ENROLLMENT, LABOR, AND EFFORT: AN ANALYSIS OF THE EDUCATIONAL CHOICES OF STUDENTS IN MEXICO

Gabrielle Vasey

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2021

Supervisor of Dissertation

Petra E. Todd
Professor of Economics

Graduate Group Chairperson

Jesús Fernández-Villaverde
Professor of Economics

Dissertation Committee

Jere R. Behrman, Professor of Economics

Holger Sieg, Professor of Economics

Arthur van Benthem, Associate Professor of Business Economics and Public Policy

ENROLLMENT, LABOR, AND EFFORT: AN ANALYSIS OF THE EDUCATIONAL CHOICES OF STUDENTS IN MEXICO

ACKNOWLEDGEMENT

There are many people who have been a part of my life over the past six years that I would like to thank. I will start by thanking the Economics faculty at Penn, who have been so knowledgeable, and provided many helpful comments during seminars and meetings. Petra Todd has given me invaluable guidance and advice throughout my PhD, and is one of my role models. Thank you for all of your time and for giving it so generously. Thank you Jere Behrman for bringing equal measures of wisdom and humor to all of our meetings. Thank you Holger Sieg for your thoughtful comments on my research, and for always making me take a step back and consider the bigger picture. Finally, thank you Arthur van Benthem and Mike Abito for involving me in my first PhD research experience. I have learned so much, not only about research, but also about how academics works, from you both.

I have met so many people during the past six years who have made Philly feel like home and the PhD experience so much fun. I feel so fortunate to call you friends. Thank you Alejandro, Camille, Emilio, Ewelina, Felipe, Görkem, Hanna, Irina, João, Marianne, Michal, Montse, Paul, Sergio, Tomás, and many others. I look forward to our next phase of being friends and colleagues, and to sharing future successes.

Lastly, thank you to my family, and my friends who knew me before this adventure began. A lot has changed during this time, and I am so grateful for you being on this journey with me. A special thanks to my mom, dad, and Owen. Your support has been instrumental in getting me here.

ABSTRACT

ENROLLMENT, LABOR, AND EFFORT: AN ANALYSIS OF THE EDUCATIONAL CHOICES OF STUDENTS IN MEXICO

Gabrielle Vasey

Petra E. Todd

This thesis consists of two chapters and examines questions centered around the educational choices of students in Mexico.

When school-age children work, their education competes for their time and effort, which may lead to lower educational attainment and academic achievement. Chapter 1 develops and estimates a model of student achievement in Mexico, in which students make decisions on school enrollment, study effort and labor supply, taking into account local schooling options and wages. These decisions affect their academic achievement in math and Spanish, which is modeled using a value-added framework. The model is a random utility model over discrete school-work alternatives, where study effort is the outcome of an optimization problem under each alternatives. The model is estimated using an administrative test score database on Mexican 6th grade students combined with survey data on students, parents and schools, geocode data on school locations, and wage data from the Mexican census. The empirical results show that if students were prohibited from working while in school, the national dropout rate would increase by approximately 20%, while achievement would increase in math and Spanish. Expanding the conditional cash transfer, either in the magnitude of the benefits or the coverage, in conjunction with prohibiting working while in school is an operational policy that would greatly reduce dropout while maintaining

the achievement gains.

In Chapter 2, my coauthor Emilio Borghesan and I analyze 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 telesecundarias relative to traditional secondary schools. The estimated MTEs reveal that school choice is not random, and that the effect of telesecundaria 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 attending a telesecundaria. We conclude by estimating the effects of counterfactual policies that expand telesecundaria availability and find that they generate improvements in academic performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1 : The Impact of Child Labor on Student Enrollment, Effort and Achievement: Evidence from Mexico

1.1. Introduction

When children participate in the labor force, it is often at the expense of their education. Globally, the International Labour Organization estimated that 144 million children under the age of 14 were working in 2012. The trade-off between working, with the benefits of receiving a wage or helping family, and attending school, in the hope of increasing future wages, is one that many children and families face worldwide. Children who attend school may also work part time and face another choice with respect to the amount of time and effort to dedicate to studying compared to working. Often, laws prohibiting child labor and requiring school enrollment exist, but they are not well enforced. Family socioeconomic status, school availability, school quality, ability and earnings opportunities all influence children's time allocation decisions and their resulting academic achievement and attainment.

This paper explores the relationship between child labor, school enrollment and academic achievement in Mexico, and analyzes the impacts of enforcing policies related to labor and education laws. I consider children who graduated from primary school (Grade 6) and who should be enrolling in middle school (Grade 7). Mexican Basic Education, defined as Grades 1 through 9, is compulsory and labor of minors under the age of 14 is legally prohibited. However, in the 2010 Census, 7.9% of children aged 12 and 13 report not enrolling in school. A nationally representative survey in 2009 found that 25.7% of Grade 7 students who are in school report working at least one day a week. Many developing countries around the world face similar struggles to keep children in school and out of the labor force.

My analysis includes not only the students who go to school full time or work full time, but also the students who combine school and work and the repercussions that working has on their academic achievement. Incorporating this more nuanced choice set is ideal, however this is one of the few empirical studies that includes these choices, which highlights how challenging it is to acquire the required data and setting. The existing studies that do allow for students to combine work and school do not consider the how this impacts academic achievement (Bourguignon, Ferreira, and Leite, 2003; Leite, Narayan, and Skoufias, 2015). I use individual-level data on school enrollment, test scores, demographics, labor choices, and effort choices to identify the parameters that define the tradeoffs that students are facing.

To study the determinants of children's time allocation decisions, I develop and estimate a model of school and labor participation decisions with endogenous school effort choices. In my model, individuals who finish primary school have a choice set of middle schools available. The choice set is determined using data on school locations and prior-year school attendance patterns. The middle schools are treated as differentiated products that vary in terms of school infrastructure and principal characteristics such as experience, as well as the type of school curriculum. The choice of school affects a student's utility directly, as well as their achievement production function and marginal cost of effort. Effort is costly, and the marginal cost of effort depends on student characteristics and on whether the student is working. Wage offers vary by student demographics and by primary school location and there are separate wage offers for working full time and working while enrolled in school.

To estimate my model, I combine several data sources: administrative data on nationwide standardized tests in math and Spanish, survey data from students, parents and principals, geocode data on school locations, and Mexican census data on local labor market wages and hours worked. Combined, these data sources create an incredibly rich dataset that has not been used by any studies to date. The administrative data tracks all students in Mexico as they complete national standardized tests and includes information on which students are beneficiaries of the conditional cash transfer Prospera. The data has been used by several recent studies to analyze the impact of Prospera on achievement (Acevedo, Ortega, and Székely, 2019; Behrman, Parker, and Todd, 2020).

In the model, students decide whether to attend school, and if they attend, they also decide what type of school to attend and how much effort to dedicate to their studies. The marginal cost of effort varies by age, gender, parental education, lagged test scores, and working status. I use the model's first-order conditions to solve for an optimal effort level that is specific to each type of school. The data provide five measures of self-reported effort, which I use within a factor model to obtain a single effort index.

The model is a discrete-continuous choice model with partially latent continuous choice variables (Dubin and McFadden, 1984). Specifically, it is a random utility model over discrete school-work alternatives, where study effort is determined as the outcome of an optimization problem under each of the school-work alternatives. Achievement is modeled using value-added equations that incorporate student's effort choices. I estimate the model via Maximum Likelihood, where the probability can be decomposed into three conditional probabilities, which each have a closed-form solution.

The identification of the parameters of interest relies mainly on geographic variation of exogenous market conditions and choice sets. To identify the value of school, I use variation in the distances required to travel to school and to identify the value of working, I use variation in local wage offers. Effort is a key mechanism in the model, and the parameters related to the cost and productivity of effort are identified using self-reported effort measures from students.

I find that traveling to a middle school is costly and that students value distance education schools (Telesecundarias) less than the other two school types (General and Technical). Students value schools with high average expected test scores, however the amount of weight they put on that component does not depend on their parents education levels or on whether they are conditional cash transfer beneficiaries. Effort is costly to students, especially when working, but less so for female students and for students with higher lagged test scores. Students are estimated to dislike working while in school overall. Effort is estimated to be an important input into both math and Spanish test score production functions.

I use the estimated model to evaluate how education and work-related policy changes would affect school enrollment, academic achievement, and children's labor-force participation rate. First, I use the model to simulate the effects of a partial labor law enforcement that removes the labor option for children who are enrolled in school. These estimates provide insight into what fraction of students only attend school if they can also work and how much achievement would increase if students did not divide their time between work and school. The second and third counterfactuals consider policies that would work in conjunction with the first, with the goal of reducing the drop out rate. The second counterfactual considers prohibiting all child labor and the third considers expanding the conditional cash transfer for school attendance, both in terms of benefits amounts and program coverage.

The results of the counterfactual analysis show that almost 10% of students who are working while enrolled in school would drop out if they were unable to combine work and school. This increases the national dropout rate by almost 20%. For students who remain enrolled, their effort increases by approximately 3% of a standard deviation, resulting in increases in their math and Spanish scores by an average of 3% of a standard deviation. Prohibiting all child labor results in a dropout rate lower than under the benchmark model. However, a similarly low dropout rate can be achieved by either increasing the conditional cash transfer amounts, or expanding the set of families who receive the conditional cash transfer to include more of those with low monthly incomes.

The paper proceeds as follows. Section 2 lists related literature and the contribution of this paper. Section 3 describes the dataset and setting and provides summary statistics for the variables of interest. Section 4 describes the model of discrete school-work alternatives with endogenous effort choice. Section 5 describes the estimation strategy and Section 6 discusses the results from the estimation. Section 7 discusses the policy implications and Section 8 concludes.

1.2. Literature

Recently, there have been several papers estimating models of school choice, where schools with differing characteristics are treated as differentiated products (Ferreyra, 2007; Epple, Jha, and Sieg, 2018; Neilson, 2014; Bau, 2019; Neilson, Allende, and Gallego, 2019). These models are similar to mine in that they include school characteristics and a student achievement production function, and the authors use the model to evaluate how policy changes impact school choices. I extend these frameworks by allowing for dropping out of school and part-time or full-time work. I also incorporate students' decisions of how much effort to devote to their studies. These extensions are necessary to make the school choice model relevant to developing country contexts where child labor is prevalent.

A large portion of the literature examining the relationship between child labor and education considers how policies, such as conditional cash transfers, affect school enrollment and child labor.[1] Dynamic models have been used to evaluate the long-term effect of such policies, however none thus far has incorporated test score production functions, time allocation decisions, and decisions about what type of school to attend (Todd and Wolpin, 2006; Attanasio, Meghir, and Santiago, 2011). There also exist some static choice models that include the options of dropping out, enrolling and working part time, or only enrolling (Bourguignon, Ferreira, and Leite, 2003; Leite, Narayan, and Skoufias, 2015). However, these models also do not examine academic achievement or how working part time affects a child's ability to study. Finally, there are some recent papers that consider the impact of labor on achievement, without incorporating school choice and enrollment decisions. Keane, Krutikova, and Neal (2018) consider many possible uses of time for students, and find that working is only harmful to achievement if it is taking away from study time.

Although there is a substantial literature in the education economics field studying teacher effort, how it affects student achievement and how it is influenced by incentive pay, there is relatively little focus on student effort, which is an important input in academic achievement. A study using an instrumental variables approach finds that school attendance has a positive causal impact on achievement for elementary- and middle-school students (Gottfried, 2010). A causal relationship between study time and grades has also been found for college students (Stinebrickner and Stinebrickner, 2008). There are very few papers that model student effort in a structural way, and

[1] There exists a related literature studying the effects of working in highschool or college, and the effects of this on educational outcomes and human capital accumulation (Stinebrickner and Stinebrickner, 2003; Eckstein and Wolpin, 1999; Buscha, Maurel, Page, and Speckesser, 2012; Le Barbanchon, Ubfal, and Araya, 2020).

estimate how it affects learning. Todd and Wolpin (2018) develop and estimate a strategic model of student and teacher efforts within a classroom setting.

The literature on CCT programs, and specifically on the Prospera program, is extensive. The program began in 1997, and since then over 100 papers have been written about it (Parker and Todd, 2017). The majority of these papers use the experimental data gathered during the first two years of the program. There is a consensus in the literature that Prospera increases enrollment in school for students in junior and senior high school (Schultz, 2004; Behrman, Sengupta, and Todd, 2005; Attanasio, Meghir, and Santiago, 2011; Dubois, De Janvry, and Sadoulet, 2012). However, studies focused on student enrollment and grade progression and not on student achievement, with the exception of two recent working papers (Acevedo, Ortega, and Székely, 2019; Behrman, Parker, and Todd, 2020).[2] Finally, there are a few studies using experimental data to estimate the impact of conditional cash transfers on child labor decisions. For example, Yap, Sedlacek, and Orazem (2009) find that the PETI program in Brazil increased academic performance and decreased child labor for beneficiary households.

1.3. Data and Setting

The data requirements for answering research questions related to child labor, school enrollment and academic achievement are high. Among other variables, the labor choice, school choice and achievement realization for each student must be observed. The data set that I use provides the above mentioned variables and more. In addition, there is quasi-random variation, since each primary school has a different choice set of middle schools, as well as a different labor market conditions. Unfortunately, there are some variables that are not available in this setting, most notably information on

[2]These papers use matching and regression-based treatment effect estimators.

parental wages. The absence of these data will inform the modeling choices that I make in the next section.

Not only does the data set provide the majority of the required variables, but Mexico as a country is an ideal setting to study this question. In 2010, the year analyzed in this paper, Mexico had Education Regulation that defined Grades 1 through 9 as compulsory, and Labor Regulation that prohibited labor of minors under the age of 14. However, 7.9% of children age 12 and 13 reported not being enrolled in school in the 2010 Census. Further, over a quarter of students in Grade 7 reported working at least one day a week in a national survey. In addition, increasing student enrollment has been a target for the Mexican government for decades, with programs such as the conditional cash transfer and the distance education schools. These ensure that the majority of students, even in rural areas, have access to a local school if they choose to enroll.

1.3.1. Data Sources

To carry out this research, I use a newly available merged dataset. This dataset is comprised of several components which come from two main sources. The first component is the Evaluación National de Logro Académico en Centros Escolares or ENLACE test scores. These tests were administered at the end of the school year to gather information on students' achievement in math and Spanish. They were given to students every year between the 2006/2007 school year and the 2013/2014 school year. The Mexican Secretariat of Public Education (SEP) was in charge of administering the test. The second component comes from the same source as the ENLACE test scores, and can be easily merged with the test score data. Every year a group of schools was randomly selected and all students enrolled in those schools were given a questionnaire. These data have recently been used for impact evaluation studies of the Prospera program (Acevedo et al., 2019; Behrman et al., 2020). The third component of the data set is comprised of a list of all schools in Mexico, and can be merged with the above data to provide the geographical location of the schools.

The test score data provides important information regarding student achievement, however whether a student took the test or not may not always be an accurate method of recording school attendance. It is possible that a student who is enrolled and attending school does not write the ENLACE test for several reasons. To ensure that these students are recorded as enrolled, even without a test score, I merge the National Student Registry (Registro Nacional de Alumnos) with the test score data. This provides information on enrollment for all students in the country.

Finally, the model requires data on wages, which are not recorded in the previously mentioned source. The 2010 Census is used to access information on children between the ages of 12 and 20, and their working status and wages. The Census also contains other personal information on the students such as their age, gender, school attendance history, parental education, living situation, and the municipality in which they reside.

Combining all of the data from the above sources, yields an incredibly rich representative sample of students across Mexico. For each student, I have their national standardized test scores, their school IDs (with associated school information), individual demographics, household demographics (including conditional cash transfer status), and the average municipal wage conditional on age, gender, family background and school attendance.

1.3.2. Estimation Sample

The analysis in this paper focuses on students across Mexico in Grade 6 in 2008 who progress to Grade 7 in 2009.[3] The sample can be divided into two groups: those who enrolled in school in Grade 7, and those who dropped out of school after Grade 6.[4] There are 229,199 students enrolled in Grade 6 in 2008 for whom I have survey answers from themselves and their parents. Of these students, 17,195, or 7.5% do not appear in either the ENLACE data or the Roster data in any of the next four years. I assume that these students have dropped out of school.

In Grade 7 in 2009, there are 107,898 students for whom we have survey answers from themselves and their parents.[5] The mean age of the students in Grade 7 is 13, and a bar graph showing the distribution of ages is shown in Figure 1. The sample is approximately equal in terms of gender, as 49.9% of the students are female. 26.1% of students are beneficiaries of the conditional cash transfer Prospera.

1.3.3. Key Variables

Test Scores

Standardized test scores in math and Spanish are used as a measure of student achievement. The test administrators (SEP) standardized the tests in the base year, 2008, to have a mean of 500 and a standard deviation of 100. The same transformation was used in subsequent years.[6] All students write the test in Grade 6 and Grade

[3]There are three states (out of 32) that are not included in the analysis. The states of Guerrero, Michoacán, and Oaxaca had many schools for which there were no ENLACE scores submitted. To prevent bias in the analysis, students in these states were not included.

[4]See Appendix A.4 for more details on how the sample for estimating the discrete choice model is constructed.

[5]Each year a different sample of schools is given the questionnaire, so the majority of these students are not in the sample of Grade 6 students from the previous year. Sample size also changes from year to year.

[6]This is not equivalent to standardizing the scores each year, as is apparent from the means presented.

7, so it is possible to see how they change relative to students in the same grade from their baseline results. Table 1 shows the mean and the standard deviation of Grade 6 and 7 test scores in math and Spanish. To compute these statistics, the cohort of Grade 7 students was used.

Labor Decision

To observe the labor decision of the students, I use a question from the student survey which asks: "On average, how many days a week do you work?". Figure 2 shows the responses, divided by gender. Boys work more than girls, and the majority of students are not working. The mean number of days a week worked for the whole sample is 0.83. However for children 13 and younger the mean is 0.80, and for children 14 and older the mean is 1.68, so older children are working substantially more than the younger children.

Although I will not be considering different occupation types in this project, it is of interest to know what kinds of labor children were engaging in during this time period in Mexico. From the Census, the most common occupation type for 12 year olds was agriculture (maize, beans, livestock, flowers, vegetables, fruits) with the next most common being a sales worker or working in a store. Other occupations reported included street vendors, food preparation and a support worker for construction. In the student survey, there is a question inquiring about the reasons for working, and 59% of students reported working for their family.

Student Effort

Achieving a high test score and earning a wage at a job both take time and energy. To capture this, and to understand how combining school and work may impact achievement, I incorporate effort into my analysis. The rich data set provides five self-reported measures related to effort. The questions are:

1. On average, how many hours a day do you spend studying or doing homework outside of school hours? Options: 0, 1, 2, 3, or 4 hours.
2. How often do you pay attention in your classes at school? Options: never, almost never, sometimes, almost always, always.
3. How often do you participate in your classes at school? Options: never, almost never, sometimes, almost always, always.
4. How often do you miss school? Options: never, almost never, sometimes, almost always, always.
5. How often do you skip your classes when you're at school? Options: never, almost never, sometimes, almost always, always.

The first measure, the number of hours studied per day, is cardinal. The other four measures are ordinal variables, as they are answered on a Likert scale. To combine them into one value, I use factor analysis. This analysis is done outside of the model estimation, and uses polychoric correlations to take into account the ordinal variables. I then compute the eigenvalue decomposition of the correlation matrix, and estimate loadings for each of the five variables. The end result is a single value of effort for each student, $\tilde{e}^{M}_{ijL}$, which combines the information from the student's responses to the five effort questions. Figure 3 presents a histogram of the new continuous effort variable. The effort values are almost all positive and the distribution appears to be approximately normal.[7] Estimation details and results are in Appendix A.2.

[7] Negative values of effort are possible, though rare, since the last two effort questions have negative loading factors.

Wages

It is necessary to know what wage each of the children could be earning if they decided to work. Unfortunately, wages are not included in the survey data, so I impute wages for all students using census data. The census contains the working status, enrollment status, and the wages earned for children across Mexico. Other important information such as the age and gender of the child, the education level of their parents, and the municipality in which they live is also recorded.

To account for non-random selection into working, a Heckman selection model is estimated. Two separate equations are estimated, the first with the outcome variable being hourly wages, and the second with the outcome variable being number of hours worked per week. Variables representing family socioeconomic levels, such as family income and home infrastructure are used as instruments that affect selection into working, but do not affect the wage offers directly. Regressions are estimated separately for girls and boys. For details on the wage estimation and parameter estimates, see Appendix A.6.

$$\begin{aligned} w_{igj} = \gamma_0 + \underbrace{\gamma_1 a_i}_{\text{Age}} + \underbrace{\gamma_2 \mathbb{1}\{j \neq 0\}}_{\text{Not enrolled}} + \gamma_3 a_i \times \mathbb{1}\{j \neq 0\} + \underbrace{\gamma_4 MomEduc_i + \gamma_5 DadEduc_i}_{\text{Parental education}} \\ + \gamma_6 a_i \times MomEduc_i + \gamma_7 a_i \times DadEduc_i + \underbrace{Geo_g}_{\text{Municipality FE}} + \nu_{igj} \end{aligned}$$

To compute the monthly wage for each student, their hourly wage is multiplied by their weekly hours, and then by four. The imputed monthly wages will be used for the remainder of this paper. Table 2 contains results from the imputations. The results are divided by gender of the child, and by the school enrollment status. The mean and standard deviation are shown for the hourly wages, the number of hours worked per week, and the monthly wage. The monetary values are in 2010 pesos. For students working full time, the imputed wages for females is 14.2 pesos per hour and for males it is 14.1 pesos per hour. The part time wages for students who are also enrolled in school are slightly lower. During this time period, the minimum wage was 15 pesos per hour, and the results are just below this value, which is feasible given that these were informal jobs. Students working while enrolled in school worked on average 19 hours a week, while those who are working full time work just under 40 hours a week.

School Types

There are four different types of middle schools in Mexico: General, Technical, Telesecundarias, and Private. Technical middle schools have a focus on vocational studies. Telesecundarias, which are wide spread and well established in Mexico, are predominately located in rural areas and offer instruction through video sessions at local centers. The purpose of these schools are to provide access to education for students in rural areas without having to incur the cost of hiring teachers specializing in each subject. Private schools are almost exclusively in urban areas, and have tuition payments. Unfortunately, I was not able to collect information on school tuition, so students attending private schools are not included in the estimation of the model.

Table 3 contains summary statistics for the four different types of schools in Mexico. From the table it is apparent that there are many small telesecundarias, predominately in rural areas. The class size of telesecundarias is also noticeably smaller than both General and Technical schools. Although all schools have a fairly equal amount of female and male students, the proportion of students who are beneficiaries of the conditional cash transfer differs drastically by school type. The majority of students enrolled in a telesecundaria are beneficiaries, while less than 15% of those in General schools are. Finally, by dropping all Private schools, only 8% of students are removed from the sample.

Distances and Choice Sets

The location of each school in the data set is known. With these locations, it is possible to compute the distance between a student's primary school and middle school, and analyze how far students are traveling. Further, it is possible to see what other options were available within a certain distance. Examining the data, it is apparent that middle schools are much more sparse than primary schools, especially in rural regions of Mexico. Figure 4 shows the geographic distribution of primary and middle schools in a region in Mexico. Although there is a small city in the top right corner, the remainder of area covered by the map is rural. Depending on which primary school a student attended, there may be a middle school at the same location, or the nearest one may be several kilometers away.

Unfortunately the home address of students is not included in the data. Given the broad coverage of primary schools, I am assuming that students attend a primary school close to their home, and therefore their primary school address is an adequate proxy for their home address. To calculate distance, a straight line is measured between the primary school and the middle school, as shown in Figure 5. It is also possible to calculate distance using roads and paths on Google Maps, but this does not capture many of the rural pathways.

For the estimation, I have to define which middle schools each student considers when making their school choice. To do this, I create a circle around the primary school and consider all middle schools within the circle to be in the choice set, as shown in Figure 5. However, choosing the same radius for all primary schools would not account for regional topography or the local availability of schools. Therefore, each primary school has a custom radius that is computed by analyzing how far students from that primary school traveled on average to attend middle school in previous years.[8]

1.3.4. Data Patterns

To quantify the impact of enforcing child labor laws on educational attainment and achievement, it is crucial to understand the relationships between working, study effort, achievement, and the many other inputs from the setting. The following subsections describe these patterns and correlations in the data. This is helpful in understanding the research question, and also informative for modelling. Finally, the estimated model should be able to reproduce these patterns and correlations, which is confirmed in the Results section.

Working and Achievement

When discussing working while in school, a main concern is that the test scores of students who are working could suffer. The simple regression shown in Table 4 does show a significant negative relationship between working and Grade 7 test scores. The covariate "Working" is a dummy variable, and in column (1) it is equal to 1 for all students who report working at least 1 day a week, in column (2) it is 1 for all students who report working at least 2 days a week, and so on. This is a descriptive regression, so the results should not be interpreted as causal, however it does control

[8]Distances are capped at 15km to get rid of outliers and students who moved. Students who changed state are also removed from the estimation sample.

for the student's lagged test scores, their gender and age, their parental education levels, and whether they are a beneficiary of the conditional cash transfer. The magnitude of the relationship increases as the number of days a week working increases. Students who work at least 3 days a week are found to have 5% of a standard deviation lower test scores than their peers. This magnitude is equivalent to the increase in test scores that students have if their mother has at least a middle school education.

Working and Study Effort

Although the relationship between working and achievement can be studied directly, it is more informative to investigate the underlying mechanisms. The main mechanism that comes to mind that connects both work and academic achievement is time and energy. For this project, I will bundle them together, and call the overall measure study effort (the variable is described in detail in Section 1.3.3). If students are working, they have another use of their time. Table 5 shows that there is a negative correlation between working at least one day a week and the effort variable. The correlation decreases in magnitude as more controls are added, however the significance remains. In the final column, students who work at least one day a week have 5% of a standard deviation lower effort values than students who are not working.

Study Effort and Achievement

For study effort to be a valid mechanism between working and achievement, there must also be a correlation between effort and achievement in the data. Figure 6 shows that higher test scores in both math and Spanish are correlated with higher values of effort (without controlling for any covariates). While controlling for variables, such as lagged test scores and parental education, does decrease the magnitude of the relationship, Table 6 shows that the significant positive relationship still exists. These results also provide evidence that the effort measure created in this paper is picking up an important input into test scores, and that this input is not captured by lagged test scores and other demographic variables.

School Accessibility and Enrollment

When deciding whether to enroll in school or not, students take into consideration both the availability of schools and their outside option of working. The farther away a middle school is from their primary school, the higher the cost of traveling there. Figure 7 shows that students who have no schools in their area are more likely to drop out than the students who have middle schools nearby.

Wages and Working

The imputed wages depend not only on student characteristics, but also based on the municipality in which the student resides. This provides geographic heterogeneity in the wages. However, this geographic variation is not useful if it is not correlated with the choices that students are making. Table 7 contains estimates from a linear probability regression, where the observations are at the municipality level. For each municipality, the average monthly wage is calculated, as well as the average lagged test scores, the average parental education level, and the mean proportion of students receiving the conditional cash transfer. Wages are normalized in this regression so that a standard deviation is equal to 1. The results show that increasing the mean wage by one standard deviation is correlated with an increase of almost 1 percentage point in the dropout rate of the municipality. This is very significant given that the national average dropout rate is less than 8%.

1.4. Model

The model captures the different choices that students make as they progress from primary school (Grade 6) to middle school (Grade 7). They choose what school, if any, they wish to attend. Based on the location of the primary school that student i attended (P_i), the student will have a choice set of available middle schools, $\mathbb{S}_{P_i}$. Middle schools are categorized into three types: General, Technical (vocational) and telesecundarias (distance education). Students also make a labor choice. If the student chooses not to enroll in school, it is assumed that they work full time. Students who choose to enroll in school may choose between working part time or focusing only on their studies. Students receive wage offers that depend on their age, gender, parental education, location and whether they are enrolled in school. Finally, students who enroll in school make an effort choice. Effort is costly, however it is an input into the achievement production function and students' utility depends on achievement.

Each student who finished Grade 6 enters the model with a set of initial conditions. These include their gender, their age, their lagged test scores and if they are a beneficiary of Prospera, the conditional cash transfer program. Also included are permanent family characteristics including the number of siblings, the parental education levels, the monthly family income, and some information about the household, such as if they own a computer. Finally, the geographic location of the primary school is included, which gives information on whether the neighbourhood is rural or urban, and also identifies the choice set of middle schools.

1.4.1. Student Utility

Students in the model are 12 years old on average, and therefore it is plausible that they are making their schooling choice along with their family.[9] Families care about student achievement, monetary compensation coming from Prospera or wages, the type of school the student attends, the cost of traveling to school, and the cost of effort. Effort may be more costly if the student has other demands on their time, such as a part time job, or if they have lower lagged test scores. The utility of student i attending school j with labor choice L is given by

$$
\begin{aligned}
U_{ijL}(e_{ijL}) = {} & \underbrace{CCT_i + \mathbb{1}\{L = PT\} w_i^{PT}}_{\text{Monetary Compensation}} + \underbrace{\alpha_1 d_{P_i j} + \alpha_2 d_{P_i j}^2}_{\text{Distance Traveled}} + \\
& \underbrace{(\alpha_3 + \alpha_4 PEduc_i + \alpha_5 \mathbb{1}\{CCT > 0\}) \left(\hat{A}_{ij}^{7,S}(e_{ijL}) + \hat{A}_{ij}^{7,M}(e_{ijL}) \right)}_{\text{Achievement}} + \\
& \underbrace{\alpha_6 + \alpha_7 PEduc_i + \sum_{k \in Type} \beta_k \mathbb{1}\{Type_j = k\}}_{\text{School Types}} + \underbrace{\alpha_8 \mathbb{1}\{L = PT\}}_{\text{Working}} + \\
& \underbrace{(\alpha_{i,9} + \mathbb{1}\{L = PT\}\alpha_{10})\, e_{ijL} + \alpha_{11} e_{ijL}^2}_{\text{Effort}} + \nu_{ijL}
\end{aligned}
$$

The monetary compensation includes the conditional cash transfer CCT_i, which student i receives if they are a Prospera beneficiary, as well as a part-time wage w_i^{PT}, which they receive if they choose to work part time. The coefficient on the monetary component is constrained to one, so that the units of the remaining utility coefficients are in terms of money (pesos). The distance between student i's primary school P_i, and middle school j is given by $d_{P_i j}$. Achievement in Spanish and math, $\hat{A}_{ij}^{7,S}(e_{ijL})$

[9] In the ideal scenario a family budget constraint would be included. Unfortunately, although the survey does contain a question on family income, the responses are in very coarse bins, and are not fine enough to include in a budget constraint. In the results section, I discuss how not including a budget constraint could bias the results.

and $\hat{A}^{7,M}_{ij}(e_{ijL})$, depend on student characteristics, middle-school characteristics, and students' effort choices e_{ijL}. Students may care differently about their scores depending on their parent's education, $PEduc_i$ and if they are a conditional cash transfer beneficiary. To capture parental education, $PEduc_i$ is equal to one if both parents have at least a middle-school education. Students receive a benefit from enrolling in school, which is captured by α_6, and this benefit may vary depending on parental education. $Type_j$ is school j's type, and can be one of telesecundaria, Technical or General. Students potential distaste for working while in school is captured by the coefficient α_8.

A linear and quadratic term for effort are included in the student utility. This allows for flexibility and also ensures a solution for the optimal effort for each student. The random coefficient $\alpha_{i,5}$ captures heterogeneity in the marginal cost of effort across students. The coefficient can be broken down into a component that is constant across students, a component that varies with student characteristics, and a random unobserved component,

$$\alpha_{i,9} = \alpha_9 + \lambda X_i + \eta_i$$

where $\eta_i \sim \mathcal{N}(0, \sigma^2_\eta)$. Student characteristics contained in X_i include the students' gender, their parental education, and their lagged test scores.

If students choose the outside option, they are choosing to drop out of school after 6th grade. It is assumed that they work full time, and receive a full time wage w_i^{FT}.

$$U_{i0} = w_i^{FT} + \nu_{i0}$$

The error terms are assumed to be iid type I extreme value, so the overall framework is a mixed logit model. The wages, w_i^{PT} and w_i^{FT} are estimated using Mexican Census data as described below.

Student i's choice set of middle schools, $\mathbb{S}_{P_i}$, is comprised of all middle schools within a certain distance of their primary school, P_i. This distance is computed by considering how far students have historically traveled from this primary school. Because of this, some choice sets cover smaller areas than others. Each school in the choice set is defined by the distance between it and student i's primary school, $d_{P_i j}$, and the type of school it is, $Type_j$. Other school-level variables from the principal survey that I include in the analysis relate to infrastructure and principal and teacher quality.

1.4.2. Wage Offers

Each student receives a full-time and a part-time wage offer. If they accept the full time wage, they are not able to enroll in school. They can also choose to not accept either offer and only enroll in school. Potential hourly wages for children are imputed using Census data (details are in Section 1.3.3). Wages are allowed to depend on age, gender, school attendance, parental education, and geographic location (either urban/rural and municipality).

1.4.3. Expected Test Scores

For students who choose to enroll in school, their test scores are generated by a value-added production function. The student inputs to the production function include lagged test scores, student characteristics (including age, gender, and family characteristics) and their effort choices. School inputs, Z_j, include the type of school, principal education and experience, if the school has internet, if the school teaching materials are sufficient, and how the principal rates the teachers.

$$\hat{A}_{ij}^{7,T}(e_{ijL}) = \delta_0^T + \underbrace{\delta_1^T A_i^{6,M} + \delta_2^T A_i^{6,S}}_{\text{Lagged Scores}} + \underbrace{\delta_3^T e_{ijL}}_{\text{Effort}} + \underbrace{\delta_4^T X_i}_{\text{Student chara.}} + \underbrace{\delta_5^T Z_j}_{\text{School chara.}} + \delta_6^T e_{ijL} Z_j + \xi_{ije}^T \text{ for } T \in \{S, M\} \quad (1.1)$$

The last term, $e_{ijL}Z_j$, is an interaction between the student's effort level and the school type, allowing for effort to be more or less productive depending on the type of school attended. The value-added equation is estimated separately for math and Spanish test scores. For each student, the math and Spanish residuals are allowed to be correlated. Students are assumed to not know the error terms when making their school choices. Working does not directly affect achievement, however, working makes study effort more costly. The benefits of effort may vary by school type.

1.4.4. Maximization Problem

Student i solves the following maximization problem for their optimal level of effort e_{ijL}^* for each possible school j and labor option L in their choice set:

$$\begin{aligned} e_{ijL}^* = \underset{e_{ijL}}{\operatorname{argmax}} \quad & U_{ijL}(e_{ijL}, \hat{A}_{ij}^{7,S}(e_{ijL}), \hat{A}_{ij}^{7,M}(e_{ijL}); X_i, Z_j, w_i^{PT}, w_i^{FT}) \\ \text{s.t.} \quad & \hat{A}_{ije}^{7,S} = f_S(A_i^{6,M}, A_i^{6,S}, e_{ijL}; X_i, Z_j) \\ & \hat{A}_{ije}^{7,M} = f_M(A_i^{6,M}, A_i^{6,S}, e_{ijL}; X_i, Z_j) \end{aligned}$$

The first-order equation of the above maximization problem yields the following expression for optimal effort:

$$e^*_{ijL} = \frac{-\left(\begin{array}{r}(\alpha_3 + \alpha_4 PEduc_i + \alpha_5 \mathbb{1}\{CCT > 0\})(\delta_3^S + \delta_6^S Z_j + \delta_3^M + \delta_6^M Z_j) \\ + \alpha_{i,9} + \mathbb{1}\{L = PT\}\alpha_{10}\end{array}\right)}{2\alpha_{11}} \quad (1.2)$$

The parameter $\alpha_{i,9}$ is a function of the student characteristics, X_i, and the random shock, η_i. The optimal effort therefore depends on student characteristics, school characteristics, labor status, and an idiosyncratic preference shock.

Define the dummy variable $D_{ijL} = 1$ if student i chooses school j and labor option L. Student i then solves the following maximization problem, given their solutions for optimal effort e^*_{ijL} and the expected achievement that the optimal effort implies ($\hat{A}^{7,S}_{ije^*}$ and $\hat{A}^{7,M}_{ije^*}$).

$$\max_{j,L} \sum_{j=1}^{J_i} \sum_{L \in \{0,PT,FT\}} D_{i,j,L} \times U_{ijL}(e^*_{ijL}, \hat{A}^{7,T}_{ij}(e^*_{ijL}), \hat{A}^{7,T}_{ij}(e^*_{ijL}); X_i, Z_j, w_i^{PT}, w_i^{FT})$$

The final result is that each student has an optimal school j and labor option L, and an optimal effort given these choices, e^*.

1.5. Estimation

Model parameters are estimated using Maximum Likelihood. Define

$$P(j, L, A^S_{ij}, A^M_{ij}, \tilde{e}^M_{ijL} | X_i, Z_j, w_{ij}, \eta_i)$$

as the joint probability of choosing school j, labor option L, having Grade 7 test scores A^S_{ij} and A^M_{ij}, and choosing effort measures $\tilde{e}^M_{ijL}$. The probability depends on student characteristics X_i, school characteristics Z_j, imputed wages w_{ij}, and the random coefficient shock η_i. Although they are not written explicitly in the above probability, there are several other shocks in the model with defined distributions: ν_{ijL} are type I extreme value and ξ_{ij}^M and ξ_{ij}^S are jointly normal.

Define $D_{ijL} = 1$ if student i chose school j and labor option L. The likelihood is then,

$$L = \prod_{i=1}^{N} \int \prod_{j=1}^{J_i} \prod_{L \in \{0,PT,FT\}} \left[P(j, L, A_{ij}^S, A_{ij}^M, \tilde{e}_{ijL}^M | X_i, Z_j, w_{ij}, \eta_i) \right]^{D_{ijL}} f_\eta(\eta_i) d\eta_i$$

The joint probability can be decomposed into the product of conditional probabilities. The variable $\tilde{e}_{ijL}^M$ is the effort variable in the data. Two of the conditional probabilities depend on e_{ijL}^* and using Equation 1.2, e_{ijL}^* can be calculated given the choice of j and L, along with the data (X_i, Z_j), the random coefficient shock (η_i) and model parameters. Conditioning variables in probabilities are dropped in the probability expressions if the probability does not depend on them.

$$\begin{aligned}
L &= \prod_{i=1}^{N} \int \prod_{j=1}^{J_i} \prod_{L \in \{0,PT,FT\}} \left[P(j, L, A_{ij}^S, A_{ij}^M, \tilde{e}_{ijL}^M | X_i, Z_j, w_{ij}, \eta_i) \right]^{D_{ijL}} f_\eta(\eta_i) d\eta_i \\
&= \prod_{i=1}^{N} \int \prod_{j=1}^{J_i} \prod_{L \in \{0,PT,FT\}} \Bigg[P(A_{ij}^S, A_{ij}^M | j, L, \tilde{e}_{ijL}^M; X_i, Z_j, w_{ij}, \eta_i) \\
&\qquad\qquad\qquad\qquad \times P(j, L, \tilde{e}_{ijL}^M | X_i, Z_j, w_{ij}, \eta_i) \Bigg]^{D_{ijL}} f_\eta(\eta_i) d\eta_i \\
&= \prod_{i=1}^{N} \int \prod_{j=1}^{J_i} \prod_{L \in \{0,PT,FT\}} \Bigg[P(A_{ij}^S, A_{ij}^M | j, L, ; X_i, Z_j, \eta_i) P(\tilde{e}_{ijL}^M | j, L, X_i, Z_j, \eta_i) \\
&\qquad\qquad\qquad\qquad \times P(j, L | X_i, Z_j, w_{ij}, \eta_i) \Bigg]^{D_{ijL}} f_\eta(\eta_i) d\eta_i
\end{aligned} \tag{1.3}$$

Consider each of the three probabilities in the likelihood. The first is the probability of observing the Grade 7 test scores in Spanish and math:

$$P(A_{ij}^{S}, A_{ij}^{M} | j, L, \tilde{e}_{ijL}^{M}; X_i, Z_j, \eta_i)$$

The errors for the two achievement production functions are distributed iid jointly normal. Given the choice of school and labor, the data and the model parameters, the measure of effort from the model e_{ijL}^{*} can be computed. Using all of these inputs, the expected test scores can be computed using Equation 1.1. Given the normality assumption, and the expected test scores computed from the model, the probability of observing the test scores from the data can be computed.

The second probability is the probability of observing the effort measure in the data, conditional on the optimal effort predicted from the model.

$$P(\tilde{e}_{ijL}^{M} | j, L, X_i, Z_j, \eta_i) = P(\tilde{e}_{ijL}^{M} | e_{ijL}^{*})$$

Equation 1.2 defines optimal effort in the model. The coefficient $\alpha_{i,9}$ in the numerator is a random coefficient with associated shock $\eta_i \sim \mathcal{N}(0, \sigma_{\eta}^2)$. Therefore effort draws can be thought of as coming from the distribution of the true underlying value of effort, $\mathcal{N}(e_{ijL}^{*}, \sigma_{e^*}^2)$. This distribution is used to estimate the probability of observing the effort value obtained from factor analysis. Because of this, I do not need to simulate in order to calculate the integral defined in the likelihood.

The third and final probability is the probability of choosing school j and labor option L.

$$P(j, L | X_i, Z_j, w_{ij}, \eta_i)$$

The errors for the utility function are distributed iid type I extreme value. The probability of a school and work combination can be written as:

$$P(j, L|X_i, Z_j, w_{ij}, \eta_i) = \frac{\exp^{Utility_{ijL}}}{\sum_{k=1}^{J_i} \sum_{h \in \{0,PT,FT\}} \exp^{Utility_{ikh}}} \tag{1.4}$$

$Utility_{ijL}$ is a function of e^*_{ijL}, the model parameters, and the data. A scale parameter is also included in the above probability. The outside option has been normalized to the value of a wage instead of zero, and the coefficient on the monetary component is set to 1. Because of this, the scale of the distribution can be estimated.

Given a set of parameter values and the data, all three of these probabilities can be calculated for each student, and the product of them is defined as the individual likelihood. The likelihood defined in Equation 1.3 can then be calculated, and maximized to find the estimated parameters.

To calculate the standard errors, I estimate a sandwich-type covariance matrix. Details are in Appendix A.5.

1.5.1. Identification

There are 51 parameters to estimate in the model in total. The list of parameters is given by

- Utility function: $\{\alpha_k\}_{k=1}^{11}$, $\{\beta_k\}_{k=1}^{2}$, $\{\lambda_k\}_{k=1}^{3}$, σ_U
- Achievement production functions: $\{\delta_k^M\}_{k=1}^{15}$, $\{\delta_k^S\}_{k=1}^{15}$, σ_M, σ_S, σ_{MS}
- Effort: σ_E

There are 33 parameters associated with achievement. They are estimated with two value added equations. Each student who attended Grade 7 has a test score in both Math and Spanish. Each student also has lagged test scores in both subjects, as well as data on the 12 other covariates. There is variation in covariates across schools, and across students within a school.

There are 16 parameters in the utility function. Two of the parameters are associated with distance. They are identified by geographic variation in distances in different children's choice sets. Each primary school has different schools in its choice set, and every option is associated with a distance (among other characteristics). School types that are far away from a specific primary school may be of a good quality, but are chosen by a small fraction of the students (or not at all), which identifies how costly students find traveling to school.

Three parameters in the utility function represent school type (General, Technical, telesecundaria). There are too many schools in the data to have intercepts for each of them. Instead of having a common intercept in the utility for attending each school, I assume that the intercept varies by school type. These coefficients are identified by variation within choice sets as well. Students may chose a certain type of school over another even though it is farther away or offers a worse expected test score, showing a preference for this type of school over the other.

Two parameters in the utility function capture how much students value expected test scores. Two factors come into play here. The first is that students with higher test scores may get more utility from going to school compared to dropping out. The second is that achievement is affected by school inputs, so some schools in the choice set may have higher expected test scores which could make students more likely to attend. Either of these things being present in the data would identify the coefficients on test scores.

There are six parameters associated with the marginal cost of effort in the utility function. The parameters involved in demographics (parental education, female, lagged test scores) are identified by the difference in mean effort choices from students with these different demographics.

1.6. Results

Estimates for the utility parameters are shown in Table 8 and for the test score production functions are shown in Table 9. All parameters in the utility function have the units of 100s of pesos per month. The key parameter estimates and patterns are discussed below. Traveling distance to a middle school is estimated to be costly. The coefficient on distance squared is positive, showing that as the school gets farther away, the marginal cost of another kilometer starts to decrease. Both estimates are significant, even with a small estimation sample compared to the full data sample.

The estimate for the benefit of attending school in the utility is large and positive, and does not seem to depend on if parents have middle-school education. Technical schools are estimated to be slightly more valuable than general schools, but the difference is not significant. Telesecundarias are estimated to be perceived significantly worse than the other two school types.

The average expected test score has a positive coefficient in the utility function, with little change depending on parental education and conditional cash transfer status. The standard deviation of the test scores is approximately 1, meaning that students and their families place approximately the same value on a school being a kilometer closer as the school improving math test scores by one and a half standard deviations.

There is a large distaste for working part time. Further, working part time is estimated to make the marginal cost of effort more negative, so more costly. The coefficient on effort squared must be negative to guarantee a solution to the optimal effort problem in the model, and it is in fact a large negative number. The marginal cost of effort is estimated to decrease, so effort is less costly, for female students and students with higher lagged test scores.

The coefficient estimates in the achievement production function, shown in Table 9, are fairly intuitive. Lagged test scores are significant, with lagged math scores contributing to math predictions, and lagged Spanish scores contributing to Spanish predictions. Females have negative coefficients in math. Students with higher ages are estimated to do worse in both math and Spanish. The value added of a Technical schools is estimated to be greater than a General school in both math and Spanish whereas telesecundarias are estimated to be worse. The school characteristics coefficients are mainly small in magnitude and insignificant. The one exception is if the school has internet, which has a positive and significant coefficient. Finally, effort has a large positive coefficient for both math and Spanish. Effort is estimated to be more productive in telesecundaria schools compared to General schools, and less productive in Technical schools, however the change in productivity is small in magnitude.

1.6.1. Model Fit

The following figures show the fit of the model with respect to the true data. There are three main outcomes to fit: school choice, achievement, and effort. Table 10 shows the model fit for the means and standard deviations of these outcomes. The simulation means are overall quite close to the means in the data.

Figure 8 shows the fraction of students that choose each of the three school type options or to drop out in both the simulation and the data. The pattern in the data is represented in the simulated data, in that General schools are most popular, followed by Technical, telesecundarias and then Dropping out. However the values are slightly off, with somewhat more students choosing to drop out in the model than in the data.

Considering only the students who choose to drop out, Figure 9 investigates the relationship between dropping out and distance to the nearest school. Students are divided into quintiles by the distance to their nearest middle school. The mean dropout rate for each quintile is then calculated in the data and the model. The overall pattern matches, but it is apparent that the model is overestimating dropout rates for students who have a middle school in the same location as their primary school, or who have a middle school very far away. The students in between are matched very closely.

Finally, Figure 10 investigates the relationship between students' effort values and their lagged test scores. In both the data and the model, students with lower lagged test scores exert less study efforts than students with higher lagged test scores. The model captures the relationship in the data very well.

1.7. Evaluation of Child Labor Policies

With my estimated model, I am able to evaluate many relevant policies involving child labor laws, conditional cash transfers and school availability. The focus for this paper is to consider the impact of enforcing child labor laws on both dropout and achievement. Working while enrolled in school is detrimental to achievement, however I find that for many students they require the income to stay in school, and if they are not allowed to work while in school they prefer to drop out and work full time. There are two ways to counter this problem. The first is to fully prohibit child labor, both while enrolled in school and if the child has dropped out. This makes the outside option less appealing and more children will stay in school. The second is to offer conditional cash transfers as an incentive for students to enroll. The cash transfers may be the more feasible policy, however program targeting can still pose a challenge and affect the results, as does the benefit amount offered.

Using the parameter estimates, I draw shocks and simulate choices under the baseline model. Then, to do the counterfactual exercises, I change either some parameters or the choice sets that the students face, and simulate again under the modified environment using the same shocks. The results from the two simulations are compared to evaluate the policy. Of interest are the change in enrollment rates, the change in achievement, which types of schools have the largest change in enrollment, the amount of money gained/lost by families, among other outcomes.

The first counterfactual involves removing the part-time labor option. The students who chose not to work originally are not affected by this policy, and neither are students who chose to drop out. However, the children who were working while enrolled in school must decide if they wish to continue studying without the income they received, or drop out and work full time. This counterfactual could represent a policy such as teachers being able to better monitor their students, or if there was an after-school program implemented so that children studied or played sports at school during hours when they may normally work. The results of this policy are shown in Table 11. In the estimated model, 22% of students enrolled in school choose to work at the same time. When working part time is not an option 7% of these students decide to dropout, increasing the over all dropout rate by 20%, from 8% to 10%. This is a drastic increase, and represents a large number of students when considering the entire student population of Mexico. There does not appear to be much change in effort, Spanish and Math scores when considering all students, so I investigate these further.

Table 12 computes statistics for the group of students who would like to work, but when they are prohibited from working, stay enrolled in school. From this table, it is clear that they are increasing their efforts, which in turn increases their math and Spanish scores. Effort increases by an average of 3.3 percent of a standard deviation, which results in a 2.9 percent of a standard deviation increase in math scores and a 3 percent of a standard deviation increase in Spanish scores.

The final analysis analyzes the characteristics of the students who are most likely to drop out because of this policy. Table 13 shows the mean value of background characteristics for the students who would like to work while in school, separated by if they stay in school or not after the policy. The students who drop out have lower lagged test scores and have a much higher rate of being a conditional cash transfer beneficiary. The gender breakdown is very similar and the students that dropout are only slightly older than those that stay enrolled. The last three rows show that the students who stay enrolled are much more likely to have one of their parents have at least a middle school education, and to be in the top half of the income distribution. Overall, this table shows that the students who are dropping out are the students who are struggling academically and come from disadvantaged backgrounds.

Increasing the national dropout rate by such a large amount is not an ideal result of a policy that prohibits working while in school. A possible way to prevent this, would be to consider prohibiting all labor. This would reduce the value of dropping out, as the students who dropped out would not receive wages. Results from this counterfactual, along with the results from the estimated model and the first counterfactual are shown in Table 14. The numbers show that prohibiting all child labor would have better impacts than only prohibiting labor while in school, in that the dropout rate is similar to the original estimation. Effort and test scores are also slightly higher than in the baseline model, since students are dedicating all of their time to their studies.

Although prohibiting all child labor may have positive educational outcomes, as a policy it would be difficult to enforce. Therefore I look to an alternative policy to encourage enrollment if working part time is prohibited. Luckily in Mexico there is a well established policy, the conditional cash transfer, that could be modified. In my third counterfactual, I consider changing the values of the conditional cash transfer and expanding eligibility for the program.

Figure 11 shows the reduction in dropout rates for three different conditional cash transfer policies. The x-axis shows an increase in the amount of the transfer, ranging from the current Prospera transfer amount, to 9 times the current amount. The three policies change who is offered the conditional cash transfer. Policy 1 is a hypothetical policy that is not operational, but shows the best that could be achieved with a cash transfer of the given magnitude. In this policy, any student who would drop out in counterfactual 1 is offered the transfer. In reality, it would be impossible to target the policy this way. Policy 2 considers increasing the transfers to the current beneficiaries, which would be very simple to implement. Policy 3 extends the transfer beneficiaries to those who currently received Prospera, and those who have an income below the median.

The results show that increasing the conditional cash transfer payment is a very effective way of decreasing the dropout rate. For payment amount similar to the current value, expanding the conditional cash transfer to other low-income families does not have a significant effect. However, if the cash transfer is increased, then extending the transfer to these families does drastically help reduce the dropout rate.

To summarize the results, I find that prohibiting students from working while in school increases achievement by approximately 3% of a standard deviation, however it also causes a substantial increase in the dropout rate. If it were possible to ban all child labor, the dropout rate would remain close to the baseline and achievement would increase. However, this would be difficult to enforce, and I find that similar dropout rates can be achieved when working part time is banned and the cash transfer is either increased, or the beneficiaries expanded.

1.8. Conclusion

Increasing human capital is thought to be one of the best ways for developing countries to achieve growth and to increase equity. Ensuring that all children attend school to a certain age and receive a high quality education is a priority. Unfortunately, in many developing countries, child labor is prevalent and it makes providing an education to all students more challenging. Although there is an extensive literature on school choice, it is necessary to extend the currently available frameworks to consider the problem of child labor and how it interacts with school choices. In my model, I include both schooling and labor choices and I provide a mechanism through which labor affects educational achievement, which is the study effort that children dedicate to their education.

Specifically, I develop and estimate a random utility model over discrete school-work alternatives, where study effort is determined as the outcome of an optimization problem under each of these alternatives. Students who do not enroll in school are assumed to work full-time, and receive the associated wage. Students who enroll in school may choose to work part-time, for which they receive the benefit of a part-time wage, but incur the cost of increased marginal cost of effort. The results show that effort is an important input to achievement, which is estimated with a value added equation. Students who work, and as a result choose to put in less, end up with lower achievement than they would if they had not chosen to work.

To estimate my model, I combine several data sources: administrative data on nationwide standardized tests in math and Spanish, survey data from students, parents and principals, geocode data on school locations, and Mexican census data on local labor market wages and hours worked. The majority of the model parameters are precisely estimated.

By removing the part-time labor option from student's choice sets, I evaluate the impact of working while in school. I find that for the majority of students, not being able to work improves their test scores. However, almost 10% of students who would prefer to work drop out of school to work full time when the part time option is no longer available. This increases the dropout rate by approximately 20%. I analyze two policies that could be used in conjunction with prohibiting labor for enrolled students. The first is to ban all child labor. This removes the incentive to drop out of school and work full time and reduces the dropout rate, however it would be a challenging policy to implement. The second policy is to increase the conditional cash transfer, in both the payment amount and the pool of beneficiaries. I find that depending on the transfer amount, these policies show considerable potential.

With the model that I have developed and estimated, it is possible to analyze many other educational policies. I incorporate school choice and locally available schools, so one possible direction is to consider questions of school access and quality. Especially in rural areas, it is of interest to understand how the conditional cash transfer interacts with another important education policy in Mexico, the distance education schools (telesecundarias). In ongoing work, I am considering a range of such policies.

1.9. Tables

Table 1: Summary Statistics for Test Scores

	Mean	Standard Deviation
Grade 6 Math	531.2	122.5
Grade 6 Spanish	524.8	108.5
Grade 7 Math	501.0	101.5
Grade 7 Spanish	499.2	101.3

Scores are for the cohort of students in Grade 7 in 2009. The distributions of the test scores are approximately normal, as shown in Appendix A.1.

Table 2: Summary Statistics for the Wage Imputations

		Female	Male
Work and School	Hourly Wage	13.6 (2.9)	13.4 (2.8)
	Weekly Hours	19.0 (4.1)	19.1 (3.6)
	Monthly Wage	1030 (294)	1029 (305)
Only Work	Hourly Wage	14.2 (2.8)	14.1 (2.3)
	Weekly Hours	39.4 (3.6)	39.7 (3.1)
	Monthly Wage	2236 (459)	2242 (431)

Summary statistics from the two wage regressions: hourly wages and hours worked per week. Monthly wage is calculated as the product of the hourly wage and the hours worked per week times four (weeks in a month). Monetary values are in 2010 pesos. The sample of children used for the imputation is the estimation sample used throughout the paper.

Table 3: Summary Statistics for School Types

	General	Technical	Telesecundaria	Private
Number of Schools	5,820	2,857	15,974	3,866
Proportion of Cohort	0.45	0.27	0.20	0.08
Proportion Female	0.50	0.50	0.49	0.50
Proportion CCT	0.17	0.20	0.64	0.01
Proportion Rural	0.15	0.23	0.87	0.02
Mean Class Size	32.3	33.9	16.6	23.6
Mean School Cohort Size	137	170	22	39

Summary statistics for the four types of middle schools. All data on Grade 7 students in 2009 is used to create this table.

Table 4: Days Worked per Week and Test Scores

	Dependent variable:			
	Grade 7 Test Score			
	(> 0 Day)	(> 1 Days)	(> 2 Days)	(> 3 Days)
	(1)	(2)	(3)	(4)
Working	−0.047***	−0.070***	−0.106***	−0.150***
	(0.012)	(0.013)	(0.015)	(0.017)
Lagged Tests Scores	0.675***	0.675***	0.674***	0.674***
	(0.003)	(0.003)	(0.003)	(0.003)
Female	0.183***	0.182***	0.183***	0.183***
	(0.010)	(0.010)	(0.010)	(0.010)
Age	−0.078***	−0.078***	−0.077***	−0.077***
	(0.008)	(0.008)	(0.008)	(0.008)
Mom Middle School	0.041***	0.041***	0.041***	0.041***
	(0.011)	(0.011)	(0.011)	(0.011)
Dad Middle School	0.072***	0.072***	0.072***	0.072***
	(0.012)	(0.012)	(0.012)	(0.012)
Prospera	0.129***	0.130***	0.130***	0.131***
	(0.012)	(0.012)	(0.012)	(0.012)
Constant	3.890***	3.890***	3.889***	3.886***
	(0.103)	(0.103)	(0.103)	(0.103)
Observations	64,356	64,356	64,356	64,356
R^2	0.505	0.505	0.505	0.506
Note:				*p<0.1; **p<0.05; ***p<0.01

Correlation between days worked per week and test scores in the data. The covariate "Working" is a dummy variable, and its definition changes depending on the column. In column (1), "Working" is equal to 1 if students work at least 1 day a week. In column (2) "Working" is equal to 1 if students work at least 2 days a week, and so on. The dependent variable is the sum of each student's Grade 7 math and Spanish test scores.

Table 5: Working and Study Effort

	Dependent variable:		
	Effort		
	(1)	(2)	(3)
Working	-0.204^{***}	-0.086^{***}	-0.070^{***}
	(0.012)	(0.011)	(0.012)
Lagged Tests Scores		0.172^{***}	0.171^{***}
		(0.003)	(0.003)
Female			0.128^{***}
			(0.010)
Age			-0.121^{***}
			(0.008)
Mom Middle School			-0.015
			(0.011)
Dad Middle School			0.062^{***}
			(0.011)
Prospera			0.234^{***}
			(0.012)
Constant	4.698^{***}	2.911^{***}	4.215^{***}
	(0.006)	(0.028)	(0.100)
Observations	64,356	64,356	64,356
R^2	0.005	0.065	0.077
Note:			*p<0.1; **p<0.05; ***p<0.01

Correlation between days worked per week and study effort in the data. The variable "Working" is a dummy variable equal to 1 if a child reports working at least 1 day a week. The effort variable is the continuous variable created with factor analysis.

Table 6: Study Effort and Achievement

	Dependent variable:		
	Grade 7 Test Score		
	(1)	(2)	(3)
Study Effort	0.430***	0.190***	0.182***
	(0.005)	(0.004)	(0.004)
Lagged Tests Scores		0.655***	0.645***
		(0.003)	(0.003)
Female			0.166***
			(0.010)
Age			−0.057***
			(0.008)
Mom Middle School			0.045***
			(0.011)
Dad Middle School			0.062***
			(0.011)
Prospera			0.082***
			(0.012)
Constant	8.026***	2.460***	3.116***
	(0.025)	(0.030)	(0.102)
Observations	64,356	64,356	64,356
R^2	0.094	0.517	0.521
Note:			*p<0.1; **p<0.05; ***p<0.01

Correlation between study effort and test scores in the data. Test scores are the sum of Grade 7 math and Spanish scores. The effort variable is the continuous variable created by factor analysis.

Table 7: Dropout Rates and Local Wages

	Dependent variable:
	DropRate
Full Time Wage	0.008***
	(0.002)
Lagged Tests Scores	−0.023***
	(0.004)
Mom Middle School	0.051***
	(0.011)
Dad Middle School	−0.024***
	(0.008)
Prospera	−0.030***
	(0.008)
Constant	0.306***
	(0.045)
Observations	936
R^2	0.222
Note:	*p<0.1; **p<0.05; ***p<0.01

Correlation between the dropout rate and the mean wage in a municipality. The observations are at the municipality level. For each municipality, the average wage is calculated, as well as the average lagged test scores, the average parental education level, and the mean proportion of students receiving the conditional cash transfer. Wages are normalized in this regression so that a standard deviation is equal to 1.

Table 8: Utility Function Parameter Estimates

Coefficients	Estimates	Std.Error
Distance	-46.49	3.22
Distance squared	1.90	0.13
School	36.97	17.98
School x Parent Educ	6.42	3.90
Technical School	11.26	2.43
Telesecundaria	-37.24	3.99
Expected Score	28.04	3.09
Expected x Parent Educ	0.39	2.77
Expected x CCT	0.42	0.13
Working Part Time	-56.62	3.71
Linear Effort	-33.81	9.62
Linear Effort - Lagged Score	1.86	0.29
Linear Effort - Female	0.94	0.30
Linear Effort - Parent Educ	-0.53	6.15
Linear Effort - Work	-0.43	0.19
Quadratic Effort	-5.23	0.86

Coefficient estimates for parameters in the utility function. The coefficient units are 100s of pesos in 2010. Estimates come from a random sample of 10,000 students.

Table 9: Achievement Functions Parameter Estimates

Coefficients	Math Estimates	Math Std.Error	Spanish Estimates	Spanish Std.Error
Intercept	-0.712	0.455	-0.511	0.493
Lagged Math	0.262	0.032	-0.06	0.03
Lagged Spanish	0.044	0.031	0.305	0.032
Female	-0.173	0.034	0.133	0.034
Age	-0.084	0.014	-0.072	0.014
Technical School	0.069	0.122	0.124	0.13
Telesecondary School	0.363	0.192	-0.815	0.21
Principal Education	-0.003	0.007	-0.026	0.007
Principal Experience	0.001	0.002	0.002	0.002
School has Internet	0.175	0.019	0.151	0.019
School has Materials	-0.034	0.015	-0.005	0.015
Teachers are Bad	-0.037	0.009	-0.037	0.009
Effort	1.12	0.156	1.103	0.156
Effort X Technical	-0.003	0.024	-0.019	0.025
Effort X Tele	-0.065	0.035	0.151	0.037
Residual Std Error	0.537		0.545	
Residual Covariance	0.253			

Coefficient estimates for parameters in the achievement functions. Estimates come from a random sample of 10,000 students.

Table 10: Goodness of Fit: Means and Standard Deviations

Outcome Variable	True Mean	Simulated Mean	True St.Dev.	Simulated St.Dev.
Math	5.03	5.28	0.98	1.59
Spanish	5.00	5.26	0.98	1.61
Effort	4.66	4.67	1.28	1.26
Fraction Drop	0.08	0.08		
Fraction General	0.44	0.44		
Fraction Technical	0.29	0.28		
Fraction Telesecundaria	0.19	0.19		
Fraction Work PT	0.25	0.22		

Model fit for relevant means and standard deviations. The values in the "True Mean" and "True St.Dev." columns come directly from the data. The values in the "Simulated Mean" and "Simulated St.Dev" come from simulations using the parameter estimates.

Table 11: Changes in Main Outcomes from Counterfactual 1

Outcome Variable	Estimated Model	Counterfactual	Percent Change
Fraction Work PT	0.22	0.00	-100.00
Fraction Drop	0.08	0.10	19.28
Mean Effort	4.67	4.68	0.26
Mean Spanish	5.26	5.28	0.40
Mean Math	5.28	5.30	0.38

Changes in outcomes from Counterfactual 1. This counterfactual partially enforces the child labor laws, so that children who are enrolled in school are not able to work at the same time (part time work decreases to 0).

Table 12: Effects of Counterfactual 1 on Part Time Workers

Outcome Variable	Estimated Model	Counterfactual	Change in SD (%)
Effort	4.65	4.69	3.27
Math	5.29	5.33	2.92
Spanish	5.27	5.32	2.96

Changes in outcomes from Counterfactual 1 for students who would like to work part time, and when they are enable to do so stayed enrolled in school.

Table 13: Demographics of Students Who Dropout After Counterfactual 1

Background Variables	Stayed In School	Dropped Out
Lagged Math	5.17	4.55
Lagged Spanish	5.16	4.51
Female	0.51	0.46
CCT	0.22	0.47
Age	12.00	12.26
Mom has Middle School	0.56	0.32
Dad has Middle School	0.58	0.33
Family Income above Mean	0.43	0.25

Difference in background variables between the students who dropped out when prohibited from working part time, and the students who stayed enrolled.

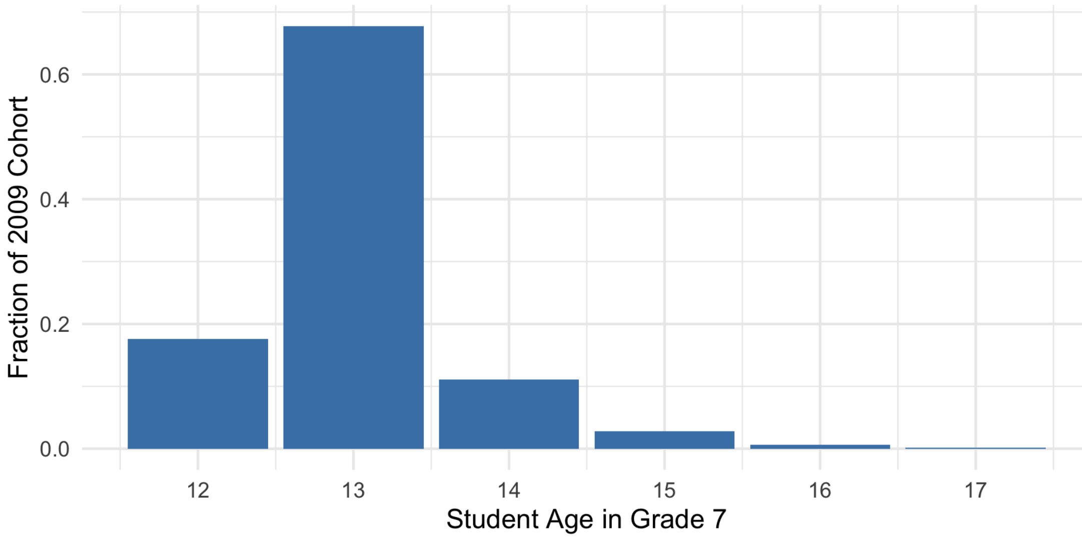
Table 14: Effects of Counterfactual 1 and Counterfactual 2

Outcome Variable	Estimated Model	Counterfactual 1	Counterfactual 2
Fraction Work PT	0.22	0.00	0.00
Fraction Drop	0.08	0.10	0.08
Mean Effort	4.67	4.68	4.68
Mean Spanish	5.26	5.28	5.26
Mean Math	5.28	5.30	5.29

Values for key variables from the estimated model, Counterfactual 1 and Counterfactual 2. All results are coming from simulations of the estimated model using the same draw of errors.

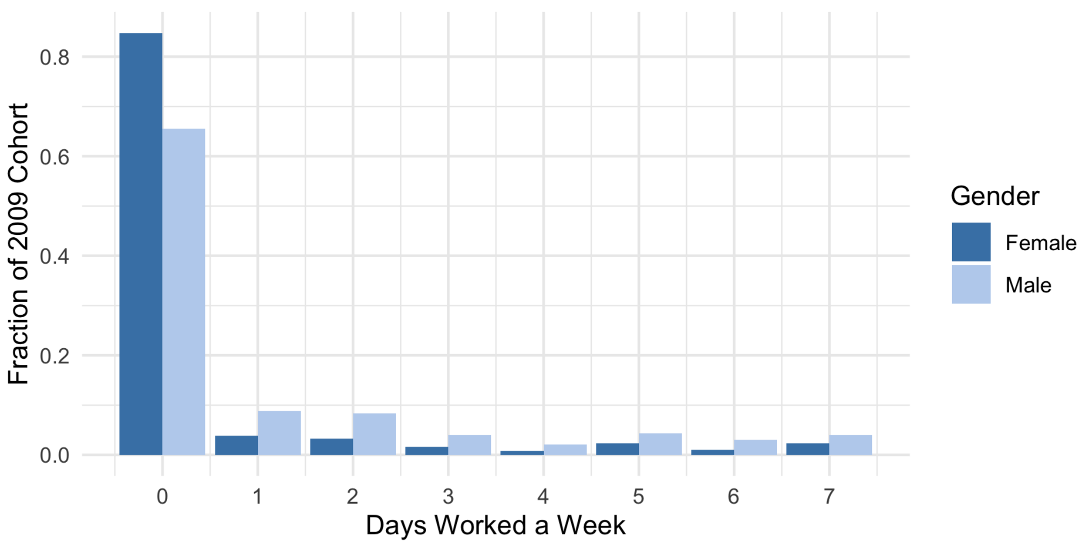
1.10. Figures

Figure 1: Age Distribution



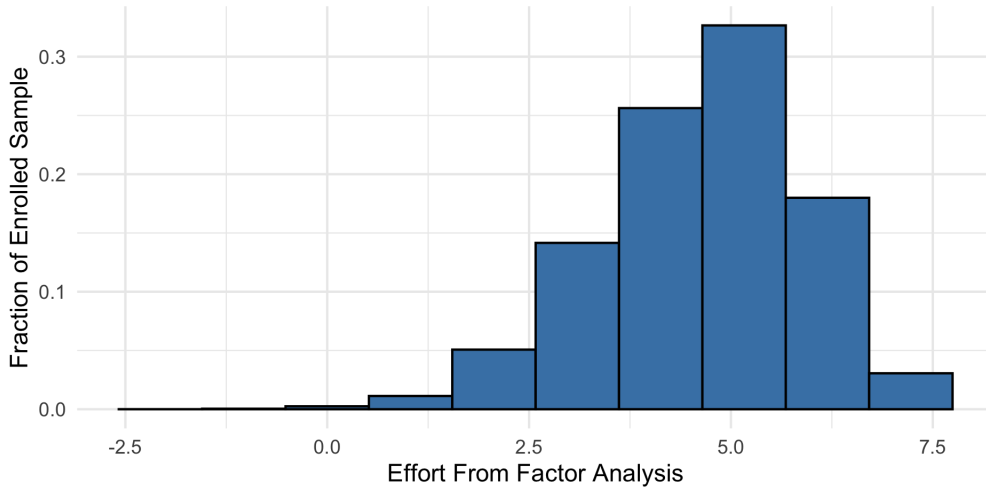
The distribution of ages of students in Grade 7 in 2009 in the estimation sample. The survey was completed in the Spring of their Grade 7 year.

Figure 2: Days Worked per Week



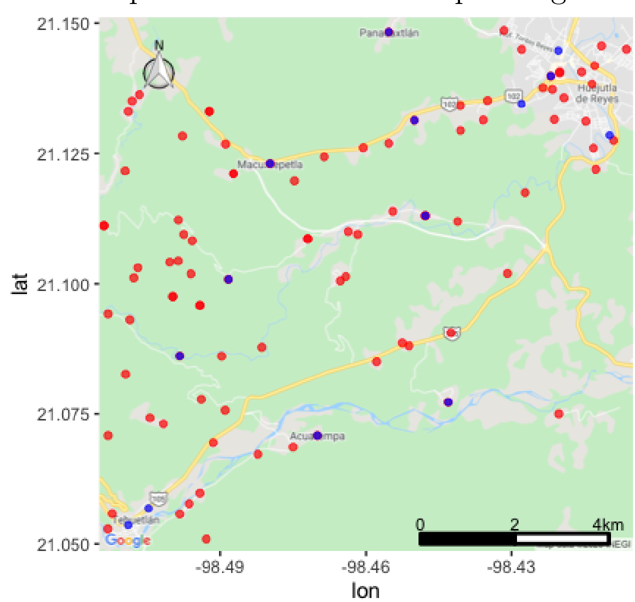
The distribution of the number of days worked per week, divided by gender, for students in Grade 7 in 2009 in the estimation sample.

Figure 3: Latent Effort Variable Distribution



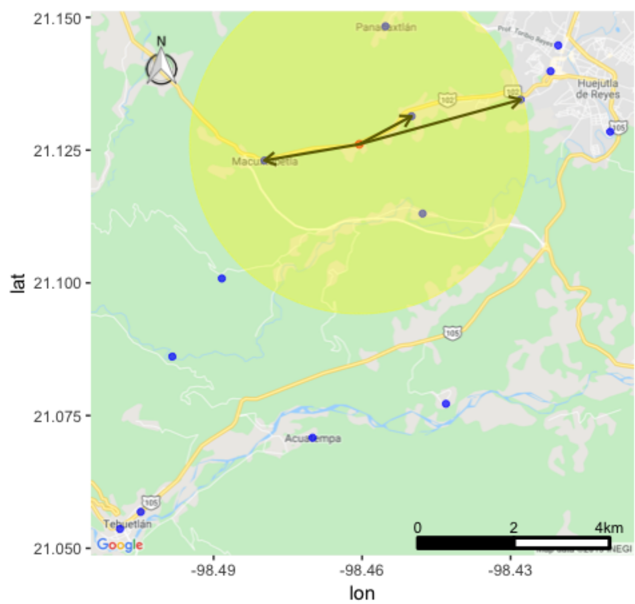
The distribution of latent effort values in the estimation sample. This variable is combines information from five questions related to study effort using factor analysis.

Figure 4: Map with Schools in Example Neighborhood



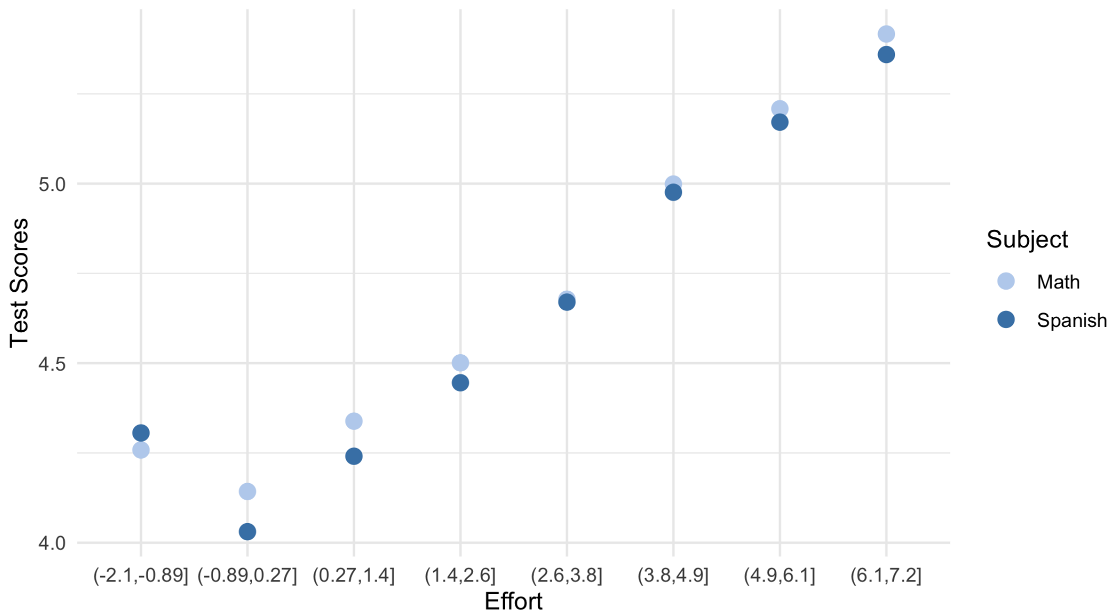
Map of all primary schools (red) and middle schools (blue) in a rural region of Mexico. The upper right corner contains a city while the remaining region is considered to be rural. There are many more primary schools than there are middle schools available.

Figure 5: Map Showing Choice Set in Example Neighborhood



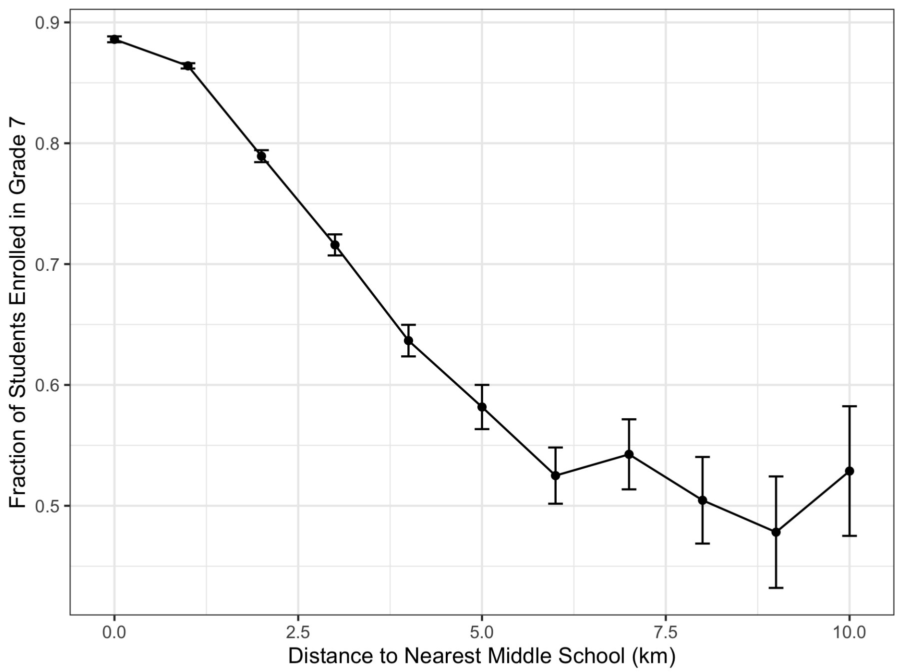
Map of a primary school (red) with the middle schools (blue) included in its choice set. The choice set is comprised of all schools included in the yellow circle. The arrows represent the actual choices of students from the primary school. There are two schools that were not chosen by students in the primary school, but are included in the choice set given their geographic proximity.

Figure 6: Study Effort and Test Scores



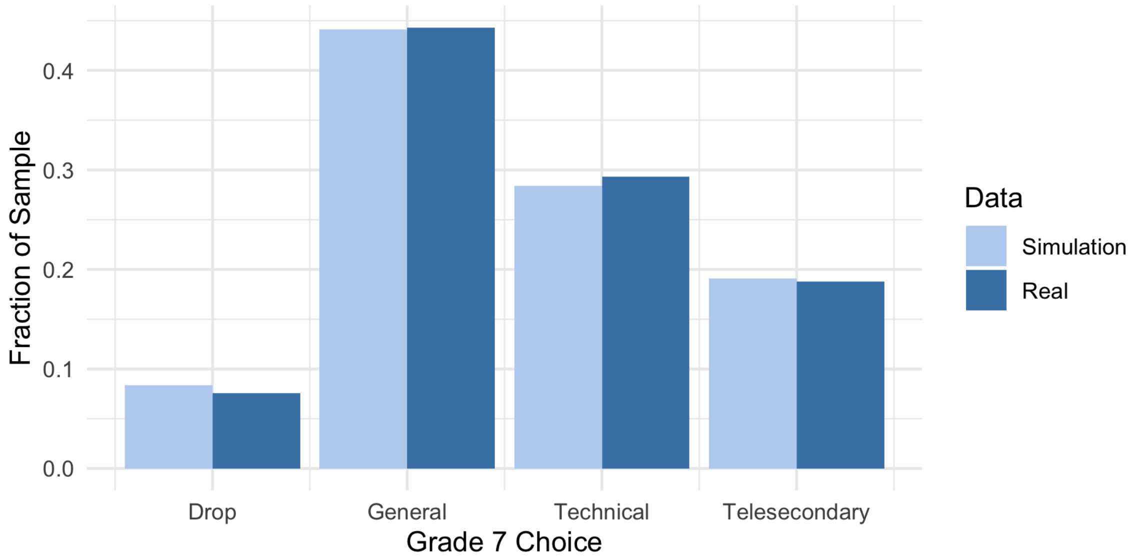
The correlation between study effort and Grade 7 test scores. The effort variable on the x-axis is binned into eight categories, and the mean test scores for students with effort values in the corresponding bin are calculated. There is a positive relationship between effort and test scores.

Figure 7: Enrollment Rates and Access to Middle Schools



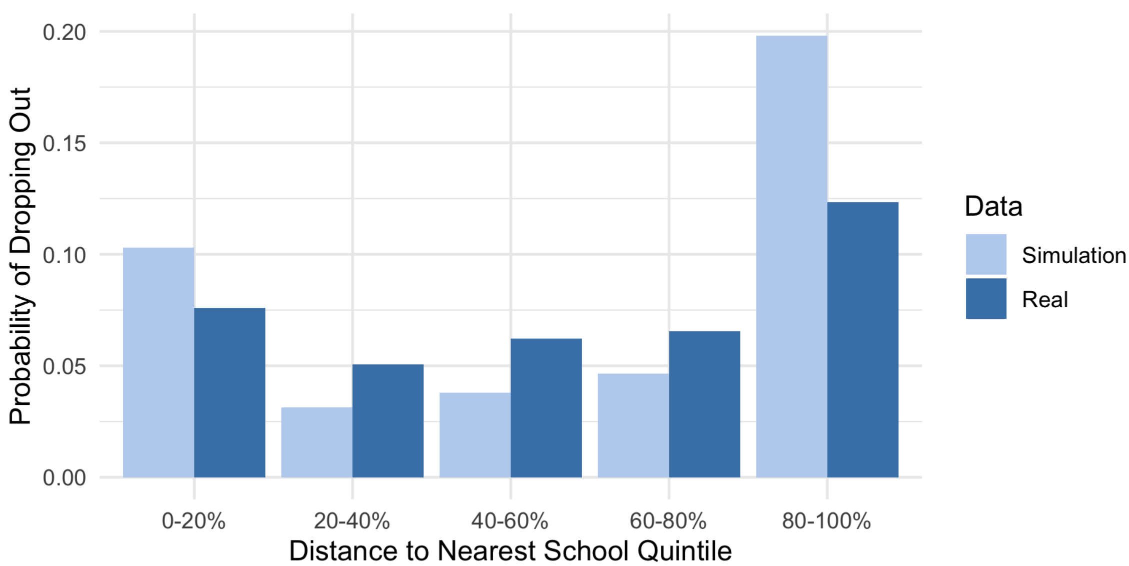
The relationship between enrollment rates and middle school accessibility. Each primary school is categorized by the distance to the nearest middle school, which is the x-axis. For each primary school, the fraction of its students who continue to Grade 7 is also calculated. The graph shows that primary schools that have middle schools near by have a higher fraction of their students enroll in middle school.

Figure 8: Goodness of Fit: School Types



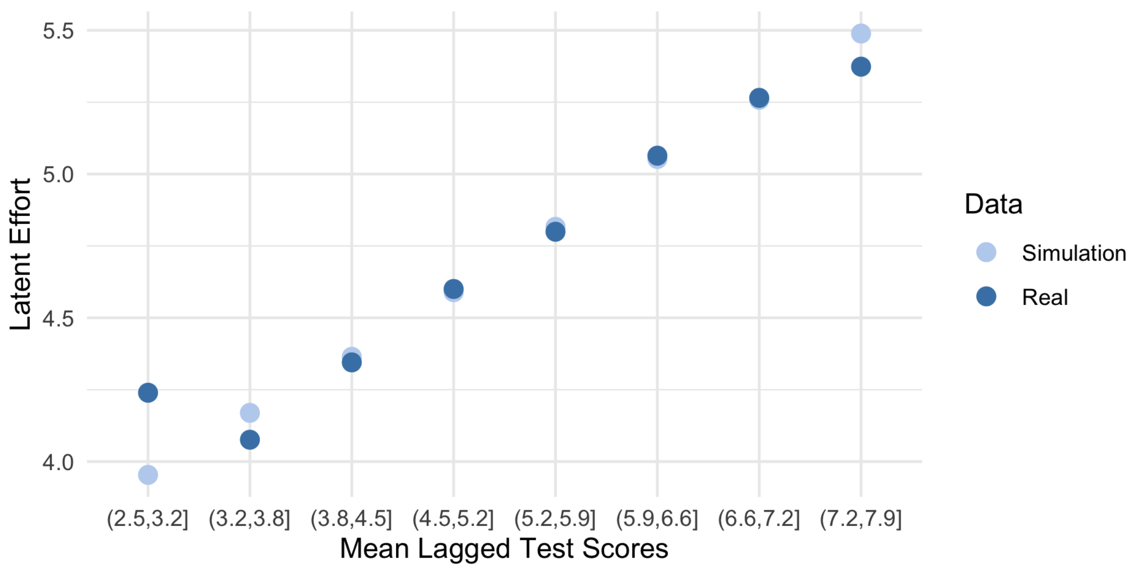
This figure shows the goodness of fit with respect to the choice of school type. The fraction of students choosing each of the three school types or dropping out in Grade 7 is included both in the data and in the estimated model.

Figure 9: Goodness of Fit: Enrollment and Accessibility



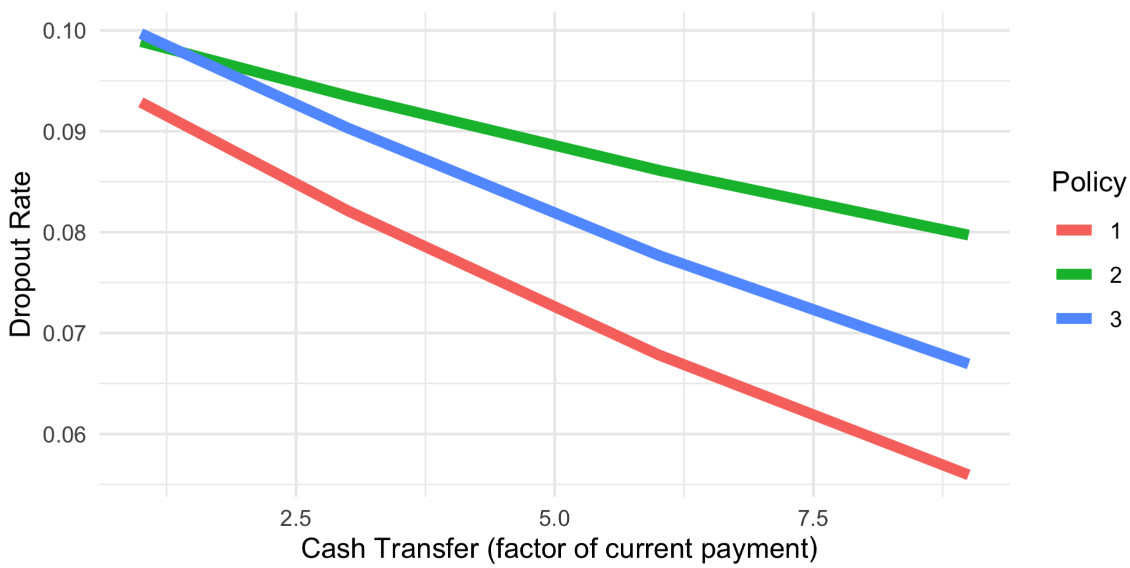
This figure shows the goodness of fit with respect to the enrollment rates and the school accessibility. The fraction of students who dropout is broken down by how far away the nearest school is from their primary school.

Figure 10: Goodness of Fit: Effort and Achievement



This figure shows the goodness of fit of the model with respect to the relationship between effort and achievement. The effort variable from the data and the effort variable generated by the model are plotted as a function of average lagged test scores.

Figure 11: Counterfactual 3: Dropout Rate and Conditional Cash Transfers



The fraction of students who dropout when considering three different conditional cash transfer policies. Policy 1 offers the transfer to any student who wants to drop out. Policy 2 offers the transfer to current beneficiaries. Policy 3 offers the transfer to current beneficiaries and students whose family earns below the median income.

CHAPTER 2 : The Marginal Returns to Distance Education: Evidence from Mexico's Telesecundarias, with Emilio Borghesan

2.1. 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 both the definition of distance learning used and 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 telesecundarias, 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

[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.

and a telesecundaria 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 telesecundaria 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 telesecundarias 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 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 observed 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 telesecundaria. The second is the distance between their primary school and the nearest traditional secondary school. Our instrument is the difference between these two measures: distance to telesecundaria minus distance to traditional school.[3]

This instrumental variable is highly predictive of attendance in telesecundarias: A one kilometer reduction, meaning the telesecundaria becomes relatively closer, causes the student's probability of attending a telesecundaria 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 travel to secondary school in our outcome equations and find that our results are unchanged.

We find that telesecundarias are highly beneficial: The average treatment effect (ATE) of telesecundaria 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 telesecundarias: some students see gains of over 0.5 standard deviations while others experience no benefit.

[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).

Our analysis uncovers a pattern of nonmonotonic selection into telesecundarias. 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 telesecundarias 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 telesecundarias.

We use our estimated MTEs to evaluate counterfactual policies that expand the availability of telesecundarias. The first policy we consider is a dramatic increase in telesecundaria availability that reduces the distance to telesecundarias for everyone in the sample by 5 km. The second policy under consideration is a school-construction program that builds a telesecundaria adjacent to the 18% of Mexican primary schools without a telesecundaria 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 telesecundaria. Nonparametric identification turns out to be important, as the MTE is nonmonotonic and parametric approaches fail to identify this feature[4]. In addition, our paper is the first to consider the impact of telesecundarias 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 telesecundarias on academic achievement.

[4]Specifically, parametric approaches based on jointly normal observables, common in this literature, would be unable to identify this feature. A fifth-order polynomial would capture the MTE curve well, however it would be impossible to know this a priori, as it requires being able to observe the shape of the nonparametric curve.

2.2. Related Literature

A large literature examines the interaction of technology and education and is surveyed in three excellent recent papers by Bulman and Fairlie (2016), Escueta, Quan, Nickow, and Oreopoulos (2017), and Rodriguez-Segura (2020). Papers that examine the effectiveness of distance learning in developing countries are much fewer in number. Beg, Lucas, Halim, and Saif (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.

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, Dustmann, Raute, and Schönberg (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.

Although telesecundarias 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 (2020), 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 telesecundarias raise educational attainment and future income, although the estimates in Navarro-Sola (2019) are substantially larger than those in Fabregas (2020). 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 telesecundarias.

2.3. 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* were seen as a solution to this problem. Telesecundarias 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 in Mexico City 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 telesecundarias 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 telesecundaria 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)).

Telesecundarias 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 17). A reform in 1993 mandated schooling through grade nine and resulted in increases in the construction of both new telesecundarias and telesecundaria enrollment. We study a single cohort, those students who were in grade six in 2007/08, after secondary schooling became compulsory.

2.4. Model

2.4.1. Student Achievement

We apply the potential outcomes framework of Rubin (1974) to a value-added model of learning. Students can either attend a telesecundaria or a traditional school.[5] We define the random variable D where $D = 1$ denotes attendance in a telesecundaria and $D = 0$ denotes attendance in a traditional school. We study the effects of telesecundaria attendance on two outcomes: math and Spanish test scores in seventh

[5]Mexico has three secondary school types: General, Technical, and Telesecundaria. We consider the choice between a traditional school (General/Technical) and a telesecundaria. Table 16 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 telesecundaria education relative to the next best alternative. A small fraction (less than 8% of full sample) of students in Mexico attend Private schools. We exclude them from the analysis. Students who drop out are also omitted (less than 8% of full sample).

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 telesecundaria, 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):

$$Y_1 = X\beta_1 + U_1 \ , \tag{2.1}$$

$$Y_0 = X\beta_0 + U_0 \ , \tag{2.2}$$

$$D = \mathbb{1}(\mathbf{Z}\gamma > V) \ , \tag{2.3}$$

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 (2.1) and (2.2) are value-added equations: X contains the previous year's test scores in both math and Spanish[6]. 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 telesecundaria relative to a traditional school on 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

[6] We assume that lagged test scores are sufficient to capture each student's human capital accumulation up to this point

econometrician only observes one of the two potential outcomes, $Y = Y_0 + D(Y_1 - Y_0)$.

The instrument, Z, is an exclusion that must simultaneously induce variation in the choice of school conditional on the covariates, X, and have no direct effect on the outcome variable. When Z is a sufficiently strong instrument, as in this paper, it can shift the probability of attending a telesecundaria continuously between 0 and 1. Such a shift would cause even the most unlikely student to attend a telesecundaria, thus allowing us to compare Y_0 and Y_1 for all individuals.

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 (2.3), represents this resistance to treatment. An individual with a higher V is, on the basis of unobservables, less likely to attend a telesecundaria. We apply the following useful transformation to equation (2.3) 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

$$MTE(x, u_D) = \mathbb{E}[Y_1 - Y_0 | X = x, U_D = u_D] \, . \tag{2.4}$$

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

2.4.2. 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 telesecundarias 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)):

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

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.

2.4.3. Parameters of Interest

A large class of parameters corresponding to the effect of telesecundaria 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:

$$PRTE_{a',a}(X)\frac{\mathbb{E}[Y_{a'} - Y_a \mid X = x]}{\mathbb{P}(D_{a'} = 1 \mid X = x) - \mathbb{P}(D_a = 1 \mid X = x)} \ . \tag{2.5}$$

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:

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

$$TT = \int MTE(X, U_D)dF_{U_D,X|D=1}(x, u_D \mid D = 1) \ , \tag{2.6}$$

$$TUT = \int MTE(X, U_D)dF_{U_D,X|D=0}(x, u_D \mid D = 0) \ , \tag{2.7}$$

$$PRTE_{a',a} = \int MTE(X, U_D)dF_{U_D,X|D_a=0,D_{a'}=1}(x, u_D \mid D_a = 0, D_{a'} = 1) \ . \tag{2.8}$$

Section 2.6 discusses the methods we use to integrate the MTE to obtain these treatment parameters.

2.5. 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 Evaluación National de Logro Académico en Centros Escolares, from now on 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 for 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 telesecundaria 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 are not available in any of our sources. In section 2.5.1, we show that the instrument that we have constructed is highly predictive of attendance in telesecundarias.

Figure 12 shows the distribution of the instrument by school attendance. A negative value on the x-axis indicates that a telesecundaria 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 2.7, we discuss how omitting dropouts influences the interpretation of our counterfactual analysis.

Table 15 lists all the outcome variables (Y), covariates (X), and the instrumental variable ($Z \backslash X$) that we use in our empirical analysis. Apart from the outcome variables, sixth grade test scores, and the instrument, all variables are categorical.

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 15: Variables Used in Estimation

We present summary statistics for these variables in Table 17 by school type. The table reveals that students who attend telesecundarias 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 mothers, and they fare worse academically in the year prior to secondary school.[7] 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 telesecundaria attendance.

2.5.1. 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 telesecundarias conditional on observable covariates, X, is easily verified. Table 18 displays the average marginal effects of each variable in Z on the probability of attending telesecundaria (estimated via Probit). The average marginal effect of relative distance on the probability of attending telesecundaria 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 child characteristics, any threats to instrument exogeneity, specifically Assumption (A-1), must be caused by a correlation between distance and unobserved *time-varying*

[7] 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.

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 2.5, 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 (2.1) and (2.2). Reassuringly, our estimates of the MTE and treatment parameters are unaffected by its inclusion.

2.6. 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 B.1, 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:

$$\begin{pmatrix} U_1 \\ U_0 \\ V \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{10} & \sigma_{1V} \\ & \sigma_0^2 & \sigma_{0V} \\ & & 1 \end{pmatrix} \right] . \tag{2.9}$$

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:

$$\begin{aligned}
\mathbb{E}(Y_1 \mid D = 1, X, Z) &= X\beta_1 + \mathbb{E}(U_1 \mid D = 1) \ , \\
&= X\beta_1 + \sigma_{1V}\left(-\frac{\phi(\Phi^{-1}(P(Z)))}{P(Z)}\right) \ , \\
\mathbb{E}(Y_0 \mid D = 0, X, Z) &= X\beta_0 + \mathbb{E}(U_0 \mid D = 0) \ , \\
&= X\beta_0 + \sigma_{0V}\left(\frac{\phi(\Phi^{-1}(P(Z)))}{1 - P(Z)}\right) \ .
\end{aligned}$$

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).[8] 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 (2.1) and (2.2) are linear, yields a conditional expectation function that is linear in X and XP and nonlinear in the propensity score, P. The conditional expectation function is defined below, where K(P) is some nonparametric function of

[8]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).

the propensity score:

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

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

$$MTE(X,P) = X(\beta_1-\beta_0) + \frac{\partial K(P)}{\partial P}\ . \tag{2.10}$$

The semiparametric estimator of (2.10) 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 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:

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

where $D_i = 1$ if an individual attends a telesecundaria and 0 otherwise, N_T denotes the number of students attending telesecundarias, 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 (2.6) - (2.8) 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 16 displays the densities used to compute ATE, TT, and TUT plotted as a function of U_D. The figures show that 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.

2.6.1. Empirical Results

Figure 14 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 15 repeats the analysis with Spanish as the outcome variable. In both figures, the horizontal axis measures the latent variable U_D, while the vertical axis measures the expected benefit to attending a telesecundaria 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 telesecundarias (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 telesecundarias 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 telesecundarias 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 telesecundaria 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 19 and 20 present estimates of standard treatment parameters for math and Spanish, respectively. All treatment parameters are positive, underscoring the findings from the MTE curve that telesecundaria 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 telesecundaria 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.

2.6.2. *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 21 and 22 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 21 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 telesecundarias is correlated with unobserved determinants of student achievement.

2.6.3. Explaining the Pattern of Selection

The MTEs for both math and Spanish are nonmonotonic and display evidence of reverse selection on gains. Reverse selection on gains occurs when students who select into the treatment (telesecundaria attendance) benefit less than students who are untreated. In this section we discuss the source of this nonmonotonicy and what it implies for the unobserved outcomes at both telesecundarias and traditional schools.

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

$$MTE(X, U_D) = X(\beta_1 - \beta_0) + k_1(U_D) - k_0(U_D) \ , \tag{2.11}$$

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

In this section, we estimate $k_1(U_D)$ and $k_0(U_D)$ separately to determine whether the shape of the MTE curve is determined primarily by variability in match qualities between students and telesecundarias 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):

$$\begin{aligned} k_1(p) &= p\frac{\partial K_1(p)}{\partial p} + K_1(p) \\ k_0(p) &= -(1-p)\frac{\partial K_0(p)}{\partial p} + K_0(p) \ . \end{aligned} \tag{2.12}$$

Figure 17 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 2.12. The bandwidths used are the same as in the estimation of the MTE function in Section 2.6: 0.28 for both math and Spanish.

Figure 17 demonstrates that variation in mean unobserved outcomes at telesecundarias, $k_1(U_D)$, 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 telesecundarias. Essentially, telesecundarias and traditional schools are equally good for students who are most likely to attend telesecundarias. But the quality of the match at traditional schools rapidly declines as students become less likely to attend telesecundarias (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 telesecundarias 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 telesecundarias, reflected in the nonzero slope of $k_1(U_D)$. It is puzzling that students who have the highest benefit from attending telesecundarias, 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 their expected future 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. A recent paper by Ainsworth, Dehejia, Pop-Eleches, and Urquiola (2020) finds that students in Romania could have chosen a school with an average of 1 standard deviation higher value-added than the one they attended, indicating that students may not be choosing schools purely based on value-added, or that families may not be aware of each school's value-added.

2.6.4. 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 (2.1) and (2.2) 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.2. We find that including distance traveled to secondary school makes no difference for math scores but causes a small reduction in Spanish scores. Figures B-2 and B-3 display the estimated MTE curves for these specifications. The pattern of selection, given by the shape of the MTE curves, as well as the point estimates of treatment parameters presented in Table B-1 are little changed for math relative to our main specification. Unlike in the main specification, the Spanish MTE has does not trend upward in the noisily-estimated right tail. We estimate an ATE of 0.324 standard deviations for math scores and 0.152 for Spanish (see Table B-2). The treatment parameters for math are all within one standard errors of the estimates generated by our main specification. The treatment parameters for Spanish are lower, by about two and half standard errors, but still significant.

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.2. 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.

2.7. 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 telesecundarias who would not otherwise attend. The treatment effect for students induced to switch attendance from traditional to telesecundarias as a result of the policy is given $PRTE_{a,a'}$, which is positive if these students learn more in telesecundarias than traditional schools.

We consider a class of policies that expand access to telesecundarias 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 telesecundarias instead of dropping out as a result of the counterfactual policy.[9] 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. In this paper we are only measuring the achievement effects of this policy, not the enrollment effects. In rural areas, distance to school is one of the main barriers to attending school, so a school building policy would

[9]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.

inevitably have a positive impact on enrollment as well.

We consider two counterfactual policies. The first is a hypothetical policy that reduces the relative distance to telesecundarias 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 telesecundaria is under 5 km. The policy counterfactual is merely analyzed as an example of the gains to a policy that can drastically raise telesecundaria attendance.

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

Figure 18 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 telesecundaria 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.

Table 23 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 telesecundarias 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.

2.8. 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 telesecundarias, but that nearly all students benefit from telesecundaria 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 telesecundarias more accessible.

Telesecundarias 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 telesecundarias play in improving student learning and can change this perception. We predict that future expansions of telesecundarias 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 telesecundaria schooling, there are likely to be significant long-term gains to a policy that expands access to telesecundarias.

2.9. Tables

Table 16: Summary Statistics by Middle School Type

	General	**Technical**	**Telesecundaria**	**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 Telesecundaria – 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 17: Summary Statistics

	Telesecundaria		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 18: 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 telesecundaria 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 telesecundaria, while the second is the distance from the student's primary school to the nearest traditional school. Relative distance is negative whenever telesecundarias 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 19: 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 telesecundaria 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 16. 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 20: 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 telesecundaria 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 16. 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 21: 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 hypothesis that the difference between adjacent LATEs is equal to zero. p-values are obtained through 50 bootstrap replications.

Table 22: 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.

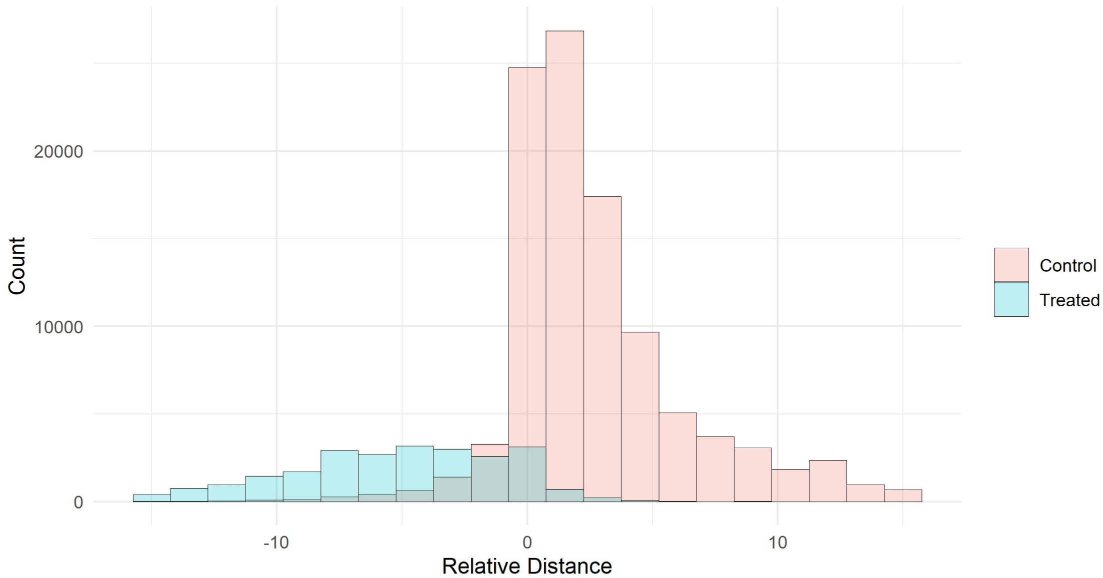
Table 23: Counterfactual Treatment Effects

	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 2.7. PRTE1 corresponds to a counterfactual policy that reduces relative distance between telesecundarias and general secondary schools by 5 km. PRTE2 corresponds a counterfactual policy that constructs telesecundarias adjacent to all primary schools that do not have a telesecundaria 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 telesecundaria attendance.

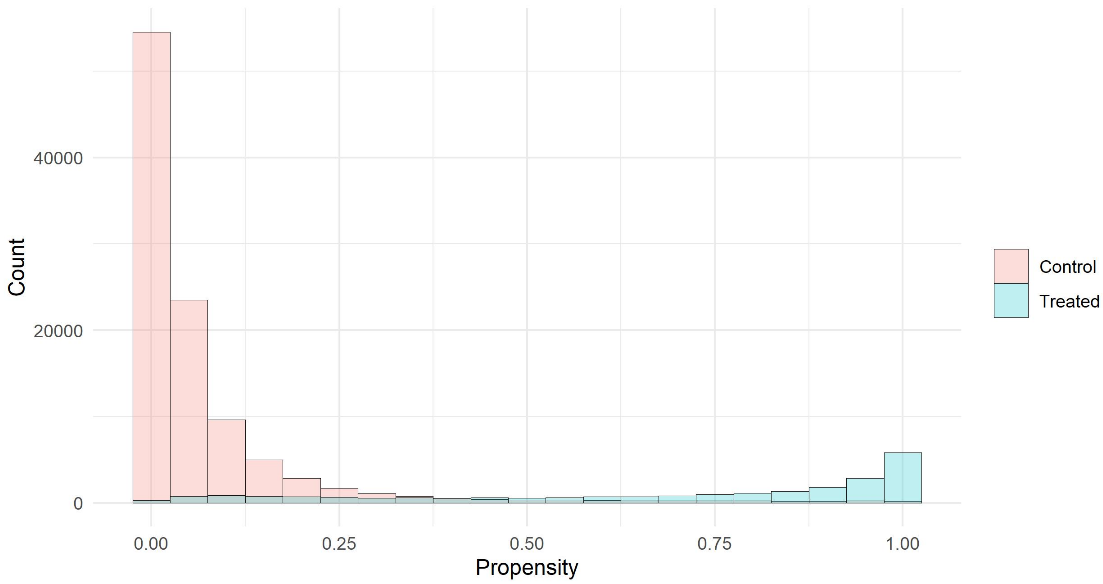
2.10. Figures

Figure 12: Relative Distance to Telesecundaria



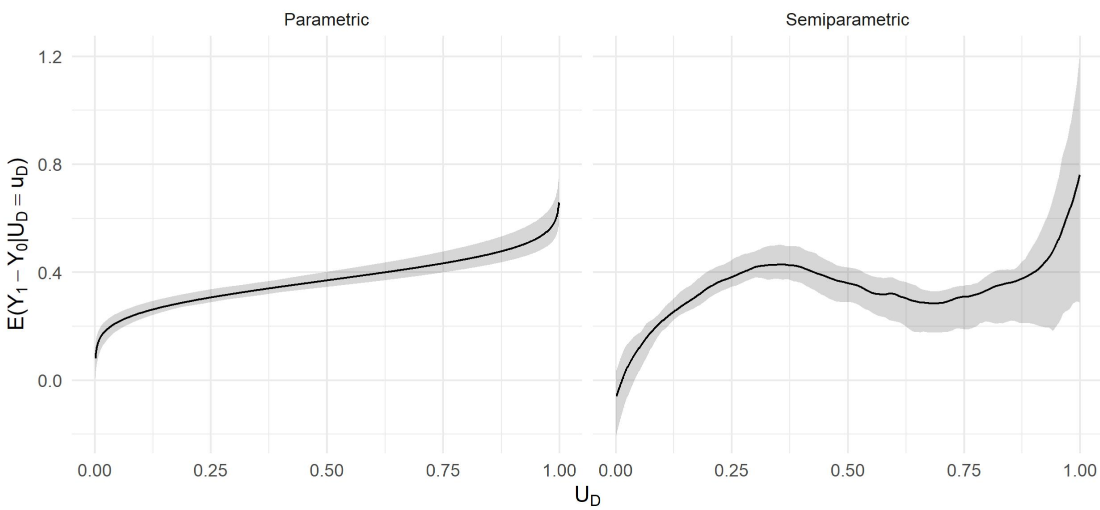
The figure plots histograms of the instrumental variable by treatment status. Control units refer to students in traditional schools, while treated units refer to students in telesecundarias. The instrument is the difference between two measures of distance. The first is the distance from the student's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

Figure 13: Estimated Propensity Score by Treatment Status



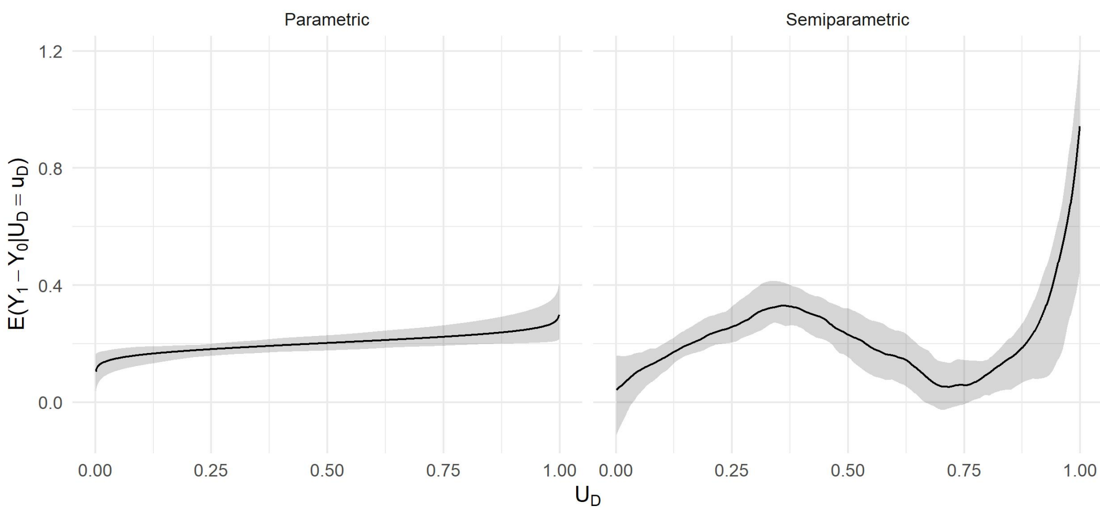
The figure plots histograms of estimated propensity scores by treatment status. Control units refer to students in traditional schools, while treated units refer to students in telesecundarias. Propensity scores model telesecundaria attendance as a function the child's sixth grade math and Spanish scores, 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 telesecundaria 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 telesecundaria, 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 14: Marginal Treatment Effect: Math



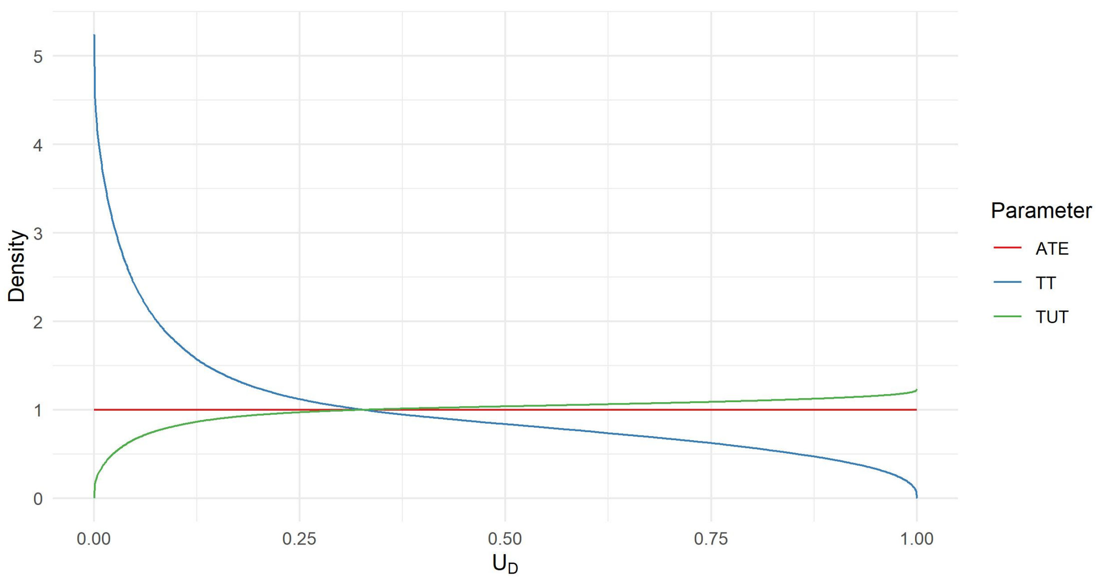
The dependent variable in the outcome equation is the score on the seventh grade nationally-standardized (ENLACE) math exam. The outcome variable has been standardized and is measured in standard deviations from the mean. The outcome equations include controls for sixth grade math and Spanish scores, 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 telesecundaria 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 15: Marginal Treatment Effect: Spanish



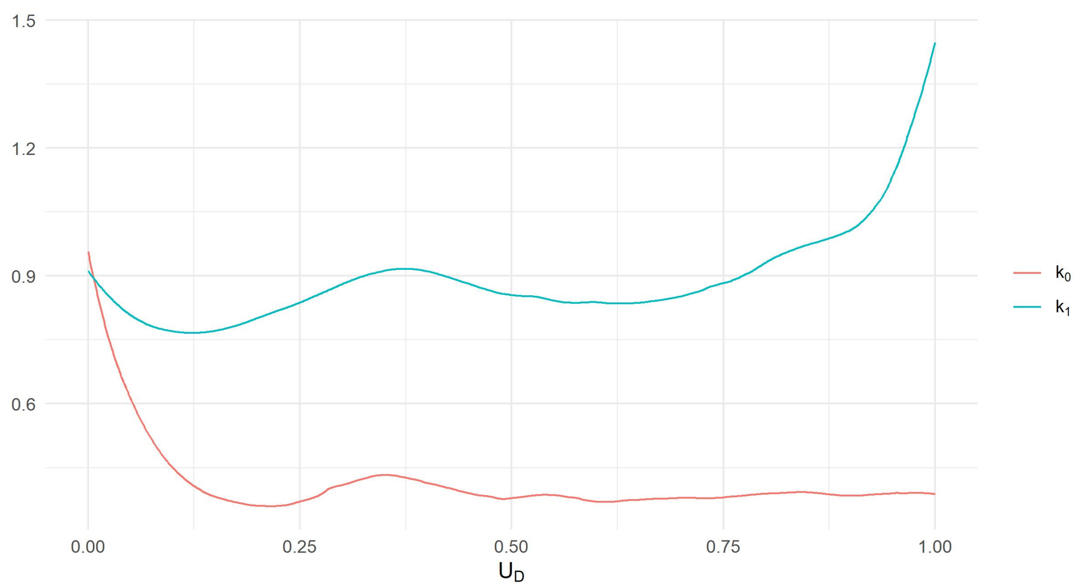
The dependent variable in the outcome equation is the score on the seventh grade nationally-standardized (ENLACE) Spanish exam. The outcome variable has been standardized and is measured in standard deviations from the mean. The outcome equations include controls for sixth grade Math and Spanish scores, 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 telesecundaria 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 16: Treatment Parameter Weights



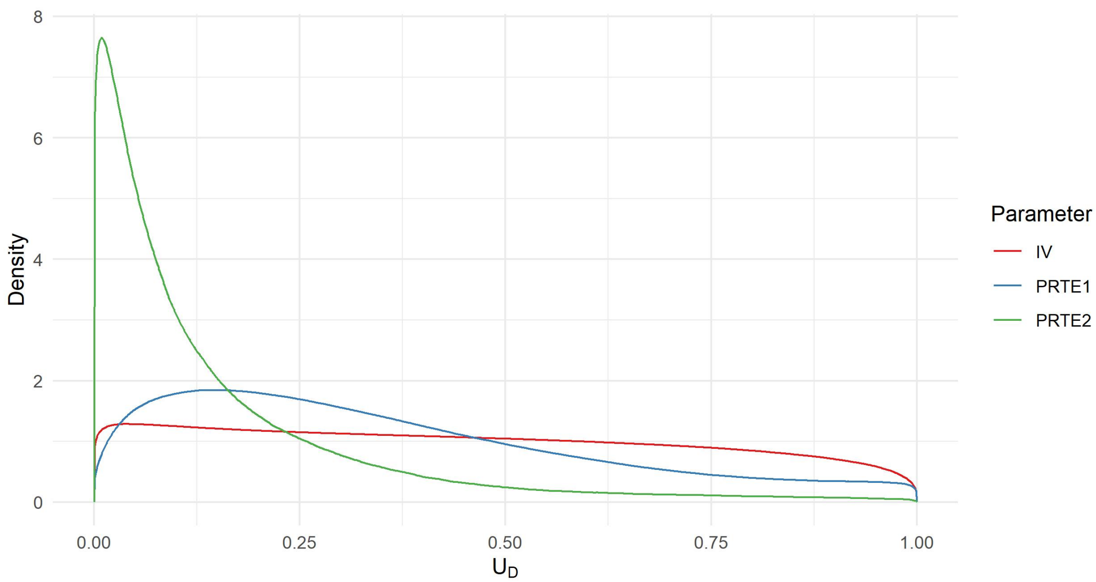
The figure shows the distribution of weighting functions used to construct three standard treatment parameters. The average treatment effect (ATE) integrates the MTE with respect to the unit uniform distribution. The average effect of treatment on the treated (TT) integrates the MTE with respect to the distribution of U_D conditional on attendance in telesecundarias, $f_{U_D,X|D=1}(x, u_D \mid D = 1)$, 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 17: The Source of Reverse Selection



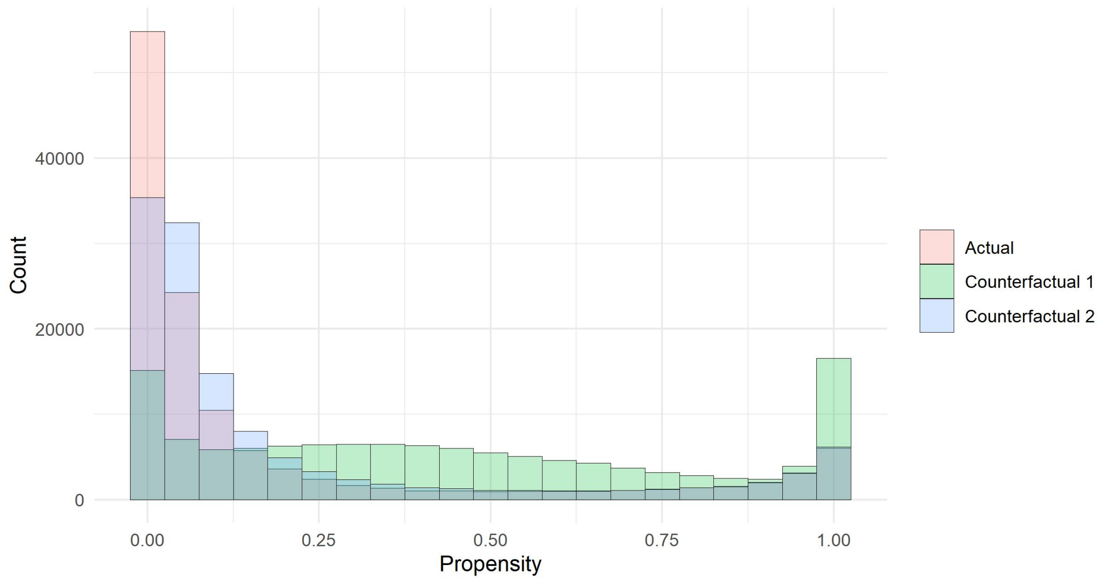
The figure plots $k_1(U_D) = \mathbb{E}[U_1 \mid U_D]$ and $k_0(U_D) = \mathbb{E}[U_0 \mid U_D]$ evaluated at $X = \overline{x}$ on the vertical axis against U_D on the horizontal axis. U_1 is the child's unobserved outcome in the equation for math value-added in telesecundarias. U_0 is the child's unobserved 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 2.6.3.

Figure 18: 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 2.7 as well as the weights induced by Two-Stage Least Squares which uses relative distance as an instrument for telesecundaria attendance. PRTE1 corresponds to the weights induced by a counterfactual policy which reduces relative distance between telesecundarias and traditional secondary schools by 5 km. PRTE2 corresponds to the weights induced by a counterfactual policy that constructs telesecundarias adjacent to all primary schools that do not have a telesecundaria within a 5 km radius.

Figure 19: Histogram of Propensity Scores under Counterfactual Policies



The figure plots histograms of the probability of attending a telesecundaria under the current policy (2008) as well as two counterfactual policies. Counterfactual 1 reduces relative distance between telesecundarias and general secondary schools by 5 km. Counterfactual 2 is a school-building policy that constructs telesecundarias adjacent to all primary schools that do not have a telesecundaria within a 5 km radius. The counterfactual policies are discussed in greater detail in section 2.7.

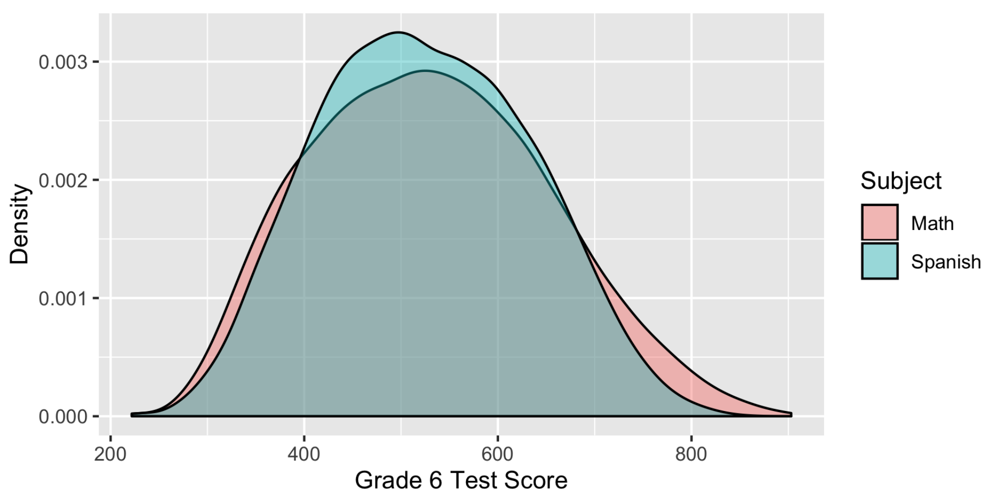
APPENDIX

A. Appendix - Chapter 1

A.1. Histograms of Raw Test Scores

Figure A-1: Grade 7 Test Score Distribution

Figure A-2: Grade 6 Test Score Distribution



A.2. Factor Analysis for Effort Questions

I use factor analysis to estimate the latent effort variable. I am assuming that there is a true unobserved latent effort variable, and that the five questions that I observe are all affected by the latent variable. Formalizing this, I assume that the unobserved latent effort variable $\hat{e}_i^M$ is connected to the five measures in the data $(e_{i1}^M, ..., e_{i5}^M)$ in the following way,

$$e_{i1}^M = \gamma_1 \hat{e}_i^M + u_{i1}$$
$$\vdots$$
$$e_{i5}^M = \gamma_5 \hat{e}_i^M + u_{i5}$$

First, I compute the correlation matrix of the five measures in the data. Because four of the measures are ordinal variables, I compute a polychoric correlation matrix. This follows the practice in the literature, and the main assumption is that the ordinal variables have an underlying joint continuous distribution. The polychoric correlation matrix for my five measures of effort is calculated to be:

Table A-1: Polychoric Correlation Matrix for Effort Variables

	Pay Attention	Participate	Miss School	Skip Class	Study Hours
Pay Attention	1.00				
Participate	0.43	1.00			
Miss School	-0.23	-0.13	1.00		
Skip Class	-0.24	-0.14	0.25	1.00	
Study Hours	0.28	0.20	-0.11	-0.08	1.00

The signs of the correlations are as would be expected, with paying attention in class, participating in class and the number of hours studied per day all positively correlated with each other, and negatively correlated with missing school and skipping class.

To compute the factor loadings and get an estimate for the latent effort variable I use the Principal Axis method. This is an iterative procedure, and iterates until the communalities of each of the measures do not vary by iteration. Communalities are defined as the component of the variance of each of the measures that are shared, and therefore can be attributed to the latent factor. The initial guess of the communality of a given variable comes from the R^2 of the regression using that variable as the independent variable, and the other four measures as the dependent variables. These initial guesses replace the diagonal elements of the correlation matrix. Then, an eigendecomposition is done of this updated correlation matrix. Using the eigenvalues and eigenvectors, new communalities can be computed. This is repeated, until the communalities stabilize. After convergence, the loadings are extracting using the eigenvalues and eigenvectors of the final matrix.

The loadings for each of the effort variables are,

Table A-2: Loading Factors for the Effort Variables

Variables	Loadings
PayAttention	0.77
Participate	0.53
MissSchool	-0.34
SkipClass	-0.34
StudyHours	0.35

To get an estimate of the latent effort variable $\hat{e}_i^M$ for each student i, I multiply their effort measures by the associated loading factor.

$$\hat{e}_i^M = l_1 * e_{i1}^M + ... + l_5 * e_{i5}^M$$

The result is a continuous effort variable for each student, that has greater variance than any of the individual measures used to compute it. Figure 3 shows a histogram of the final effort measures.

A.3. Estimation Strategy Details

1. Guess parameters. There are 51 parameters in this version of the model:

- 15 coefficients for each of the achievement value added equations
- 3 parameters in the variance-covariance matrix for the value added equations
- 17 coefficients in the utility equation
- 1 parameters for the standard deviation of the effort distribution

2. For each student, compute their individual likelihood given the guessed parameters and data:

   - Compute the effort implied by the model (for all options in the student's choice set) using Equation 1.2.
   - Compute expected math and Spanish scores using effort and Equation 1.1 (without the error terms since it is an expectation).
   - For students who enrolled in Grade 7, compute the achievement and effort probabilities. (For students who did not enroll, assign a value of 1 to these probabilities.)
   - For all students, compute the multinomial logit probability given in Equation 1.4.
   - Take the product of the three probabilities.

3. Take the log of each individual likelihood, and sum them. Maximize this value with respect to all of these parameters.

A.4. Creating Estimation Sample

For students to be in my main estimation sample, I require data on their Grade 6 and 7 ENLACE tests, as well as survey responses from the student and their parent in Grade 7. I am using the survey responses in Grade 7, since I need to know if the students are working or not in that year.

Unfortunately, this sample excludes any students who dropped out between Grade 6 and Grade 7, and I want to model this behavior as well. To incorporate these students into the sample, I randomly select students who wrote the ENLACE tests in Grade 6, have student and parent surveys from Grade 6, and dropped out after Grade 6 and include them in my estimation sample. The number of students I include is chosen so that the dropout ratio is the same as in the full dataset. In doing this, I am assuming that some of the background information from the survey, such as parental education, are constant over these two years.

A.5. Calculating the Standard Errors

Standard errors are calculated using a sandwich-type covariance matrix (Yuan et al., 2014). Define the log likelihood for student i given parameters Ω as $L_i(\Omega)$. As detailed in the estimation section, I am able to calculate such probabilities using the data and parameters. To estimate the covariance matrix with a sample of n students, I use the following formula:

$$\hat{\text{Cov}} = \frac{\hat{A}^{-1}\hat{B}\hat{A}^{-1}}{n} \tag{A-1}$$

where

$$\hat{A} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 L_i(\hat{\Omega})}{\partial\hat{\Omega}\partial\hat{\Omega}'}$$

$$\hat{B} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial L_i(\hat{\Omega})}{\partial\hat{\Omega}}\right]\left[\frac{\partial L_i(\hat{\Omega})}{\partial\hat{\Omega}}\right]'$$

The matrix $\hat{A}$ is an estimation for the Hessian, and the matrix $\hat{B}$ is an estimation of the outer product of the gradient. I calculate the gradient and the Hessian numerically in R, using the functions *grad()* and *hessian()*. To get the final standard errors, I take the square root of the diagonal elements of the covariance matrix.

A.6. Wage Regressions

The data used to estimate the wage regressions comes from the Mexico 2010 Census, and can be accessed through the IPUMS site: https://international.ipums.org/international-action/variables/search. The variables that are downloaded are:

- Age of subject (MX2010A AGE)
- Whether or not the subject currently attends school (MX2010A_SCHOOL)
- Income of individual for the last month (MX2010A_INCOME)
- Household's income from work (MX2010A_INCHOME)
- Number of hours worked by individual in the last week (MX2010A_HRSWORK)
- Educational attainment level of individual in number of years (MX2010A_EDATTAIN)
- Educational attainment level of mother in number of years

MX2010A_EDATTAIN_MOM)

- Educational attainment level of father in number of years (MX2010A_EDATTAIN_POP)
- Gender (MX2010A_SEX)
- Employment status (MX2010A_EMPSTAT)
- Position at work (MX2010A_CLASSWK)
- State code (GEO1_MX2010)
- Municipality code (GEO2_MX2010)
- Urban-rural status (URBAN)

In order to compute the regressions, we recode several new variables from the ones listed above:

- **INCOME_PER_HOUR** (Created by dividing income last month by 4 times the number of hours worked last week)
- **familyworker** (Dummy for whether the individual is an unpaid family worker)
- **mom_edattain_missing** (Created from MX2010A_EDATTAIN_MOM variable; 1 = mom's educational attainment is missing, 0 = mom's educational attainment is not missing)
- **dad_edattain_missing** (Created from MX2010A_EDATTAIN_POP variable; 1 = dad's educational attainment is missing, 0 = dad's educational attainment is not missing)

- **north_dummy** (Dummy for whether municipality is in the North or South region of Mexico ; 1 = North, 0 = South)

The next steps are to clean and filter the data:

1. Exclude individuals with an undefined age (include only MX2010A_AGE != 999)

2. Exclude individuals with undefined school attendance status (include only MX2010A_SCHOOL == 1 |MX2010A_SCHOOL == 2)

3. Assign 0 to missing or unknown values for monthly personal income (MX2010A_INCOME), monthly family income (MX2010A_INCHOME), and hours worked in the last week (MX2010A_HRSWORK)

4. Create Income Per Hour variable by dividing MX2010A_INCOME (monthly income) by 4 times MX2010A_HRSWORK (hours worked in the last week) and assign infinite and undefined values to 0

5. Reassign educational attainment variables (MX2010A_EDATTAIN, MX2010_EDATTAIN_MOM, MX2010A_EDATTAIN_POP) values with continuous values

6. Create mom_edattain_missing and dad_edattain_missing variables by assigning a 1 for these variables if the MX2010A_EDATTAIN_MOM and MX2010A_EDATTAIN_POP are missing or unknown, respectively, and a 0 if not

7. Include only individuals with educational attainment levels equal to or below 13 (MX2010A_EDATTAIN <= 13) in order to exclude students who have finished

high school

8. Create a dummy variable (mun_dummy) that indicates whether (1) or not (0) the municipality the individual is in also contains a city with at population of at least 100,000 (merged with CityCoordinates_withMunicipalities file)

9. Create a north/south dummy (north_dummy) that indicates whether the municipality is in the northern (1) or southern (0) region of Mexico

10. Filter by age to only include inviduals between the age of 12 and 20 inclusive (MX2010A_AGE <= 20 & MX2010A_AGE >= 12)

11. Create cutoffs for INCOME_PER_HOUR, MX2010A_HRSWORK, MX2010A_INCHOME and exclude entries for each variable with values above the 99th quantile

12. Create a yeswork variable where yeswork = 1 if one of the following criteria are met:

    - MX2010A_EMPSTAT == 10
    - MX2010A_HRSWORK != 0
    - MX2010A_INCOME != 0
    - MX2010A_CLASSWK == 1
    - MX2010A_CLASSWK == 2
    - MX2010A_CLASSWK == 3
    - MX2010A_CLASSWK == 4

- MX2010A_CLASSWK == 5
- MX2010A_CLASSWK == 6

and MX2010A_HRSWORK >= 5 and INCOME_PER_HOUR >0

13. Create family net income (netincome) variable by subtracting individual's income from their entire family's income (which includes the individual's income): MX2010A_INCHOME - MX2010A_INCOME

14. Then create dummies for family income where

    - family_income1 includes netincome <1500
    - family_income2 includes 1500 <= netincome <3000
    - family_income3 includes 3000 <= netincome <7500
    - family_income4 includes 7500 <= netincome <15000
    - family_income5 includes 15000 <= netincome <30000
    - family_income6 includes netincome >= 30000

15. Create separate nonzero_data data set by filtering yeswork ==1

16. Separate into two data sets based on gender

Wage regressions are estimated on the two data sets separately, using a Heckman selection model. The first step is to run a probit model on the probability of working. The full dataset is used to estimate this probit. The wage regressors include: age, school attendance, educational attainment, parental educational attainment, parents are missing, urban-rural dummies, north-south dummies, and municipality dummies. In addition, the following variables are assumed to influence selection into working, but not the wage offers, and are included as exclusion restrictions: family income, home electricity, home piped water, home internet and home computer.

Using the results from the probit, it is possible to create control functions for each student. These are included as a regressor in the next step of the estimation process, which is a fixed effect linear regression model with hourly wages as the independent variable, and the regressors listed in the probit (without the exclusion restriction variables). The fixed effects in the model are the municipality fixed effects, which allow for great geographic heterogeneity. This regression is estimated using the subset of students who report working and earning a positive wage.

Results for the wage regressions are shown in Table A-3.

Figure A-3: Wage Regression Results: Hourly Wages

	Dependent variable:	
	INCOME_PER_HOUR	
	Boys	Girls
MX2010A_AGE	0.682^{***} (0.173)	0.593^{**} (0.239)
MX2010A_SCHOOL	5.109^{***} (1.346)	5.301^{***} (1.885)
mom_edattain_missing	−1.483 (1.517)	3.561 (2.438)
dad_edattain_missing	0.080 (1.317)	−0.051 (2.097)
MX2010A_EDATTAIN	0.394^{**} (0.175)	-0.459^{*} (0.279)
MX2010A_EDATTAIN_MOM	-0.535^{***} (0.163)	0.389 (0.265)
MX2010A_EDATTAIN_POP	−0.190 (0.179)	−0.039 (0.291)
URBAN	−0.096 (1.719)	5.753^{**} (2.442)
lamda_boys	0.023 (0.017)	
lamda_girls		−0.015 (0.029)
MX2010A_AGE:MX2010A_SCHOOL	-0.252^{***} (0.081)	-0.292^{***} (0.111)
MX2010A_AGE:mom_edattain_missing	0.080 (0.085)	−0.216 (0.135)
MX2010A_AGE:dad_edattain_missing	0.006 (0.074)	0.001 (0.117)
MX2010A_AGE:MX2010A_EDATTAIN	−0.014 (0.010)	0.042^{***} (0.015)
MX2010A_AGE:MX2010A_EDATTAIN_MOM	0.034^{***} (0.009)	−0.019 (0.015)
MX2010A_AGE:MX2010A_EDATTAIN_POP	0.013 (0.010)	0.006 (0.016)
MX2010A_AGE:URBAN	0.038 (0.102)	-0.306^{**} (0.141)
MX2010A_SCHOOL:URBAN	-2.634^{***} (0.786)	-3.666^{***} (1.104)
mom_edattain_missing:URBAN	0.840 (0.943)	−2.197 (1.459)
dad_edattain_missing:URBAN	0.081 (0.827)	0.105 (1.280)
MX2010A_EDATTAIN:URBAN	−0.019 (0.110)	0.072 (0.169)
MX2010A_EDATTAIN_MOM:URBAN	0.229^{**} (0.097)	−0.203 (0.154)
MX2010A_EDATTAIN_POP:URBAN	0.137 (0.106)	−0.021 (0.168)
MX2010A_AGE:north_dummy	-0.585^{***} (0.156)	−0.302 (0.231)
MX2010A_SCHOOL:north_dummy	-4.866^{***} (1.170)	0.438 (1.743)
mom_edattain_missing:north_dummy	−0.161 (1.621)	−2.976 (2.625)
dad_edattain_missing:north_dummy	−0.092 (1.461)	-4.095^{*} (2.409)
MX2010A_EDATTAIN:north_dummy	−0.080 (0.187)	−0.301 (0.301)
MX2010A_EDATTAIN_MOM:north_dummy	0.177 (0.159)	-0.504^{*} (0.259)
MX2010A_EDATTAIN_POP:north_dummy	0.113 (0.162)	−0.037 (0.260)
URBAN:north_dummy	0.144 (0.124)	−0.341 (0.213)
MX2010A_AGE:MX2010A_SCHOOL:URBAN	0.128^{***} (0.047)	0.190^{***} (0.064)
MX2010A_AGE:mom_edattain_missing:URBAN	−0.012 (0.053)	0.162^{**} (0.081)
MX2010A_AGE:dad_edattain_missing:URBAN	−0.006 (0.047)	0.012 (0.072)
MX2010A_AGE:MX2010A_EDATTAIN:URBAN	0.0001 (0.006)	−0.007 (0.009)
MX2010A_AGE:MX2010A_EDATTAIN_MOM:URBAN	-0.013^{**} (0.005)	0.013 (0.009)
MX2010A_AGE:MX2010A_EDATTAIN_POP:URBAN	−0.008 (0.006)	0.001 (0.009)
MX2010A_AGE:MX2010A_SCHOOL:north_dummy	0.278^{***} (0.068)	−0.029 (0.099)
MX2010A_AGE:mom_edattain_missing:north_dummy	0.021 (0.090)	0.159 (0.144)
MX2010A_AGE:dad_edattain_missing:north_dummy	−0.002 (0.081)	0.210 (0.133)
MX2010A_AGE:MX2010A_EDATTAIN:north_dummy	0.006 (0.010)	0.012 (0.016)
MX2010A_AGE:MX2010A_EDATTAIN_MOM:north_dummy	−0.007 (0.009)	0.028^{*} (0.014)
MX2010A_AGE:MX2010A_EDATTAIN_POP:north_dummy	−0.007 (0.009)	0.001 (0.014)
Observations	174,905	76,978
R^2	0.216	0.251
Adjusted R^2	0.205	0.226
Residual Std. Error	5.878 (df = 172417)	5.747 (df = 74506)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure A-4: Wage Regression Results: Hours Worked per Week

	Dependent variable:	
	MX2010A_HRSWORK	
	Boys	Girls
MX2010A_AGE	2.988*** (0.382)	3.731*** (0.592)
MX2010A_SCHOOL	39.657*** (2.972)	41.688*** (4.668)
mom_edattain_missing	0.205 (3.350)	5.856 (6.038)
dad_edattain_missing	−0.937 (2.907)	−9.230* (5.193)
MX2010A_EDATTAIN	−0.548 (0.387)	0.152 (0.690)
MX2010A_EDATTAIN_MOM	−0.865** (0.361)	−0.161 (0.655)
MX2010A_EDATTAIN_POP	−0.116 (0.395)	−0.980 (0.721)
URBAN	−8.085** (3.794)	−3.035 (6.047)
lamda_boys	−0.052 (0.037)	
lamda_girls		−0.160** (0.072)
MX2010A_AGE:MX2010A_SCHOOL	−1.499*** (0.179)	−1.689*** (0.274)
MX2010A_AGE:mom_edattain_missing	0.042 (0.187)	−0.556* (0.335)
MX2010A_AGE:dad_edattain_missing	0.028 (0.164)	0.475 (0.291)
MX2010A_AGE:MX2010A_EDATTAIN	0.042** (0.021)	−0.010 (0.038)
MX2010A_AGE:MX2010A_EDATTAIN_MOM	0.046** (0.020)	0.016 (0.037)
MX2010A_AGE:MX2010A_EDATTAIN_POP	0.012 (0.022)	0.054 (0.040)
MX2010A_AGE:URBAN	0.622*** (0.225)	0.258 (0.350)
MX2010A_SCHOOL:URBAN	1.999 (1.735)	2.944 (2.735)
mom_edattain_missing:URBAN	−1.320 (2.082)	−2.742 (3.613)
dad_edattain_missing:URBAN	0.270 (1.826)	6.199* (3.170)
MX2010A_EDATTAIN:URBAN	0.391 (0.242)	−0.345 (0.419)
MX2010A_EDATTAIN_MOM:URBAN	0.195 (0.214)	−0.540 (0.381)
MX2010A_EDATTAIN_POP:URBAN	−0.094 (0.234)	0.439 (0.417)
MX2010A_AGE:north_dummy	1.479*** (0.345)	1.338** (0.571)
MX2010A_SCHOOL:north_dummy	9.461*** (2.582)	5.570 (4.315)
mom_edattain_missing:north_dummy	1.222 (3.579)	4.526 (6.499)
dad_edattain_missing:north_dummy	1.838 (3.225)	2.504 (5.964)
MX2010A_EDATTAIN:north_dummy	−0.306 (0.414)	0.329 (0.746)
MX2010A_EDATTAIN_MOM:north_dummy	−0.353 (0.351)	0.609 (0.642)
MX2010A_EDATTAIN_POP:north_dummy	0.369 (0.357)	−0.131 (0.643)
URBAN:north_dummy	−1.337*** (0.274)	0.688 (0.527)
MX2010A_AGE:MX2010A_SCHOOL:URBAN	−0.092 (0.104)	−0.158 (0.160)
MX2010A_AGE:mom_edattain_missing:URBAN	0.063 (0.116)	0.208 (0.200)
MX2010A_AGE:dad_edattain_missing:URBAN	0.011 (0.103)	−0.339* (0.177)
MX2010A_AGE:MX2010A_EDATTAIN:URBAN	−0.031** (0.013)	0.019 (0.023)
MX2010A_AGE:MX2010A_EDATTAIN_MOM:URBAN	−0.012 (0.012)	0.021 (0.021)
MX2010A_AGE:MX2010A_EDATTAIN_POP:URBAN	0.001 (0.013)	−0.028 (0.023)
MX2010A_AGE:MX2010A_SCHOOL:north_dummy	−0.743*** (0.151)	−0.444* (0.246)
MX2010A_AGE:mom_edattain_missing:north_dummy	−0.066 (0.198)	−0.180 (0.356)
MX2010A_AGE:dad_edattain_missing:north_dummy	−0.087 (0.180)	−0.130 (0.329)
MX2010A_AGE:MX2010A_EDATTAIN:north_dummy	0.016 (0.023)	−0.020 (0.040)
MX2010A_AGE:MX2010A_EDATTAIN_MOM:north_dummy	0.021 (0.020)	−0.032 (0.035)
MX2010A_AGE:MX2010A_EDATTAIN_POP:north_dummy	−0.019 (0.020)	0.009 (0.036)
Observations	174,905	76,978
R^2	0.230	0.217
Adjusted R^2	0.219	0.191
Residual Std. Error	12.975 (df = 172417)	14.231 (df = 74506)
Note:		*p<0.1; **p<0.05; ***p<0.01

B. Appendix - Chapter 2

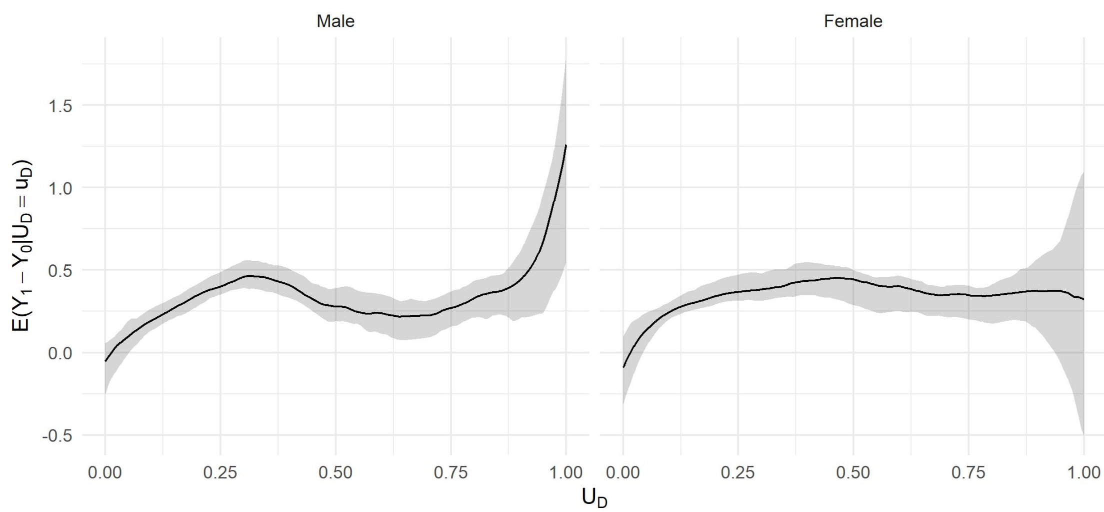
B.1. Letting the MTE vary by observables

Identification of the Marginal Treatment Effect does not require that X be exogenous. As we explain in Section 2.6, assuming that X is exogenous has several practical advantages, although it imposes a cost as well, that the slope of $MTE(X, U_D)$ in U_D does not depend on X. There are a number of ways to investigate this assumption. In this section, we divide the sample into groups based on values of X and estimate $MTE(X, U_D)$ separately for each subsample.

We first partition the sample into two subsamples on the basis of gender. For each subsample, we estimate $MTE(X, U_D)$ semiparametrically as in the main text, conditioning on all the variables in X apart from gender. This procedure generates two MTE functions, one for boys and another for girls, that we present side-by-side in Figure B-1. The figure displays the average difference in math scores between Telesecondary and traditional schooling on the vertical axis and the latent variable denoting resistance to treatment, U_D, on the horizontal axis. Both MTEs are mostly upward-sloping, but there is greater variability among boys than girls. For boys with the lowest values of U_D as well as the very highest values, the MTE has a much steeper slope than for girls. The MTE is not precisely estimated for the highest values of U_D, so it is difficult to say whether this difference is significant, although it does seem to be significant for lower values of U_D. This suggests that the pattern of reverse selection is driven much more by heterogeneity in schooling outcomes among boys than among girls.

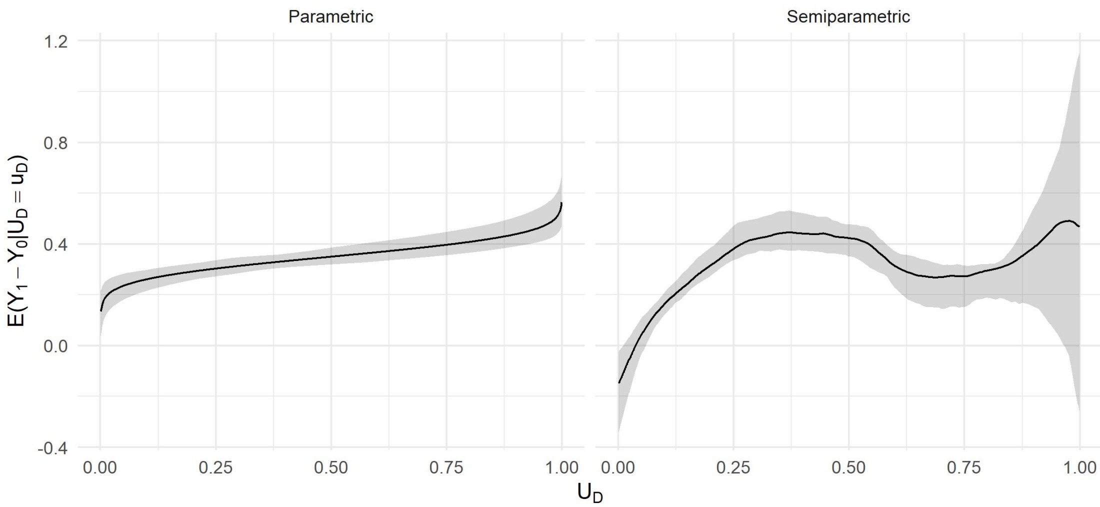
We are also considering partitioning the sample by whether students had below or above median test scores prior to attending secondary school as well as for rural vs nonrural residence.

Figure B-1: Gender-Specific MTEs



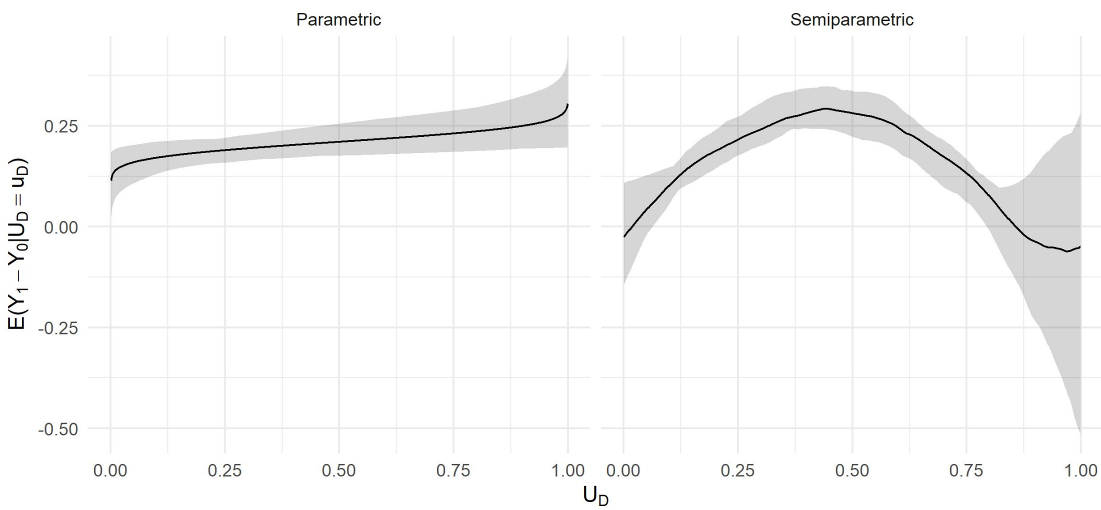
The dependent variable in the outcome equation is the score on the seventh grade nationally-standardized (ENLACE) math exam. The outcome variable has been standardized and is measured in standard deviations from the mean. The outcome equations include controls for sixth grade Math and Spanish scores, the age, 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, dummies for rural residence and residence in a Northern state, and the distance actually traveled to secondary school. 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 sample is divided into subsamples by gender before estimating the MTE using using Local Quadratic Regression and an Epanechnikov kernel with a bandwidth of 0.28, the same as in the figures in the main text. Both MTEs are evaluated at the mean value of the covariates, $X = \overline{x}$. Confidence intervals are computed from boostrapping using 50 draws.

Figure B-2: Controlling for Distance Traveled: 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, dummies for rural residence and residence in a Northern state, and the distance actually traveled to secondary school. 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 B-3: Marginal Treatment Effect Controlling for Distance Traveled: 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, dummies for rural residence and residence in a Northern state, and the distance actually traveled to secondary school. 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.

Table B-1: Estimated Treatment Effects Controlling for Distance Traveled: Math

	Parametric		**Semiparametric**	
	Estimate	Standard Error	Estimate	Standard Error
Average Treatment Effect	0.35	(0.0191)	0.324	(0.0298)
Treatment on the Treated	0.304	(0.0177)	0.266	(0.0206)
Treatment on the Untreated	0.358	(0.02)	0.338	(0.0336)

The table displays three treatment parameters corresponding to the effect of Telesecondary attendance on seventh grade math scores, measured in standard deviations. It differs from Table 19 in the main text in that this specification conditions on the distance actually traveled to secondary school in the outcome equations. The three treatment parameters are obtained by integrating the MTE with respect to the densities displayed in Figure 16. 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 B-2: Estimated Treatment Effects Controlling for Distance Traveled: Spanish

	Parametric		**Semiparametric**	
	Estimate	Standard Error	Estimate	Standard Error
Average Treatment Effect	0.21	(0.0222)	0.152	(0.0262)
Treatment on the Treated	0.19	(0.0191)	0.137	(0.0192)
Treatment on the Untreated	0.214	(0.0233)	0.156	(0.0302)

The table displays three treatment parameters corresponding to the effect of Telesecondary attendance on seventh grade Spanish scores, measured in standard deviations. It differs from Table 20 in the main text in that this specification conditions on the distance actually traveled to secondary school in the outcome equations. The three treatment parameters are obtained by integrating the MTE with respect to the densities displayed in Figure 16. 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 B-3: Estimated Treatment Effects Omitting Mexico City: Math

	Parametric		**Semiparametric**	
	Estimate	Standard Error	Estimate	Standard Error
Average Treatment Effect	0.362	(0.0173)	0.328	(0.0209)
Treatment on the Treated	0.308	(0.0146)	0.283	(0.0193)
Treatment on the Untreated	0.377	(0.0185)	0.34	(0.0248)

The table displays three treatment parameters corresponding to the effect of Telesecondary attendance on seventh grade math scores, measured in standard deviations. It differs from Table 19 in the main text in that this specification omits students who attend secondary school in Mexico City. The three treatment parameters are obtained by integrating the MTE with respect to the densities displayed in Figure 16. 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 B-4: Estimated Treatment Effects omitting Mexico City: Spanish

	Parametric		**Semiparametric**	
	Estimate	Standard Error	Estimate	Standard Error
Average Treatment Effect	0.198	(0.0176)	0.225	(0.0269)
Treatment on the Treated	0.177	(0.0161)	0.176	(0.0215)
Treatment on the Untreated	0.203	(0.0186)	0.238	(0.031)

The table displays three treatment parameters corresponding to the effect of Telesecondary attendance on seventh grade Spanish scores, measured in standard deviations. It differs from Table 19 in the main text in that this specification omits students who attend secondary school in Mexico City. The three treatment parameters are obtained by integrating the MTE with respect to the densities displayed in Figure 16. The simulation method of Carneiro, Lokshin, and Umapathi (2017) is used to integrate the semiparametric MTE. Standard errors are obtained through 50 bootstrap replications.

BIBLIOGRAPHY

I. Acevedo, A. Ortega, and M. Székely. Rendimiento Escolar y Transiciones Laborales con Transferencias Condicionadas en México. *Working Paper*, 2019.

R. Ainsworth, R. Dehejia, C. Pop-Eleches, and M. Urquiola. Information, Preferences, and Household Demand for School Value Added. *National Bureau of Economic Research*, 2020.

O. P. Attanasio, C. Meghir, and A. Santiago. Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progresa. *The Review of Economic Studies*, 79(1):37–66, 2011.

N. Bau. Estimating an Equilibrium Model of Horizontal Competition in Education. *Working Paper*, 2019.

S. A. Beg, A. M. Lucas, W. Halim, and U. Saif. Beyond the Basics: Improving Post-primary Content Delivery through Classroom Technology. *National Bureau of Economic Research*, 2019.

J. R. Behrman, P. Sengupta, and P. Todd. Progressing Through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico. *Economic Development and Cultural Change*, 54(1):237–275, 2005.

J. R. Behrman, S. W. Parker, and P. Todd. Impacts of Prospera on Enrollment, School Trajectories and Learning. *Working Paper*, 2020.

N. Bianchi, Y. Lu, and H. Song. The Effect of Computer-Assisted Learning on Students' Long-Term Development. *National Bureau of Economic Research*, 2020.

A. Björklund and R. Moffitt. The Estimation of Wage Gains and Welfare Gains in Self-Selection Models. *The Review of Economics and Statistics*, pages 42–49, 1987.

F. Bourguignon, F. H. Ferreira, and P. G. Leite. Conditional Cash Transfers, Schooling, and Child labor: Micro-Simulating Brazil's Bolsa Escola Program. *The World Bank Economic Review*, 17(2):229–254, 2003.

C. N. Brinch, M. Mogstad, and M. Wiswall. Beyond LATE with a Discrete Instrument. *Journal of Political Economy*, 125(4):985–1039, 2017.

G. Bulman and R. W. Fairlie. Technology and Education: Computers, Software, and the Internet. In *Handbook of the Economics of Education*, volume 5, pages 239–280. Elsevier, 2016.

F. Buscha, A. Maurel, L. Page, and S. Speckesser. The Effect of Employment While in High School on Educational Attainment: A Conditional Difference-in-differences Approach. *Oxford Bulletin of Economics and Statistics*, 74(3):380–396, 2012.

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

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

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

P. Carneiro, J. J. Heckman, and E. J. Vytlacil. Estimating Marginal Returns to Education. *American Economic Review*, 101(6):2754–2781, 2011.

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

T. Cornelissen, C. Dustmann, A. Raute, and U. Schönberg. Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance. *Journal of Political Economy*, 126(6):2356–2409, 2018.

J. Currie and E. Moretti. Mother's Education and the Intergenerational Transmission of Human Vapital: Evidence from College Openings. *The Quarterly Journal of Economics*, 118(4):1495–1532, 2003.

J. A. Dubin and D. L. McFadden. An Econometric Analysis of Residential Electric Appliance Holdings and Consumption. *Econometrica: Journal of the Econometric Society*, pages 345–362, 1984.

P. Dubois, A. De Janvry, and E. Sadoulet. Effects on School Enrollment and Performance of a Conditional Cash Transfer Program in Mexico. *Journal of Labor Economics*, 30(3):555–589, 2012.

Z. Eckstein and K. I. Wolpin. Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities. *Econometrica*, 67(6):1295–1339, 1999.

D. Epple, A. Jha, and H. Sieg. The superintendent's Dilemma: Managing School District Capacity as Parents Vote with their Feet. *Quantitative Economics*, 9(1): 483–520, 2018.

M. Escueta, V. Quan, A. J. Nickow, and P. Oreopoulos. Education Technology: An Evidence-based Review. *National Bureau of Economic Research*, 2017.

R. Q. Estrada. Telesecundaria: Los Estudiantes y los Sentidos Que Atribuyen a Algunos Elementos del Modelo Pedagógico. *Revista Mexicana de Investigación Educativa*, 8(17), 2003.

R. Fabregas. Broadcasting Education: The Long-term Effects of Mexico's Telesecundarias. *Working Paper*, 2020.

J. Fan and I. Gijbels. *Local Polynomial Modelling and its Applications: Monographs on Statistics and Applied Probability*, volume 66. CRC Press, 1996.

M. M. Ferreyra. Estimating the Effects of Private School Vouchers in Multidistrict Economies. *American Economic Review*, 97(3):789–817, 2007.

M. A. Gottfried. Evaluating the Relationship Between Student Attendance and Achievement in Urban Elementary and Middle Schools: An Instrumental Variables Approach. *American Educational Research Journal*, 47(2):434–465, 2010.

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

J. J. Heckman and E. J. Vytlacil. Local instrumental variables. In K. M. C. Hsiao and J. L. Powell, editors, *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, pages 1–46. Cambridge University Press, New York, 2001a.

J. J. Heckman and E. J. Vytlacil. Policy-relevant Treatment Effects. *American Economic Review*, 91(2):107–111, 2001b.

J. J. Heckman and E. J. Vytlacil. Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, 73(3):669–738, 2005.

J. J. Heckman and E. J. Vytlacil. 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, 2007.

J. J. Heckman, D. Schmierer, and S. Urzua. Testing the Correlated Random Coefficient Model. *Journal of Econometrics*, 158(2):177–203, 2010.

G. W. Imbens and J. D. Angrist. Identification and Estimation of Local Average

Treatment Effects. *Econometrica: Journal of the Econometric Society*, pages 467–475, 1994.

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

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

M. P. Keane, S. Krutikova, and T. Neal. The Impact of Child Work on Cognitive Development: Results from Four Low to Middle Income Countries. *IFS Working Papers*, 2018.

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

T. Le Barbanchon, D. Ubfal, and F. Araya. The Effects of Working while in School: Evidence from Uruguayan Lotteries. *Working Paper*, 2020.

P. Leite, A. Narayan, and E. Skoufias. How do Ex Ante Simulations Compare with Ex Post Evaluations? Evidence from the Impact of Conditional Cash Transfer Programs. *Journal of Poverty Alleviation & International Development*, 6(1), 2015.

F. Martinez Rizo. La Telesecundaria Mexicana. Desarrollo y Problematica Actual. *Coleccion de Cuadernos de Investigacion*, (16), 2005.

L. Navarro-Sola. Secondary School Expansion Through Televised Lessons: The Labor Market Returns of the Mexican Telesecundaria. *Working Paper*, 2019.

C. Neilson. Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students. *Working Paper*, pages 1–62, 2014.

C. Neilson, C. Allende, and F. Gallego. Approximating the Equilibrium Effects of Informed School Choice. *Working Paper*, 2019.

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

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

D. Rodriguez-Segura. Educational Technology in Developing Countries: A Systematic Review. *Working Paper*, 2020.

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

T. P. Schultz. School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program. *Journal of Development Economics*, 74(1):199–250, 2004.

R. Stinebrickner and T. R. Stinebrickner. Working During School and Academic Performance. *Journal of Labor Economics*, 21(2):473–491, 2003.

R. Stinebrickner and T. R. Stinebrickner. The Causal Effect of Studying on Academic Performance. *The BE Journal of Economic Analysis & Policy*, 8(1), 2008.

P. Todd and K. Wolpin. 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, 2006.

P. Todd and K. I. Wolpin. Accounting for Mathematics Performance of High School Students in Mexico: Estimating a Coordination Game in the Classroom. *Journal of Political Economy*, 126(6):2608–2650, 2018.

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

Y.-T. Yap, G. Sedlacek, and P. F. Orazem. Limiting Child Labor through Behavior-based Income Transfers: An Experimental Evaluation of the PETI Program in Rural Brazil. In *Child Labor and Education in Latin America*, pages 147–165. Springer, 2009.

K.-H. Yuan, Y. Cheng, and J. Patton. Information Matrices and Standard Errors for MLEs of Item Parameters in IRT. *Psychometrika*, 79(2):232–254, 2014.