to a counterfactual policy that reduces relative distance between Telesecondaries and general secondary schools by 5 km. PRTE2 corresponds a counterfactual policy that constructs Telesecondary schools adjacent to all primary schools that do not have a Telesecondary within a 5 km radius. PRTE1 and PRTE2 are calculated using the semiparametric MTE combined with the simulation method of Carneiro et al (2016). Standard errors for these two treatment parameters are displayed in parentheses and are obtained through 50 bootstrap replications. The IV estimates are obtained by a Two-Stage Least Squares regression that uses relative distance as an instrument for Telesecondary attendance.