ABSTRACT

Title of Dissertation: ESSAYS ON TREATMENT EFFECTS FROM MULTIPLE UNORDERED CHOICES

Camila Galindo
Doctor of Philosophy, 2021

Dissertation Directed by: Professor Sergio Urzúa
Department of Economics

I study some of the methodological and empirical challenges associated with estimating treatment effects of one option *versus* another, in contexts where agents can choose from many alternatives with no clear rank (*i.e.*, one option is no better than the other, for everyone). In particular, I focus on educational decisions throughout the life-cycle, such as parental choice of childcare, students' choice of high school, and college enrollment. First, I present a strategy to overcome a limitation of instrumental variables in these settings. I use this strategy to provide empirical evidence from an early childhood development intervention in Colombia, where parents can choose among different childcare options (*e.g.*, small centers, large centers, or home care). In the third chapter, I focus on the Chilean high school context where students can choose from three types of schools: academic, vocational, or hybrid. I find that, while academic schools seem to improve the student's academic achievement, the effects of hybrid and vocational schools depend on the student's fallback option (*i.e.*, what they would have chosen if their preferred option was

not available). Last, in the Colombian context, jointly with Maria Marta Ferreyra and Sergio Urzúa, I examine the labor market returns to short-cycle degrees *versus* bachelor's degrees and *versus* obtaining a high school diploma.

In Chapter 2, I present an identification strategy that exploits the joint effect of discrete and continuous instruments on the probability of choosing an option. When the identifying variation stems from multiple instruments, agents can switch into different options and from many initial states. I discuss how to use conditional choice rules to estimate the shares of agents switching at well-defined margins of choice and their treatment effects.

I develop an identification strategy consistent with this framework and apply it to assess the impact of childcare choice in Colombia on children's development. Parents can choose between home care and public care at small or large centers. I exploit two sources of exogenous variation: a lottery that provides information and encourages winning parents to switch to large centers, jointly with the geographical distance between the child's home to the nearest center. Parental responses to the experimental variation can differ depending on the distance to the center. This feature uncovers heterogeneous responses along two margins of choice: small *versus* large centers, and small centers *versus* home care. Previous methods would attribute all the experimental variation to the small *versus* large centers margin. I find that, on average, 15-18% of parents are induced to switch from small towards large centers as the lottery outcome and proximity vary. My results suggest that, on average, switching towards large centers might benefit some children who live far from large centers but have more educated mothers.

In Chapter 3, I estimate the effects of different high school types on educational achievement, such as high school completion and higher education enrollment. I find evidence that suggests that attending a vocational high school does not have a differential effect on the probability of enrolling in a vocational college. Moreover, while hybrid schools seem to foster student enrollment in bachelor's programs, this effect largely depends on the student's fallback option. In particular, there is no evidence of improvements in educational achievement among students who would have chosen academic schools instead of hybrid schools.

In Chapter 4, with Maria Marta Ferreyra and Sergio Urzúa, we provide evidence of diversion and expansion effects of changes in the local supply of Short-Cycle degrees, in the context of higher education for Colombia. Our results suggest that most students would divert from bachelor's- and into short-cycle- degrees as the local supply of short-cycle degrees changes. For these students we find significant gains, particularly among women, in terms of participation in the formal labor market and years of experience.

ESSAYS ON TREATMENT EFFECTS FROM MULTIPLE UNORDERED CHOICES

by

Camila Galindo

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor Sergio Urzúa, Chair/Advisor
Professor Guido Kuersteiner
Professor Nolan Pope
Professor Judith Hellerstein
Professor Erich Battistin

Dedication

Para Consuelo y José, por apoyarme siempre.

Acknowledgments

I want to thank Sergio Urzúa for his dedication, support, and commitment to my progress and success throughout the Ph.D. His teachings and academic advice have been inspirational; with his guidance, I have pursued challenging projects in the topics closer to the core of my research interests. I am also grateful for his kindness and patience, for listening and motivating me. His confidence in my potential, more so when I doubted myself, helped me keep learning and finding ways to improve my projects, even during a global pandemic.

I am also grateful to Guido Kuersteiner for his advice and support from early on. His encouragement increased my confidence to explore applied econometrics questions, which has been one of my favorite research fields from the start. He has also taught me that through consistent work and effort, everything starts to fall into place. These words helped me get through working on my job market paper and this dissertation, more so during the covid pandemic. I am also thankful to Nolan Pope, from whom I learned both as a student and as RA. I am incredibly grateful for his feedback and for inspiring me to study research questions rigorously and wholeheartedly.

I also thank Erich Battistin for agreeing to be part of my dissertation committee. Judith Hellerstein sparked my interest in labor economics during field day and throughout her course during my second year. I am grateful for her questions and comments during my presentations, which always contributed to gaining a deeper understanding of my projects' economic content. John Shea has also been supportive from the start. A big thank you goes to Vickie, Mark, and Terry for all their help these years. And to Vickie for her messages of support and for finding a way to keep us connected while apart throughout this year.

I made it all the way from Colombia, through DC, and to Maryland until the completion of my Ph.D. with the support from many. Raquel Bernal has been a mentor and an inspiration since my undergrad. I am thankful for fellow students, from whom I have learned a great deal, for their friendship beyond the Ph.D.: Gonzalo Garcia, Macarena Kutscher, Nathalie Gonzalez, Fernando Saltiel, Cristián Sánchez, Mateo Uribe-Castro, and Shanjukta Nath. And to lifelong friends for being there, always: Valentina Duque, Amalia Rodriguez, Lina Diaz, Cesar Mantilla, Cristina Martinez, Christian Salas, Rafael Hernandez, Paula Gonzalez, Karen Ortiz, Maria Adelaida Martinez, Cynthia van der Werf, and Diana Martinez. And to Matteo, my biggest thank you.

Most of all, I am grateful to my mother and my father. This has been a challenging and arduous process. We accomplished it together. It is and has always been because of both of you. Gracias.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

In this dissertation, I study the identification and estimation of treatment effects with instrumental variables in contexts where agents can choose from multiple options. Recent methodological literature (Caetano and Escanciano, 2020; Feng, 2020; Heckman and Pinto, 2018; Lee and Salanié, 2018, 2020; Xie, 2020) provides novel and richer frameworks to understand the gains, or losses, from different combinations of choices in settings with many endogenous alternatives. There is also increasing interest in this type of settings in the economics of education (Dean and Jayachandran, 2019; Kirkeboen, Leuven, and Mogstad, 2016; Kline and Walters, 2016; Mountjoy, 2019; Pinto, 2019) with both empirical and methodological contributions. Yet, several identification and implementation challenges remain. Given these challenges, the empirical evidence that considers the role of multiple alternatives for different educational contexts remains scant.

The goal of this dissertation is twofold. First, I present a methodological contribution to the literature on the identification and estimation of the effect of choosing an option in contexts where there are multiple alternatives available. Second, I provide empirical evidence of heterogeneous effects for three educational settings: parental choice of childcare, high school choice, and college enrollment. To start, Chapter 2 proposes an identification strategy to overcome a limitation of standard Instrumental Variables (IVs). Namely, when multiple options are available IV estimates the effect of one option *versus* the next-best alternative (*i.e.*, a mixture of the remaining options available). The main contribution of this chapter is to present a strategy to uncover the effect of choosing one option *versus* another, or pairwise choice combinations, on an outcome of interest. The novelty of this strategy is that I exploit the combined effect of discrete and continuous instrumental variables on the probability of choosing each option. In the second part of Chapter 2, I employ this identification strategy to assess the effects of an early childhood development intervention in Colombia.

Empirically, I study educational investments throughout the life cycle, such as parental choice of childcare (in Chapter 2), students' choice of high school, and college enrollment. Chapter 3 focuses on the context of high school choice in Chile, where students can choose from three alternatives (Academic, Vocational, or Hybrid). The role of Hybrid schools has remained largely unexplored; I find that students shifting from Vocational and into Hybrid schools might benefit from choosing the latter in terms of their academic achievement and performance. Last, Chapter 4 shows how changes in the local supply of short-cycle degrees in Colombia seem to divert students from bachelor's degrees but have virtually no effect on inducing students to enter the higher education system. All in all, using Instrumental Variables (IVs), I provide evidence on how agents with different fallback options can reap differential gains from their educational choices.

A critical aspect of the IV approach is that it estimates local effects for agents who change their behavior as the IV changes (*i.e.*, compliers). With multiple unordered choices, compliers are heterogeneous, and agents can switch along many margins (Heckman, Urzua, and Vytlacil, 2006; Heckman and Urzúa, 2010; Kirkeboen et al., 2016; Mountjoy, 2019). For instance, consider the case of parental choice of childcare and its effect on children's development among low-income families in Colombia. In 2011, some parents won a random offer to transfer from a small childcare center to new and large integral centers. Uncovering the effect on children's development of choosing these large centers largely depends on parents' fallback option (*e.g.*, alternative care centers or home care). For example, if a large childcare center opens close to home, parents could switch between home and large centers or small and large centers. Importantly, parents responding along each margin of choice can be systematically different, and their children could derive gains or losses accordingly. Thus, identifying the effects of parental choice of childcare is an empirical challenge with potential policy implications.

Chapter 2 addresses some of the empirical challenges associated with estimating treatment effects when agents can choose from multiple alternatives (such as childcare choice). My main contribution is to present an empirical strategy that exploits the combined effect of discrete and continuous instruments on the probability of choosing an option. These combined effects have been recognized in the binary choice case (Mogstad, Torgovitsky, and Walters, 2020a,b). In turn, current methods for multiple unordered choices implicitly assume that the behavior of compliers (or responses) to one instrument is the same across the distribution of other instruments.

Instead, I allow for the response to the variation in one instrument (for example, an offer of a slot at a center) to differ depending on other instruments (for example, proximity to the center). To do so, I employ a latent utility framework and model responses to the instruments through their effect on each option's costs. With assumptions motivated by economic theory (*i.e.*, convexity of cost functions), I define *conditional vectors* consisting of combinations of potential choices that satisfy monotonicity (Imbens and Angrist, 1994) locally. I use *conditional vectors*, and the methodological approach of Heckman and Pinto (2018) to estimate conditional local average treatment effects (LATEs).

The researcher can apply the identification strategy of Chapter 2 to many contexts, such as experimental settings with noncompliance. When additional sources of exogenous variation are available, such as costs, the researcher can exploit the combined effects of random variation and costs to identify LATEs for different margins of choice. For example, suppose the response to the experimental variation along a specific margin of choice is only prevalent when costs are lower. In that case, the researcher can directly identify the conditional LATE of these options. Otherwise, I combine the joint effects of multiple IVs with an assumption of homogeneity (Hull, 2018; Kline and Walters, 2016; Lee and Salanié, 2020).[1] I extend the homogeneity assumption to the case of conditional estimation and propose a test of differences in the average of baseline variables across margins of choice to determine its feasibility.

[1]Homogeneity states that when agents would react to the IVs by switching along multiple margins but towards the same option, the average counterfactual of that option would be the same across all initial states.

The literature on multivalued treatment effects offers alternative identification strategies. Whereas I focus on contexts where the available IVs stem from different sources (*e.g.*, random assignment combined with distance to an option), Mountjoy (2019) presents important identification results when the IVs follow the same logic. The author studies the case of community colleges in the U.S. and uses variation in proximity to two- and four-year colleges. The main assumption is comparable compliers, for those students who would change their choice between two- and four-year college as the distance to each option changes. Both Chapter 2 and Mountjoy (2019) rely on variation from multiple IVs. In turn, Hull (2018); Kline and Walters (2016); Lee and Salanié (2020) deal with contexts where the available instruments are not enough to identify well-defined effects. For instance, Hull (2018) employs a homogeneity assumption that requires exploiting variation in one instrument and its interaction with background variables.[2] Overall, whether Chapter 2 or another strategy is more suitable for identifying LATEs depends on the context and the nature of the IVs available.

The main advantage of the identification strategy that uses *conditional vectors* is to separately identify local effects for pairwise combinations of the choices (rather than *versus* the next-best), but there are some limitations. First, with *conditional vectors* I can identify conditional Local Average Treatment Effects (LATEs) that can vary across the distribution of the IVs. While it is possible to recover unconditional LATEs for pairwise comparisons of different options, the strategy I present might

[2] It is worth noting recent empirical work using Machine Learning (Dean and Jayachandran, 2019; Rodriguez and Saltiel, 2020) to identify the fallback option of the agents, as well as the literature on identification and estimation of effects in settings where individual rankings of the options are available (Kirkeboen et al., 2016).

exacerbate the consequences of weak instruments by exploiting local (conditional) variation. Second, defining the *conditional vectors* over a continuous instrument requires some level of discretization, such that the researcher can construct bins at which the behavior of the agents would change. On the other hand, without enough support in the conditioning IV, it would not be feasible to identify effects separately for different combinations of the options. For instance, using a binary IV and conditioning on another binary IV might not be enough to define potential responses that change across the IV distribution.

I apply the identification and estimation strategy of Chapter 2 to estimate the impact of childcare choice in Colombia on children's development. Parents can choose between home care, small centers, or large centers. My sample consists of children who were initially at small centers, before large centers were available. I exploit experimental variation in a lottery that provides information about large centers jointly with the geographic distance between the child's home and the nearest large center. Previous methods would attribute all the experimental variation to the small versus large centers margin. In turn, I find evidence that parents also respond by switching along the small center *versus* home care margin when large centers are relatively more expensive. Among some of the mechanisms that might explain why some parents choose home care when winning the lottery is a disruption of the public childcare system due to the closure of some small centers.

By disentangling effects at two margins (small *versus* large centers, and small centers *versus* home care), I show that some children might benefit from choosing large centers in terms of their cognitive development. In contrast, standard IV methods would have uncovered an overall negative effect of large centers on children's cognitive development. Importantly, the LATEs I estimate are further restricted to children already in public care (since all children in my sample were previously in small centers) and cannot provide information for children who would enter the system through large centers. In addition, the estimation largely depends on defining thresholds of the distance to large centers where the parental response to the lottery outcome would change. Throughout my main specification, I use an interval (rather than a point) threshold between 0.75-0.85km to estimate the LATEs. I assess the sensitivity of the results by using the extreme points of this interval as thresholds. The results are overall consistent in terms of significance and direction, with some differences in magnitude only for the LATEs that are not statistically significant. While encouraging, the definition of the thresholds requires further work such that the researcher can choose them in a disciplined manner rather than from exploratory results.

In Chapter 3, I study the effect of high school-type decisions on short-term academic achievement and college enrollment after graduation in the Chilean context. I analyze a setting where students can select between three schools (academic, vocational, or hybrid). Hybrid schools, plus features of the high school system, make Chile an interesting context to study the effects of vocational *versus* traditional or general education. In contrast to purely Academic or Vocational schools, Hybrid schools can offer both tracks, which allows for higher heterogeneity in class composition. In addition, attending a Hybrid school might allow students to postpone track choice compared to Academic or Vocational schools. While recent literature has found labor market gains from increasing the general curricula in vocational tracks and delaying the age of tracking, among other features of several reforms, (Bertrand, Mogstad, and Mountjoy, 2021; Canaan, 2020; Ollikainen, 2021; Zilic, 2018), the evidence on educational trajectories remains mixed.

In this Chapter, I also address identification and estimation limitations that have received less attention in high school choice contexts. When choosing a type of school, students face many alternatives with no inherent ranking, and they can have different fallback options. To address these challenges, I exploit (i) data from Chile that links student educational trajectories, high school type decisions, college enrollment and completion, and (ii) local variation in average distance to different high schools. I find that the variation to estimate effects of school-type choice using changes in travel time is about 8.4% to 14%. This fraction corresponds to students who would change their high school choice due to changes in travel time to each school type. On average, they are more disadvantaged than students who are not responsive to variation in travel time to a school-type in terms of household income, maternal level of education, and baseline academic performance.

Among some of my findings, I observe that attending a Vocational school seems to increase the probability of completing high school but has no effect on the likelihood of enrolling in a vocational college. These effects are in line with previous findings, such that students who attend vocational schools are less likely to keep studying (Bassi and Urzua, 2010; Brunello and Checchi, 2007; Brunner, Dougherty, and Ross, 2019). In contrast, I uncover positive LATEs on the probability of enrolling in an academic college for students in Hybrid schools. This effect is likely to stem from students who would have enrolled in Vocational schools otherwise (if Hybrid high schools were not available) rather than those who would have enrolled in an Academic school.

Last, in Chapter 4, jointly with Maria Marta Ferreyra and Sergio Urzúa, we extend the analysis of multiple unordered choices to the context of higher education enrollment in Colombia. Following the literature on two-year college enrollment in the U.S. (Acton, 2020; Denning, 2017; Leigh and Gill, 2003; Mountjoy, 2019; Rouse, 1995), we ask whether students in the Colombian context have also experienced a democratization effect, rather than a diversion from bachelor's degrees, of increasing access to short-cycle programs. We use an Instrumental Variables strategy and exploit variation at the municipal level in the availability of higher education institutions that only offer short-cycle degrees.

Our results differ from the evidence for the U.S. We observe that changes in the local availability of short-cycle degrees increase the probability of choosing a short-cycle degree at the expense of enrollment in bachelor's degrees, with no effect on students who do not enter the higher education system. In contrast, (Acton, 2020; Denning, 2017; Leigh and Gill, 2003; Mountjoy, 2019) have found that most of the response to increasing access to two-year college has shifted students to enter the higher education system, rather than diverting them from four-year college enrollment. Similarly, these authors have uncovered gains of two-year college among students who would not have enrolled in higher education and adverse effects among those who divert from bachelor's degrees (Mountjoy, 2019). Differences in the role of short-cycle degrees in Colombia as a way to acquire skills and join the labor market fast *versus* the role of two-year colleges in the U.S. as a pathway towards four-year enrollment might explain some of the differences in our findings.

The methodological and empirical results of this dissertation provide some lessons for the identification and estimation of LATEs. First, I exploit the variation in discrete and continuous IVs throughout Chapter 2 to Chapter 4. While continuous IVs are likely to provide enough support to identify a richer set of LATEs (*i.e.*, one option *versus* another, rather than *versus* the next-best), they also require analyzing the specification of the functional form. In Chapter 3, in the main specification, I assume that travel time to each type of high school enters the choice function linearly. Next, I present robustness checks to the functional form of travel time and show how the LATEs of vocational schools *versus* the next-best are the most sensitive to including quadratic and cubic terms. As a result, with a more flexible functional form, I find that the linear effect of vocational schools on fostering high school completion largely disappears.

In contrast, Chapter 4 uses variation in a discrete IV, limiting the scope of the strategy in Chapter 2 and the set of LATEs that can be identified. In this last Chapter, I observe that the response to the discrete IV (which is a binary variable of supply of higher education institutions offering short-cycle programs at the municipal level) is very small (at most three percent). This low response to the variation in the local supply of higher education measured with a discrete variable also highlights the limitations of IVs with small support in contexts with multiple options.

Throughout the dissertation, I employ the variation from a cost-shifter (distance to an option), which is frequently used in the literature of the economics of education but has some drawbacks. From the work of Card (1995) on the returns to education to more recent work by Mountjoy (2019) for the U.S. context, distance or proximity to an option is assumed to affect outcomes only through its effect on choices. There are two important critiques to this source of variation. First, distance to an option (*e.g.*, childcare centers) can be correlated with local factors that in turn explain outcomes (*e.g.*, children's development). Violations of the former (exclusion) restriction threaten the identification of causal effects. With this caveat in mind, I control for local conditions with fixed effects. Also, I present a robustness check in Chapter 3 by including measures of proximity to amenities (such as hospitals, police stations, or libraries) as controls. I observe that the LATEs of Academic high schools are relatively the same in size and significance, even after controlling for local amenities. On the other hand, the effects for Vocational and Hybrid schools show some difference in size, but not in their significance or direction. Hence, the LATEs using proximity measures might be miss-estimating the effects of the different options and warrant caution in their interpretation.

Second, agents who respond to the variation in the distance are likely to differ from the population on average (for instance, they can be more likely to be children from the most disadvantaged backgrounds). I present estimates of the average of baseline variables for different complier groups across all the empirical analyses to address this concern. With this information, I show in Chapter 4 that, for instance, women who would divert from bachelor's and into short-cycle degrees because of changes in the local supply of higher education are older and more likely to belong to low-income households.

Last, an important takeaway from Chapters 2 to 4 is that estimating effects for different fallback options matters, regardless of the context. For example, chapter 2 shows that by disentangling the next-best LATEs into margin-specific LATEs, I can recover positive effects of choosing large centers for some children. Instead, a standard IV approach would have found overall adverse effects. Similarly, Chapter 3 shows that Hybrid high schools might improve the likelihood of enrolling in a bachelor's program for some students. I find that most of the improvement likely stems from students who would divert from Vocational high schools and into Hybrid high schools as travel time to the latter decreases. Previous empirical work focuses on the track (binary) choice or disregarded Hybrid high schools altogether (Larrañaga, Cabezas, and Dussaillant, 2013). Hence, by including the third alternative in my analysis, I uncover important results which show the role of an often overlooked option in fostering students to keep investing in their education.

Chapter 2: Empirical Challenges of Multivalued Treatment Effects

2.1 Introduction

In contexts where agents can choose from many endogenous options (such as career paths, migration decisions, or parental choice of childcare) the researcher would like to estimate average treatment effects of different choices (on average, do large childcare centers improve cognitive development relative to small care centers? or, relative to home care?). While instrumental variables emerge as an alternative to deal with multiple endogenous choices (Caetano and Escanciano, 2020; Feng, 2020; Heckman and Pinto, 2018; Lee and Salanié, 2018, 2020; Xie, 2020), with this strategy the researcher often uncovers the effect of one option *versus* the next-best (*e.g.*, the effect of large centers compared to a mixture of home care and small centers) (Heckman et al., 2006; Kirkeboen et al., 2016). These effects might be uninformative and mask heterogeneous effects for agents with different preferences, (*e.g.*, large centers might be beneficial for children's development when compared to home care, but have detrimental effects when compared to small centers).

This chapter studies a general case of multivalued treatment effects where the identifying variation stems from discrete and continuous instruments (IVs). I exploit the joint effect of multiple instruments on the probability of choosing an option to identify the effect of one option *versus* another, rather than the effect of one option *versus* the next-best. If multiple instruments can affect the same margins of choice, how do agents sort into and out of alternatives as the instruments change? For instance, suppose that the researcher has access to experimental variation that encourages agents to choose option 1 instead of 2 or 3, but only if they face lower costs. In turn, when costs are higher, the responses between options 1 and 2 disappear. In this scenario, the experimental variation alone might not be enough to uncover the full set of responses. Therefore, within my framework, I allow for the response to the variation in one instrument (e.g., an offer of a slot at a childcare center) to differ depending on other instruments (e.g., proximity to the center) that affect the same choice. In contrast, while joint responses to multiple instruments have been recognized in the literature of binary treatments (Mogstad et al., 2020a,b), current identification strategies for multiple options implicitly assume that the behavior of compliers (or responses) to one instrument is the same across the distribution of other IVs.[1]

I employ an empirical strategy consistent with a framework of multivalued treatment effects. The instruments affect choices through their effect on potential costs.[2] Following long-standing literature of shape restrictions that stem from economic theory, [3] I assume that the cost function determining individual's decisions

[1]For the binary treatment case, Mogstad et al. (2020a,b) show that the standard monotonicity assumption in Imbens and Angrist (1994) imposes homogeneity in the responses to multiple instruments.

[2]I model agents' decisions with a latent utility model with additive separability between the unobserved preference heterogeneity and the cost function. This is standard in the unobserved heterogeneity literature (Vytlacil, 2002). Recently, Heckman and Pinto (2018) extended this notion for unordered choices; Lee and Salanié (2020) model changes in relative mean utility through changes in relative costs.

[3]See, for example, a nonparametric model of multinomial choice Matzkin (1991), with restric-

is weakly convex. This is a necessary condition for a concave objective function. In addition, the role of weak convexity in identification is twofold. First, without any restrictions on the shape of the function, multiple instruments can affect the costs in different directions. Convexity provides a framework to restrict the set of potential choices such that agents who are induced to change their behavior do so in the same direction (towards the cheaper option). In contrast, without convexity, there could be multiple, not consecutive, intervals of the distribution of the instrument where one option can switch from having the lower to the highest cost. Hence, agents would change into, or away, from the same option at multiple realizations of the instruments. Second, convexity also allows the researcher to model a richer set of cost functions where multiple instruments affect the same choices. Thus, instruments can enter the cost function nonlinearly.

I use the methodological approach of Heckman and Pinto (2018) in which rules of behavior from revealed preference analysis define vectors used to identify the prevalence of compliers (i.e., agents who change their behavior as the IV changes) and their counterfactuals. I extend this notion to the case of *conditional vectors* defined by an instrument that can change across the support of other IVs. Namely, *conditional vectors* consist of combinations of potential choices that differ depending on the variation of other instruments. These combinations are consistent with the restrictions of convexity on the cost function.

I show that the *conditional vectors* satisfy a weaker version of monotonicity

tions of concavity and monotonicity to achieve identification. Recent work by Freyberger and Horowitz (2015) impose shape restrictions such as convexity and monotonicity on the functional form of the outcome equation. For a review of shape restrictions in applied work see Chetverikov, Santos, and Shaikh (2018).

than in the literature of Imbens and Angrist (1994). To be specific, they satisfy conditional/partial monotonicity for multiple treatments (Mountjoy, 2019).[4] This assumption requires that agents who react to the variation in one instrument do so in the same direction (towards or away from the same choices), conditional on the variation in other IVs. Although this assumption has been previously employed, I further exploit a well-known, but frequently overlooked, property of monotonicity. It requires that changes in an instrument from z to z' induce all agents in the same direction; meanwhile, changes from z' to z'' do not have to induce agents in the same direction as changes in z to z'. For instance, if agents would switch towards option 1, and away of option 2, when the instrument goes from z to z' it is not required that they do so when the instrument changes from z' to z''. With *conditional vectors*, I identify a richer set of compliers that result from allowing the responses at z *versus* z' and z' *versus* z'' to differ.

I discuss the role of *conditional vectors* to estimate shares of compliers at different margins of choice and their treatment effects. I show how to estimate complier shares and their counterfactuals as response-weighted averages of conditional propensity scores and choice-outcomes interactions. Under the assumption of strict convexity of costs, the method can be implemented with a nonparametric approach,[5]

[4] Das (2005) imposes this assumption conditional on covariates for nonlinear IVs; Heckman et al. (2006) impose conditional monotonicity on a second instrument, for the binary treatment case. More recently, Mogstad et al. (2020a,b) and Goff (2020) present identification results for the binary treatment case.

[5] Nonparametric estimation of IV strategies such as 2SLS suffer from the ill-posed inverse problem. Nonetheless, the strategy I present here deals with discrete endogenous variables and does not suffer from this problem. The method can be implemented by reduced form estimation of choice-by-outcome interactions, which can be estimated nonparametrically, as well as propensity scores from the first-stage.

without imposing restrictions on the relation between the instruments on the choice equation. This approach fits with nonlinear instrumental variables estimation for discrete endogenous variables presented in Das (2005) and, more generally, in Chen, Chernozhukov, Lee, and Newey (2014); D'Haultfœuille and Février (2015); Imbens and Newey (2009) and Chernozhukov, Imbens, and Newey (2007).[6] Moreover, for a linear cost structure, the method can also be implemented with a linear decision model by incorporating interactions between the instruments. Overall, the treatment effects that can be identified are conditional Local Average Treatment Effects.[7]

The strategy I present can be applied, but is not limited, to experimental designs with noncompliance when costs are also available. Noncompliance is frequent in applied work (e.g., Dean and Jayachandran, 2019; Kline and Walters, 2016; Pinto, 2019), and poses the challenge of identifying counterfactual choices or fallback options that can largely vary across individuals. Moreover, the researcher can often uncover at most local treatment effects of one option *versus* the next-best (i.e., what the agent would have chosen if that option is no longer available). Previous literature has employed assumptions on homogeneity of the counterfactuals to disentangle treatment effects along well-defined margins of choice (Hull, 2018; Kline and Walters, 2016; Lee and Salanié, 2020).[8] I extend this assumption to the case of conditional margins of choice and present a test of homogeneity that takes advantage

[6]Nonetheless, to impose the conditional rules of behavior the researcher needs to restrict the dimension of the nonparametric component since a fully nonparametric approach would threaten the assumption on convexity of costs.

[7]Heckman et al. (2006) discuss the scope of conditional estimation and conditional LATEs for binary treatment. Das (2005) shows how to identify conditional LATEs by conditioning on covariates.

[8]In turn, a different approach to secure separate identification can be found in Pinto (2019) and Mountjoy (2019).

of changes in the prevalence of some complier groups conditional on the variation of other IVs. Using *conditional vectors*, I show how to recover treatment effects at well-defined margins of choice and how to assess whether or not the homogeneity assumption is plausible.

The estimation strategy I present has some limitations. First, IV estimates local effects (LATEs) for agents who would change their choices as the instruments change. With a conditional approach, I can directly identify conditional LATEs and recover the unconditional LATE by integrating over the distribution of the conditioning IV. Although this approach allows the researcher to disentangle effects at different margins of choice, it might do so at the expense of exacerbating the consequences of weak IVs. Second, the definition of *conditional vectors* requires discretizing a continuous IV to define bins across the distribution where the agents' behavior would change. Nonetheless, conditioning on IVs with not enough support (*e.g.*, one binary IV conditional on a binary IV) will not allow the researcher to identify effects for pairwise combinations of the options separately.

I use the identification results to empirically assess the case of childcare choice in Colombia and its impact on children's development. This is a unique setting to analyze IVs in contexts of unordered multivalued treatments. First, there are three childcare alternatives that have no clear rank across dimensions of child development. Parents of children between six months to five years of age can choose between home care, small centers, or large centers. These options differ in terms of infrastructure, caretaker to children ratios, nutritional services, and the classrooms' age composition. Neither small nor large centers fare better across all dimensions that could potentially improve the development of children. For instance, small centers have lower teacher to child ratios compared to large centers. But large centers have age-specific classrooms while small centers group children of all ages in the same room. Second, the childcare system in Colombia underwent a supply expansion, increasing the available options from small centers to both small and large centers. In my sample, I observe children during this expansion and have measures of their characteristics before and after large centers were available. I also use institutional features of the changes in supply to explain how parents sort between alternatives.

In my application, the source of variation stems from discrete and continuous IVs that jointly affect the same margins of choice. Childcare choices are endogenous, with parents self-selecting into their preferred alternative. To estimate treatment effects, I exploit experimental variation jointly with the geographical distance to the nearest center. The former is a lottery with noncompliance that provided information and encouraged winning parents to switch to large centers. Consistent with the general framework of this chapter, I assume that the instruments affect choices through their effect on costs. I impose a shape restriction of convex and increasing costs on the distance to the large center. This restriction translates into potential parental responses to the lottery outcome along the margins of home care *versus* large centers, and small *versus* large centers.

My framework allows the parental response to the experimental variation to change with distance to the nearest large center. Convenience (i.e., low travel time or distance) has been recognized as an important predictor of childcare choice (Attanasio, Maro, and Vera-Hernández, 2013; Bernal and Fernández, 2013; Hojman and López Bóo, 2019). Therefore, despite winning the lottery, some parents might be discouraged from choosing large centers if the distance from home is relatively large. In practice, I observe that some parents without a large center nearby react to the lottery outcome variation by switching towards home care and away from small centers. With these findings, I define conditional rules of parental behavior that vary with distance to the nearest large center. This feature empirically uncovers heterogeneous responses along two margins of choice: small *versus* large centers and small centers *versus* home care. Importantly, previous methods would attribute all the experimental variation to the small *versus* large centers margin. I find that, on average, 15-18% of parents are induced to switch from small towards large centers as the lottery outcome and proximity vary. In turn, about 7% of parents without a large center nearby are induced to switch between small centers and home care due to the lottery outcome variation.

My results suggest that, on average, switching towards large centers might benefit some children who live far from large centers but have more educated mothers. Although there is considerable evidence on the effect of better infrastructure on children's development in developed countries, the evidence in developing contexts remains scant. The paper closer to my analysis is Bernal, Attanasio, Peña, and Vera-Hernández (2019), who estimate the effect of small *versus* large centers for a similar sample using variation in the lottery, and find negative treatment on the treated effects on cognitive development and positive effects on nutrition. The authors present an exhaustive analysis of noncompliance with the lottery and find that one of the principal factors was child's age. In turn, I find large baseline differences in children's age across a richer set of compliers, with those closest to the eligibility limit benefiting the least from the transfer to large centers. I also show that part of the detrimental local effect on children's development of enrolling in a large center is due to a bias term that stems from children who would switch between small centers and into home care due to the variation in the lottery and proximity to the large center.

The application results have some caveats. First, the variation of proximity to the large centers might be correlated with other factors that can affect the development of children. Given the nature of public care, it is plausible that the location of the centers depends on the prevalence of eligible children and, hence, it is not random. While I control for local factors in the main specification, my analysis is limited by a lack of data to determine the characteristics of the population around the large centers. Second, the lottery was assigned at the small centers level, and the workers at winning centers could inform parents about the large centers. Hence, the lottery and the opening of large centers disrupted the supply of public care, and some small centers closed as a result. The latter implies that some parents lost their current care option and might have been discouraged from finding alternative public care. Last, I only observe a subsample of children who were initially (before large centers were available) at small centers. Therefore, I can only recover LATEs specific to children already in the public care system; my results cannot be generalized for children who would enter public care as the supply expands. Despite these limitations, I focus on this context to show the role of *conditional vectors* and why they matter in the estimation of effects in complex settings.

Last, one important component of the estimation of conditional LATEs is the definition of a threshold of the conditioning IV (*e.g*, distance to large centers) at which the *conditional vectors* would change. In the application, rather than defining a point value for the threshold, I use an interval of 0.75-0.85km. I assess the sensitivity of the conditional LATEs to two alternative (point) thresholds: 0.75km, and 0.85km. Overall, the main specification findings seem consistent in direction and significance across the different thresholds.

This chapter is organized as follows. Section 2.2 presents a general framework of multivalued treatment effects and instrumental variables, describes the shape restrictions that secure identification, and the resulting conditional rules of behavior. Here, I present the estimation strategy. Section 2.3 provides background on childcare choices in Colombia, describes the data, and presents descriptive results. It includes a latent utility model of choice of childcare, as well as restrictions on the cost function. Next, I derive conditional rules of parental behavior and discuss their empirical implementation. Section 2.4 discusses the empirical results. Section 2.5 concludes.

2.2 General Framework

Consider a decision maker, i, who can choose between different alternatives. Each of these alternatives are no better than the others *a priori* (i.e., they are unoredered), but agents can have preferences over them. Let Y_i denote an outcome of interest for individual i. Options are represented by d, with $d \in \{1, 2, ..., k\}$, where k is the total number of choices, and in contexts of unordered choice higher values do not represent better options. Vector $\mathbf{Z}_i$ contains L instrumental variables, such that $\mathbf{Z}_i = \{Z_{i,1}, ..., Z_{i,L}\}$ where each $Z_{i,l}$ takes values in $\text{supp}(Z_{i,l})$. I place no restrictions on their support, which can be continuous or discrete. I also place no assumptions on the joint support of the instruments in $\mathbf{Z}$ and denote it as $\mathcal{Z}$. For simplicity, I refer to an element $z \in \mathcal{Z}$ as a combination of $(Z_{i,1}, ..., Z_{i,L})$.

Potential outcomes are denoted by $Y_{di}(z)$, which represent the outcome agent i would derive from choosing option d at value $z \in \mathcal{Z}$. Define $D_i(z)$ as potential choices at value z of the instruments, i.e. the choice agents would select when faced with value z. As in Imbens and Angrist (1994), the instruments in $\mathbf{Z}$ satisfy:

A1 *(Independence)* $(\{Y_d\}_{d \in 1,..,k}, \{D(z)\}_{z \in \mathcal{Z}}) \perp\!\!\!\perp \mathbf{Z} | \mathbf{X}$,

where $\mathbf{X}$ denotes a vector of covariates. Assume that agents preferences can be represented by a discrete choice problem, where $U_{id}(z)$ represents latent utility for agent i at a combination of values of the instruments $z \in \mathcal{Z}$. Utility depends on unobserved preferences and costs: $U_{id}(z) = \mu_{id} - V_{id}(z)$, where μ_{id} represents unobserved preference heterogeneity for choice d and $V_{id}(z)$ is the disutility, or costs, of choosing d at z. Agents select the option with the highest latent utility,

$$D_i(z) = \underset{d \in \{1,..,k\}}{\text{argmax}} \left(\mu_{id} - V_{id}(z) \right) \tag{2.1}$$

Potential choices in equation (3.1) depend on $\mathbf{Z}_i$ only through V_{id}, or costs. That is, changes in the instruments shift costs but do not alter the underlying preferences of agents for choice d.[9] Costs in V_{id} can depend on one or many instruments, and I can model how changes in a given $Z \in \mathbf{Z}$ will affect costs for option d. For example, increasing the geographical distance to option $1 \in d$ increases the costs of choosing 1. I am interested in agents who will change their choices as the instruments vary and the margins at which they change their behavior.

To understand how agents switch into, and away, from choices d as the instruments vary the researcher can define a set of rules of behavior. These rules stem directly from the properties of costs in V_{id}, and how they affect the relative mean utility of the different options. For instance, in the context of my childcare choice application, if changes in the distance to option 1 make this option more costly relative to other options then the researcher could assume that changes in distance move agents away from choosing 1. In the case of binary treatment with one binary instrument, agents can either move into (i.e., compliers) or away from (i.e., defiers) treatment as the instrument varies.[10]

I focus on contexts where multiple instruments can affect the same options, and agents can switch along many margins of choice. This is the case in settings where, for example, the researcher has access to several costs for an option, such as tuition, fees, travel time or distance. In addition, experimental variation can also be

[9]The assumption of separability between μ_{id} and V_{id} in equation 3.1 is standard in latent index selection models (Heckman et al., 2006; Lee and Salanié, 2020; Vytlacil, 2002).

[10]For example, suppose that treatment is enrolling, or not, in any college. The instrument is a variable that takes the value of one if the student grew up in a county with any college (as in Card, 1995). Always-takers (Never-takers) are students who would choose (not) to enroll in college whether or not they grew up in a county with any college. Neither always-takers nor never-takers respond to changes in the presence of college; they don't provide any variation to estimate treatment effects. Compliers are students who would enroll in college when their county had a college, and would choose to not enroll when there is no college present. Lastly, defiers are students who choose to enroll in college if they grew up with a college in their county, and to enroll if there was no college.

available along with information on the costs of different options. Heckman et al. (2006); Mogstad et al. (2020a,b) and Mountjoy (2019) show that if the researcher simultaneously increases one instrument (e.g. tuition) while decreasing another (e.g. travel time), agents can move towards, and away from, the same option. This motivates exploiting the variation in one instrument, while keeping the other instrument(s) fixed.

A2 *Conditional (Partial) Monotonicity* - Let z, z' be two values in $\text{supp}(Z_l)$, with $Z_l, Z_{l^-} \in \mathbf{Z}$. For all $d \in \{1, ..., k\}$ either $D_{id}(z, Z_{l^-}) \geq D_{id}(z', Z_{l^-})$ for all i, or $D_{id}(z, Z_{l^-}) \leq D_{id}(z', Z_{l^-})$ for all i. This a weaker version of monotonicity in Imbens and Angrist (1994).[11,12]

For binary treatment, Mogstad et al. (2020a) employs this assumption for identification. Similarly for binary treatment, Heckman et al. (2006) show that conditioning on other instruments can produce one-way flows in the variation of a selected IV. Mountjoy (2019) extends partial monotonicity to the case of multiple unordered options, and secures identification with an additional assumption of comparable compliers. The latter requires that compliers induced along the same margin of choice by two different instruments would have the same average potential outcomes. Instead, in my framework, I secure identification by imposing restrictions on the cost function such that instruments $Z \in \mathbf{Z}$ would satisfy *A2*.

[11]Monotonicity in Imbens and Angrist (1994) states that for any values z, z' in $\text{supp}(Z_l)$ either $D_i(z) \geq D_i(z')$ for all i, or $D_i(z) \leq D_i(z')$ for all i. It is a stronger assumption, since it has to hold regardless of changes in instruments other than Z_l.

[12]Das (2005) employs a conditional monotonicity assumption to the case of nonlinear IV. The author shows that, conditioning on covariates, nonparametric estimation can recover a conditional LATE parameter.

While the implications of uncontrolled variation have been previously analyzed, as well as the assumption of conditional or partial monotonicity to secure identification, I further extend this notion by focusing on the case where the researcher can define subsets of the support of Z_{l^-} such that agents move across well-defined margins. Heckman et al. (2006) show the case of conditional monotonicity for binary treatment, while in my context there can be multiple margins of choice. The latter complicates identification, since movements along margins of choice that violate monotonicity can be a result of incentives provided by the instruments and cannot be necessarily ruled out as defiers.

2.2.1 Restrictions on the cost function

One approach to determine if an instrument changes the relative utility of option d is to compare costs at different values of the instrument. This follows from the assumption that changes in $\mathbf{Z}$ only affect the utility of choice d through their effect on costs V_{id} (see equation (3.1)). If an instrument increases (or decreases) the costs of option d, while the costs of other options remain unchanged, then agents that respond to the variation should do so towards option d (i.e., the relatively cheaper option). By imposing restrictions on the shape of the cost function, the researcher can rule out combinations of potential choices and determine which margins of choice are affected by $\mathbf{Z}$. I impose the following restrictions:

R1 *(Cost restrictions)* Let $V_{id}(Z)$ represents the costs of alternative d which depend on instruments $Z \in \mathbf{Z}$. Let $Z_l \in \mathbf{Z}$ and define Z_{l^-} as an instrument in $\mathbf{Z}$

other than Z_l. Assume that the function $V_{id}(Z)$ is (weakly) convex in $Z_l \in \mathbf{Z}$ (i.e., $\frac{\partial^2 V_{id}(Z_l, Z_{l^-})}{\partial Z_l^2} \geq 0$).

R1.1 *(Joint effect)* Denote V'_{d,ll^-} as the cross-derivative of the cost function with respect to Z_l and Z_{l^-},

$$V'_{d,ll^-} = \frac{\partial^2 V_{id}(Z_l, Z_{l^-})}{\partial Z_l \partial Z_{l^-}}$$

If Z_{l^-} is discrete, then for two realizations z, z' the joint effect can be defined as

$$\begin{aligned} V'_{d,ll^-} = & \\ & [V_{id}(Z_l = z_l, Z_{l^-} = z) - V_{id}(Z_l = z_l, Z_{l^-} = z')] \\ & - [V_{id}(Z_l = z'_l, Z_{l^-} = z) - V_{id}(Z_l = z'_l, Z_{l^-} = z')] \end{aligned}$$

Thus, if $V'_{d,ll^-} \neq 0$, Z_l and Z_{l^-} jointly affect the costs of option d.

R1.2 *(Exclusion Restriction for Multiple Treatments)* Let Z_d be a set of instruments that affect choice d. If the conditions on the derivatives above hold then instrument Z_l and Z_{l^-} are in the set Z_d. Moreover, let Z_{d^-} represent the set of instruments that affect choices other than d. There exist at least one instrument in Z_d and not in Z_{d^-} such that it does not affect choices other than d.

The assumption on weak convexity of costs is frequent in economic theory and can be extended to applications that employ exogenous variation in cost-shifters as instrumental variables. The condition of weak convexity is necessary for a concave objective function, and therefore for a unique solution to the utility maximization problem. Moreover, imposing convexity on the cost function serves two main purposes. First, without any restrictions on the shape of the function multiple instruments can affect the costs in different directions. As a result, two agents experiencing a change of the same magnitude and sign in an instrument Z_l could face increasing or decreasing costs and could be induced to change their choices along different margins (i.e., towards, and away from, choosing different options). In turn, convexity imposes restrictions such that agents who would be induced to change their choices because of a change in Z_l would do so towards, or away from, the same options. The latter states that the assumption of convex costs translate into combinations of counterfactual choices that satisfy *A2*. Importantly, the restriction on costs does not require that $V_{id}(Z)$ is joint convex. Instead, it suffices that $V_{id}(Z)$ is element-wise convex in $Z_l \in \mathbf{Z}$.

Second, convexity also allows to model a richer set of cost functions where multiple instruments affect the same choices, than when each instrument can only affect one option. Joint effects are defined formally in *R1.1* . If the cross-derivatives are nonzero, such that they satisfy the convexity assumption, the effect on costs of one instrument can depend on other IVs. This can be the case in contexts where there is random assignment to receive a treatment but agents' reaction to it can be stronger or weaker depending on travel costs (time or distance). Lastly, *R1.2* imposes a type of exclusion restriction for multiple choice models (such as in Heckman et al., 2006; Heckman, Urzúa, and Vytlacil, 2008). It states that there should be at least one instrument that affects choice d but not other choices. Importantly, if the restrictions in *R1* hold, then the instruments satisfy conditional (partial) monotonicity and the researcher can identify a set of local treatment effects.

Thus, if the above conditions hold then Z_l satisfies *A2*, conditional on Z_{l^-}.

2.2.1.1 Example

To illustrate how changes in costs affect different margins of choice, consider the case of three choices, $d = \{1, 2, 3\}$, and two IVs, $\mathbf{Z} = \{Z_B, Z_C\}$, where Z_B is a discrete instrument. Let $U_d(Z_B, Z_C)$ represent the utility of option d, which depends on a binary instrument, Z_B, and a continuous instrument, Z_C. For instance, Z_B could represent random assignment with noncompliance to a treatment d and Z_C a measure of cost (e.g., price, travel time, travel distance). I impose the following restrictions on how changes in the instruments translate into changes in costs:

(i) Z_B decreases the cost of choosing option 3: $V_{d-}(0, Z_C) = V_{d-}(1, Z_C)$ for $d_- = \{1, 2\}$, and $V_3(0, Z_C) > V_3(1, Z_C)$

(ii) Z_C increases the cost of option 3: $\frac{\partial V_{d_-}(Z_B,Z_C)}{\partial Z_C} = 0$ and $\frac{\partial V_3(Z_B,Z_C)}{\partial Z_C} \geq 0$

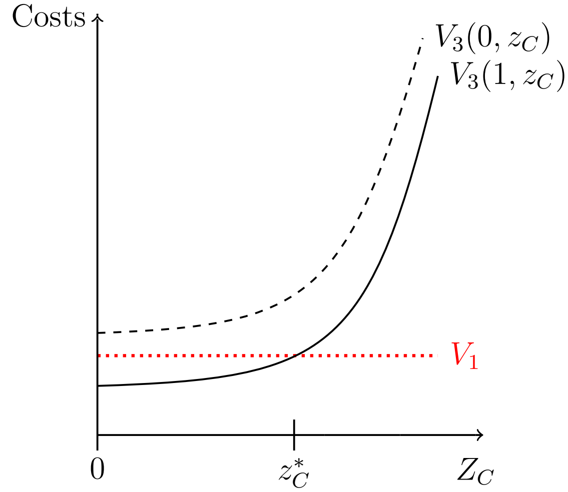
(iii) (Weak) Convexity of $V_3(Z_B, Z_C)$: $\frac{\partial^2 V_3(Z_B,Z_C)}{\partial Z_C^2} \geq 0$

The first restriction imposes that the binary instrument, Z_B, does not affect the costs of option 1 and 2 but it decreases the costs of choosing 3. Similarly, the second restriction requires that Z_C only affects the costs of option 3. I further assume that the costs of option 3 are convex in Z_C, or that at low Z_C costs of option

3 remain largely unchanged but increase fast as Z_C becomes larger. For instance, if Z_C is distance from home to option 3 convexity translates into larger increases in costs of choosing option 3 as it gets relatively far. Changes in Z_B and Z_C translate into changes in utility through their effect in costs. Hence, a higher Z_B and lower Z_C induce agents to move along the same margins of $1 \rightarrow 3$ and $2 \rightarrow 3$.

I depict a set of costs that satisfy the previous restrictions for option 1 and option 3 in Figure 2.1. For simplicity, I focus on one margin, $1-3$, but the analysis can be readily extended to all other margins. I also depict a case of interest, such that changes in the instruments would induce agents to change their choices. The dashed and solid convex curves represent the costs of option 3 when $Z_B = 0$ and $Z_B = 1$, respectively. The red dotted line represents the costs of option 1, which are constant for all Z_C. Note that a higher Z_B decreases the cost of 3 in the same magnitude, for a low Z_C. That is, the reduction in costs from $Z_B = 0$ to $Z_B = 1$ remains constant for low values of Z_C, and it narrows as Z_C increases. Thus, Z_B changes the intercept and the slope of the cost curve. For instance, suppose that Z_B is a random offer that encourages choosing 3 while Z_C is distance to option 3. For low Z_C the reduction in costs from the random offer remains about the same. As Z_C increases the effect on costs of the random offer largely disappears. Formally, $\frac{\partial U_3(1,Z_C)}{\partial Z_C} \neq \frac{\partial U_3(0,Z_C)}{\partial Z_C}$.

Figure 2.1: Relative costs that induce agents along the $1-3$ margin



Following Figure 2.1, what would be the choices of an agent with these costs? Note that for $Z_C < z_C^*$, the cost of option 1 lies between the cost of option 3 when $Z_B = 1$ and $Z_B = 0$, or that $V_3(Z_B = 0|0 < Z_C < z_C^*) > V_1 > V_3(Z_B = 1|0 < Z_C < z_C^*)$. Hence, if $Z_B = 0$ agents would choose 1 over 3 (since it has a higher cost), while if $Z_B = 1$ agents would choose 3 over 1. For a relatively low Z_C the variation in Z_B would induce agents to change their choice from option 1 to option 3. Meanwhile, if $Z_C > z_C^*$ the cost of option 1 is always lower and the incentives provided by Z_B would not be enough to change agents' choices. Moreover, this holds regardless of the assumptions on the slope of V_3. Assuming that cost curves also differ in their slope is nonetheless important, since it captures potentially stronger effects of Z_B on the costs of option 3 when the second instrument, Z_C, is low. All in all, this example shows how costs restrictions can rule out responses along margin $1-3$ for a high Z_C.

2.2.2 Conditional rules of behavior

The restrictions on costs translate into rules of behavior, or admissible potential choices that agents could make as a response to changes in an instrument Z_l. If there are binary or discrete IVs then there are limited states that the researcher has to analyze to define potential choices, and in turn complier groups.[13] However, in contexts with multiple instruments these rules alone are not sufficient to secure identification unless the researcher controls for the variation in a second instrument, Z_{l^-}. With this challenge in mind, I employ the approach and econometric model in Heckman and Pinto (2018) in which rules of behavior from revealed preference analysis define vectors that are used to identify the prevalence of compliers and their counterfactuals. In this section, I extend their method to the case of conditional vectors defined by an instrument Z_l. This vector changes along the support of a second instrument Z_{l^-}. Next (in Section 2.2.3), I present the econometric strategy and describe the role of the conditional vectors in the estimation of treatment effects.

An important component for setting the conditional rules is the definition of thresholds where the agents' behavior can change. For each agent i, the threshold $z^*_{il^-}$ consists of a realization of instrument Z_{il^-} where the costs of different options intersect. Importantly, the thresholds depend on the cost structure and the effect of the instruments on the costs of different options. If *A1* holds, then the thresholds

[13] For contexts where the instruments available are discrete, Heckman and Pinto (2018) show that there are subsets of complier groups that are justified by the mechanisms of the instruments and it is possible to identify treatment effects. Similarly, Lee and Salanié (2020) show that if there are choices targeted by the discrete instruments (e.g., Z_1 increases the mean relative utility of option 1), it is possible to estimate treatment effects for different complier groups.

must not depend on the unobserved components that affect outcomes and choices. What defines the thresholds are changes in costs such that the cheaper option is not the same above and below the threshold. Formally, $z^*_{il^-}$ is a threshold such that $V_{id}(Z_l|Z_{l^-} = z^*_{il^-}) = V_{id^-}(Z_l|Z_{l^-} = z^*_{il^-})$ and $V_{id^-}(Z_l|Z_{l^-}) = \min_d V_{id}(Z_l|Z_{l^-} = z^*_{il^-} + \epsilon)$ for an option d^- other than d.

Different agents can have distinct cost structures which can complicate the definition of a unique threshold $z^*_{l^-}$. I assume that $z^*_{l^-} = \max_i\{z^*_{il^-}\}$, which means that the conditional rules of behavior should only change once the share of agents responding to the instrument along one margin is sufficiently small. One threat to identification would occur if agents switch in and out of alternatives at multiple intervals of Z_{l^-}. If the cost function were to cross at multiple points with agents switching into and out of different options, such behavior would violate monotonicity. Convexity can help in reducing this threat such that, for instance, the costs of two options do not cross multiple times. The exclusion restriction for multiple treatments (*R1.2*) can also help overcome this challenge by avoiding multiple crossings of costs that would induce some agents into an alternative and other agents out of that alternative in the same conditioning interval.

The conditional rules consist of combinations of potential choices at different realizations z_l of instrument Z_l, conditional on Z_{l^-}. Let $D(z_l|z_{l^-})$ denote potential choices when the agent faces value $Z_l = z_l$ given $Z_{l^-} = z_{l^-}$. Combinations of $D(z_l|z_{l^-})$ form conditional response vectors denoted by S_l. Consider z_l, z_l'[14] two

[14] For binary variables, z_l, z_l' correspond to 0 and 1. For continuous variables, rather than defining potential responses at all, infinite, values of the continuous IV the researcher can place restrictions on choice behavior at marginal increases of the instrument.

realizations in $\text{supp}(Z_l)$, a response vector S_l denotes a combination of potential choices at z_l and z_l', given $Z_{l^-} = z_{l^-}$. The latter allows the response vector to vary depending on Z_{l^-}. Formally, $S_l(z_{l^-}) = [D(z_l|z_{l^-}), D(z_l'|z_{l^-})]'$. For instance, the costs structure in Figure 2.1 would imply a response vector $S_B(Z_C < z_C^*) = [1,3]'$ and $S_B(Z_C > z_C^*) = [1,1]'$. That is, below z_C^* agents that change their behavior do so along the $1-3$ margin; above $z_{l^-}^*$ changes in Z_l do not induce agents to change their choice of option 1. The total number of conditional response vectors, $N_{S,z_{l^-}}$, depends on the total possible choice combinations that follow from *R1* and satisfy *A2* at a given $Z_{l^-} = z_{l^-}$.[15]

Combinations of conditional response vectors that satisfy *A2* can be summarized in a matrix $R_l(Z_{l^-})$. I explicitly allow R to vary with a second instrument Z_{l^-}. The matrix $R_l(Z_{l^-})$ contains the conditional rules of behavior, and it describes the type of responses induced by changes in Z_l. For instance, in the binary-treatment-binary-IV case R would contain always-takers, never-takers, and compliers. The dimension of $R_l(Z_{l^-})$ is $l \times N_{S,z_{l^-}}$, where l is the number of values of the instrument at which agents' potential behavior is evaluated. For a binary instrument, l is equal to two. Similarly, for continuous instruments the researcher could analyze marginal changes in the instrument such that $l = 2$. For discrete instruments with more than two possible realizations, such as assignment to multiple treatment arms, $l > 2$.

To illustrate, in the example in 2.2.1.1 there would be two matrices containing conditional rules of behavior. Let $R_B(Z_C < z_C^*)$ and $R_B(Z_C > z_C^*)$ denote the

[15] Moreover, Heckman and Pinto (2018) show that the total number of response vectors cannot exceed $1 + [(k-1) \times l]$ where k are the total number of choices and l the number of values of the instrument that define the response vectors (e.g., $l = 2$ for a binary instrument).

rules of behavior for Z_C below, and above the threshold z_C^*, respectively. Below z_C^* instrument Z_B induces agents to change their behavior along the $1-3$ and $1-2$ margin, hence:

$$R_B(Z_C < z_C^*) = \begin{vmatrix} 1 & 2 & 3 & 1 & 2 \\ \\ 1 & 2 & 3 & 3 & 3 \end{vmatrix}$$

Above z_C^*, responses along the $1-3$ are not prevalent such that:

$$R_B(Z_C > z_C^*) = \begin{vmatrix} 1 & 2 & 3 & 2 \\ \\ 1 & 2 & 3 & 3 \end{vmatrix}$$

In both matrices, the first three columns denote always-takers for option 1, option 2, and option 3, respectively. The last two columns in matrix $R_B(Z_C < z_C^*)$ denote $1-3$ compliers and $2-3$ compliers induced by the variation in Z_B when Z_C is relatively low. The last column in $R_B(Z_C > z_C^*)$ represents $2-3$ compliers induced by the variation in Z_B when Z_C is relatively high.

In the next section, I describe the role of $R_l(Z_{l^-})$ in the identification of treatment effects.

2.2.3 Econometric Model

Let Y represent observed outcomes, $\mathbf{X}$ denotes baseline variables, and $\mathbf{V}$ are unobserved characteristics. Assume that instruments in $\mathbf{Z} = \{Z_1, ..., Z_L\}$ satisfy *A1*; and that the restrictions in *R1*, which imply *A2*, also hold. Combined with these assumptions, the following equations represent a standard IV model of multivalued choices:

$$\text{Outcome equation: } Y = f(D, \mathbf{X}, \mathbf{V}) \tag{2.2}$$

$$\text{Choice equation: } D = g(\mathbf{Z}, \mathbf{X}, \mathbf{V}) \tag{2.3}$$

where D is a categorical variable that takes values in $d \in \{1, 2, ..., k\}$, $\mathbf{V}$ is an unobserved component that affects choices and outcomes. Moreover, $\mathbf{Z}$ only affects outcomes through its effect on choices (i.e., exclusion restriction). Let $\mathcal{Z}$ represent the joint support of instruments in $\mathbf{Z}$, and denote an element $z \in \mathcal{Z}$ as a combination of $(z_1, ..., z_L)$. Let $Z_l, Z_{l^-} \in \mathbf{Z}$ represent instrument l while l^- represents instruments other than l. Potential choices $D(z_l, z_{l^-})$ satisfy *A2* conditional on Z_{l^-}. Combinations of conditional potential choices, $D(z_l|z_{l^-})$, define response vectors S_l.

In what follows, I keep implicit the variables in $\mathbf{X}$. To highlight the role of conditional response vectors, note that potential choices when $Z_l = z_l$ given $Z_{l^-} = z_{l^-}$ can be rewritten using equation (2.3) as $D(z_l|z_{l^-}) = g(z_l, \mathbf{V}|Z_{l^-} = z_{l^-})$. This translates into conditional response vectors as conditional functions of the unobservables $\mathbf{V}$:

$$\begin{aligned}
S_l(z_{l^-}) &= [D(z_l|z_{l^-}), D(z_l'|z_{l^-})]' \\
&= [g(z_l, \mathbf{V}|Z_{l^-} = z_{l^-}), g(z_l', \mathbf{V}|Z_{l^-} = z_{l^-})]' \\
&= g_S(\mathbf{V}|Z_{l^-} = z_{l^-})
\end{aligned}$$

Hence, conditional response vectors generate partitions of the unobservables, $\mathbf{V}$, at a given Z_{l^-}. Combinations of $S_l(z_{l^-})$ define conditional response matrices, $R_l(Z_{l^-})$ with dimension $l \times N_{S,z_{l^-}}$. Each column in $R_l(Z_{l^-})$ denotes a type of response, denoted as $g \in \{1, ..., N_{S,z_{l^-}}\}$. For instance, in the binary-treatment-binary-IV case, g_1 denotes always-takers, g_2 denotes never-takers, and g_3 denotes compliers.

Let $B_l^d(Z_{l^-})$ denote a binary matrix that takes the value of one every time option d is chosen in matrix $R_l(Z_{l^-})$. Define $B_l(Z_{l^-})$ as a matrix that stacks option specific B_l^d such that:

$$B_l(Z_{l^-}) = \left(B_l^1(Z_{l^-})', B_l^2(Z_{l^-})', .., B_l^k(Z_{l^-})'\right)'$$

Importantly, matrix $B_l(Z_{l^-})$ depends on Z_{l^-}. This matrix weights observed choices and outcomes based on admissible potential responses (i.e., conditional rules of behavior) to changes in the instruments. Appendix A.2 describes how to estimate the probability of each type of response in $R_l(Z_{l^-})$, as well as their average baseline characteristics. The main contribution of this framework, in contrast to the methodological approach in Heckman and Pinto (2018), is to define local response matrices in $R_l(Z_{l^-})$ rather than a general response matrix R_l that does not vary with changes in the other instruments available.

The identification of counterfactuals for each response in matrix $R_l(Z_{l^-})$ depends on the margins and choices agents are induced to choose as a result of variation in Z_l. For instance, if Z_l induces agents away from choosing option 1 and option 2 and towards choosing option 3 there is not enough variation to separately identify the counterfactuals of choosing option 3 along both margins. Therefore, the researcher can identify the counterfactual of option 3 *versus* the next-best (i.e., what the agent would have chosen if option 3 was no longer available).

To define the counterfactuals that can be identified, let $\sum_{dl}(q|Z_{l^-})$ be a set that contains responses in matrix $R_l(Z_{l^-})$ where d appears q times, with $q \in \{1, ..., l\}$. Let $b_{dl}(q|Z_{l^-})$ be a binary vector that takes the value of one every time option d appears q times in matrix $R_l(Z_{l^-})$. Let YD_d denote a vector of interactions between outcome Y and choice $d \in \{1, 2, ..k\}$. Define $L^d_{Yl}(Z_{l^-}) = [E[YD_d|z_l, Z_{l^-}], E[YD_d|z'_l, Z_{l^-}]]'$ as the average of outcome Y when option $d \in \{1, 2, ..k\}$ is chosen, evaluated at realizations z_l, z'_l conditional on Z_{l^-}. If assumption *A2* holds, then the following counterfactuals can be identified:

$$E\left(Y_d \mid g \in \sum\nolimits_{dl}(q|Z_{l^-})\right) = \frac{b_{dl}(q|Z_{l^-})B_l^{d+}(Z_{l^-})L^d_{Yl}(Z_{l^-})}{b_{dl}(q|Z_{l^-})B_l^{d+}(Z_{l^-})\Pr_{Z_l}(D = d|Z_{l^-})} \tag{2.4}$$

where $B_l^{d+}(Z_{l^-})$ is the Moore-Penrose pseudoinverse; and $\Pr_{Z_l}(D = d|Z_{l^-})$ is the propensity score of choice d evaluated at realizations $z_l, z'_l \in \text{supp}(Z_l)$, conditional on Z_{l^-}. That is, $\Pr_{Z_l}(D = d|Z_{l^-}) = [\Pr_{Z_l}(D = d|z_l, Z_{l^-}), \Pr_{Z_l}(D = d|z'_l, Z_{l^-})]$. Appendix A.1 formally shows how partial monotonicity in *A2* translates into the identification of conditional counterfactuals in equation (2.4).

If the set $\sum_{dl}(q|Z_{l^-})$ for $g \in [1, l-1]$ contains exactly one element then response specific counterfactuals can be identified. Otherwise, the counterfactuals for d, such that set $\sum_{dl}(q|Z_{l^-})$ for $g \in [1, l-1]$ is not unique, are a combination of next-best alternatives. Consequently, in cases as the latter the researcher cannot separately identify local treatment effects for each choice d. Instead, it is possible to estimate local treatment effects of choice d *versus* the next-best.

2.2.3.1 Beyond the next-best

One limitation of the standard IV framework of multiple unordered choices for the identification of treatment effects is that in some contexts the researcher can at most identify local treatment effects of an option d *versus* the next-best (Heckman et al., 2006; Kirkeboen et al., 2016). However, these effects can be uninformative and difficult to interpret. This is specially the case when the alternatives that belong to the next-best are systematically different such that agents can derive different gains or losses depending on which of these alternatives is their fallback option. Separate identification of these effects is an empirical challenge, that might have policy implications.

Previous literature has employed assumptions on the counterfactuals to disentangle treatment effects along well-defined margins of choice. Kline and Walters (2016); Lee and Salanié (2020) employ an homogeneity assumption such that the potential outcome from choosing option d is assumed to be the same along all margins at which agents are induced towards choosing option d.[16] For instance, if an instrument induces agents along the $1-3$ and $2-3$ margins the homogeneity assumption

[16] Along the lines of achieving separate identification, Mountjoy (2019) exploits variation in a second instrument but assumes that compliers induced along the same margin of choice by two different instruments would have the same average potential outcome. The author refers to this assumption as comparable compliers, and employs it in the case of two-*versus* four-year college enrollment. In turn, Pinto (2019) uses properties of the response matrix, revealed preference, and an extension of the Local Instrumental Variables model to multiple choices to achieve separate identification of counterfactuals. Hull (2018) presents a method to recover margin-specific LATEs by exploiting variation in one instrument and its interaction with baseline covariates.

can be summarized as $E(Y_3|D(z_l) = 1, D(z_l') = 3) = E(Y_3|D(z_l) = 2, D(z_l') = 3)$. I extend this assumption to the case of conditional responses and propose a test to empirically assess its feasibility in the context of the approach of this chapter.

Let d^- denote options other than d. Let d' be an alternative in d^-. If $\sum_{d'l}(q'|Z_{l^-}) \subset \sum_{dl}(q|Z_{l^-})$ for $q, q' \in \{1, ..., l-1\}$, then local treatment effects of $E\left(Y_d - Y_{d^-}| \sum_{dl}(q|Z_{l^-})\right)$ are identified. The conditional homogeneity assumption states that:

$$E\left(Y_d \mid g \in \sum\nolimits_{dl}(q|Z_{l^-})\right) = E\left(Y_d \mid g' \in \sum\nolimits_{dl}(q|Z_{l^-})\right)$$

for all responses $g, g' \in \sum_{dl}(q|Z_{l^-})$. In the example of responses along the $1-3$ and $2-3$ margins, g refers to the former and g' refers to the latter. More generally, g, g' can be referred to as switchers. One important caveat is that as the number of groups g becomes larger, homogeneity might be less plausible. Nonetheless, the assumption is easy to implement and can, at least, bound treatment effects at specific margins of choice.

More important, responses g and g' can stem from agents that are systematically different and assuming homogeneity can under-or over-estimate their treatment effects. For instance, agents switching along the $1-3$ margin might largely differ from those switching along the $2-3$ margin. I propose a test that exploits differences in the conditional response matrix $R_l(Z_{l^-})$ along the distribution of Z_{l^-}. The test can be implemented in contexts where $R_l(Z_{l^-})$ contains only one type of switchers in an interval of Z_{l^-}. Suppose that for $Z_{l^-} < z^*_{l^-}$ switchers consist of responses along $1-3$ and $2-3$ such that the researcher can identify LATE of option 3 *versus* the next-best. Now suppose that responses along the $1-3$ margin are not prevalent for $Z_{l^-} > z^*_{l^-}$. Then, at this interval of Z_{l^-} treatment effects along margin $2-3$ can be separately identified. The intuition of the test I propose is to compare counterfactuals and average of baseline variables for agents along the $2-3$ margin around a neighborhood of threshold $z^*_{l^-}$.

More generally, for a response g that belongs to matrix $R_l(Z_{l^-})$ above and below threshold $z^*_{l^-}$,[17] I assume comparable counterfactuals:

$$\lim_{Z_{l^-} \to z^{*}_{l^-}{}^{+}} E(Y_d | g \in R_l(Z_{l^-})) = \lim_{Z_{l^-} \leftarrow z^{*}_{l^-}{}^{-}} E(Y_d | g \in R_l(Z_{l^-}))$$

Implicitly, this assumption states that agents who exhibit the same potential behavior above and below a threshold of the second instrument, Z_{l^-}, are similar in their observed and unobserved characteristics. Differences in unobserved characteristics cannot be tested. However, using the average of baseline variables it can be tested whether the following holds,

$$\lim_{Z_{l^-} \to z^{*}_{l^-}{}^{+}} E(X_{gl} | g \in R_l(Z_{l^-})) = \lim_{Z_{l^-} \leftarrow z^{*}_{l^-}{}^{-}} E(X_{gl} | g \in R_l(Z_{l^-}))$$

An example gives some intuition. If agents who switch along margin $2-3$ appear above and below $z^*_{l^-}$ then, in the limit around the threshold, they are largely indistinguishable except from differences in Z_{l^-}. By *A1*, differences in instrument Z_{l^-}

[17]To be precise, $\sum_{dl}(q|Z_{l^-} = z^{*}_{l^-}{}^{-}) \cap \sum_{dl}(q|Z_{l^-} = z^{*}_{l^-}{}^{+}) = \{g\}$ and either $\sum_{dl}(q|Z_{l^-} = z^{*}_{l^-}{}^{-}) \subset \sum_{dl}(q|Z_{l^-} = z^{*}_{l^-}{}^{+})$ or $\sum_{dl}(q|Z_{l^-} = z^{*}_{l^-}{}^{+}) \subset \sum_{dl}(q|Z_{l^-} = z^{*}_{l^-}{}^{-})$. That is, g is the only response in both sets and it is unique for either set.

should not affect potential outcomes, and g switchers above and below the threshold should be comparable.

2.3 Empirical Application: Parental choice of childcare

In this section, I study the case of childcare choices in Colombia and their impact on the cognitive, socio-emotional, and nutritional development of children. I begin by describing the public childcare system in Colombia and the data I employ.

2.3.1 Background

The public childcare system in Colombia provides free care for children of low income families between the ages of six months to five years. The majority of the childcare supply consists of small nurseries (s) and large centers (l). Small nurseries are run by one caretaker that serves 12-15 children in the same space, typically the home of the provider. From the 1980s to 2011, these small nurseries were the main providers of public childcare for low income families. In 2011 the public provision of childcare was expanded to offer services through both s and l. The latter serve around 300 children who receive care in age-specific classrooms, with one teacher per 25 children, a nutritionist and a psychologist in each center, and administrative staff.[18]

The expansion of the public provision of childcare that started during 2011 has interesting institutional features that help explain how parents sort into the different

[18]For further details on the supply of childcare in Colombia and the characteristics of each center, see Bernal et al. (2019) and Bernal and Ramírez (2019).

alternatives. First, eligible children living nearby to a new l were given priority to obtain a slot. Parents living near l are both more likely to be aware of the new option of childcare and have a higher chance of obtaining a slot. Second, given that s and l are publicly provided, caretakers who were initially at s could switch to work for l. If the caretaker of an existing s would decide to transfer to l it would force parents to either transfer to l or find an alternative childcare option. Third, the expansion was part of the national early childhood strategy whose goal was to provide high-quality integrated services (e.g., nutrition, care, health, psychological services) to low-income children ages 0-5. Most of these services were not offered in s, which mainly provided care, psycho-social stimulation, and covered a fraction of the children's nutritional needs.

Despite the fact that l centers offer nutritional and psychological services and have better infrastructure than s centers, there are other factors that could improve or deter children's development. Age-specific classrooms may improve cognitive development, but higher teacher to student ratios may have a negative impact. Having a nutritionist could be beneficial for children at risk, but they can be supervised more closely in smaller classrooms. The heterogeneity of the potential benefits of s and l makes it difficult to rank them *a priori*. Moreover, parents with preferences for centers that provide nutritional services could prefer l over s, while those concerned with socio-emotional development could rank s (where children of all ages are interacting) above l. Also, parents with strong preferences for low caretaker to children ratios could prefer to take care of their child at home than either at s or l. Rather than imposing a rank that would restrict parent's preference heterogeneity, I assume that these alternatives are unordered (e.g., l centers are no better than s ones across all dimensions of children development).

2.3.2 Data

I exploit rich data containing a random sample of children who were enrolled in s, before l was available (see Bernal et al., 2019). That is, all children in my sample were enrolled in s at baseline and I observe their enrollment at follow-up which could either be home care (h), s, or l. Given that I follow an IV approach, the effects I can uncover are local but even more so in my case given that I do not observe children who were out of the public care system prior to the expansion to l centers. The dataset contains socioeconomic variables collected at baseline such as children's age and sex, mother's years of education, number of children in the household, and household income. It also has age-standardized measures of child development on cognitive, socio-emotional, and nutritional dimensions, which were collected before and after l was available.[19] In particular, cognitive development comes from measures on communication, problem solving, and the measure of fine and gross motor skills from the Ages and Stages Questionnaire (ASQ).[20] The measure of socio-emotional development is a composite of scores on compliance, adaptive functioning, affect, and interaction from the ASQ. Nutritional development is measured using weight-for-age, height-for-age, and arm-circumference for age.

[19] Details on data collection procedures and the timeline are in Bernal et al. (2019). In summary, collection of baseline data took place between November 2010 and May 2011. Collection at follow-up took place in two stages: (i) November-December 2011, and (ii) September-November 2012.

[20] The ASQ is a tool to screen the development and socio-emotional progress of children younger than 6. Higher scores on the cognitive measures indicate higher cognitive development; higher scores on the socio-emotional measures indicate behavioral problems.

At follow-up, more than half of the children in the sample attend l, compared to 32% in s and 13% who receive care at home (h). Table 2.3 presents summary statistics of baseline (before the expansion took place) characteristics and childcare choices. The sample contains children who were 13 to 55 months of age at baseline, with an average of 37 months of age. Half of the sample are male children, living at homes with an average of 1.5 children between 0-5 years of age. They have mothers with an average of 9 years of education, which corresponds roughly to incomplete secondary. The sample only contains families living in poverty, who are eligible for public childcare, with about 44% of children belonging to the poorest income levels.

Although children at h belong to the most disadvantaged families, their cognitive development level at baseline was higher than for children enrolled in alternatives l or s (Table 2.4). In contrast, socio-emotional and nutritional development for those in h was below than for children enrolled in s and l. The differences in baseline development for children in h, s, and l are indicative of heterogeneity in sorting patterns. Those at h are about one month older, on average, than those in l and s centers. There is a higher fraction of girls among those in h, at 63%, than in s (47%) and l (48%). Average years of education of the mother are lower by almost one year for children in h, compared to those in l and s. In terms of average outcomes, children in l show higher cognitive and nutritional development than those in s and h. The latter show higher average socio-emotional development, while those in l perform the worst in this dimension.

Many observed and unobserved factors determine the childcare option that parents select. Family size, labor market participation, and the level of education of the parents are some of the observed factors that could influence the type of care they choose. Parents can have preferences for low costs, for the nearest alternative to their home, or for options with more staff per children. Unobserved factors, such as how concerned parents are about their child's development, affect both their care decisions and the child's outcomes. Hence, childcare choices are endogenous with parents self-selecting into their preferred alternative.

In order to estimate treatment effects I exploit two sources of exogenous variation: an experiment aimed at providing information to families about the supply of centers, and the geographical distance between the child's home and the different options. Thus, I analyze two potential instruments: random assignment to transfer from s to l, and distance to l from the child's home. The former is a lottery assigned at the s level that offered the chance to transfer to l, for caretakers and children at winner s. The lottery did not provide financial incentives to transfer. Instead, it increased the likelihood of being informed about l and transferring to l. This transfer was not enforced. Distance to l is the straight line distance from the child's home to the nearest l center, measured in kilometers.

About 74% of children in the sample were randomly assigned to transfer with their caretaker directly from s to l ("Wins lottery", Column 5 in Table 2.5), while the remaining 26% are control children who did not win the lottery. Information on this transfer opportunity was given to caretaker in winner s centers, who could decide whether or not to contact and inform parents. Neither caretakers nor parents at winner centers were required to transfer, and caretakers and parents in control s centers could also choose to switch to the nearest l. In this sense, there is noncompliance in the parental response to the lottery. Table 2.5 shows that 80% of l children and h children won the lottery, while this fraction is 61% among those remaining in s.[21]

On average, parents who choose s live almost twice as far from l centers than those who select l. In this sense, distance to l is an important predictor of parents choosing to switch to l. In contrast, average distance to s is similar across types of care and for the full sample, at about 0.3km. This is expected, since all children in the sample were enrolled in s centers at baseline.

The distribution of choices at different outcomes of the lottery shows that, on average, the lottery encouraged parents to switch from s into l and h. Figure 2.3 shows that 60% of children who were assigned to transfer choose l, and this fraction is almost 20 percentage points higher than for those in the control group. The fraction of children at h is also higher among those who won the lottery, but only by about four percentage points in comparison to the control group. Enrollment in s centers shows the opposite pattern: 49% of parents of control children choose s, and this fraction decreases to 27% among lottery winners. These patterns are consistent with the reshuffling of childcare options caused by the experiment. Recall that the sample consists of children originally at s, and in the scenario of winning the lottery parents could choose to switch to l or h, or remain in s if the caretaker did not switch to l.

Figure 2.4 shows choice of childcare across quintiles of distance to l. En-

[21] Bernal et al. (2019) show that 75% of caretakers at winner s centers transfer to l, while only 40% of caretakers at control s do so.

rollment in l decreases as the distance to those centers increases. In particular, enrollment at the lowest distance quintile is almost four times higher than that of the highest distance quintile. Enrollment in s displays the opposite pattern, with higher enrollment at high levels of distance to l. These patterns follow the logic of substitution of care such that as l becomes more costly (i.e., as distance to l increases) enrollment in s increases. Lastly, the fraction of children at h shows an inverted U shape, which suggests that there is a turning point at which parents go from substituting l with s *and* h, and start to substitute l *and* h with s. One way to explain why parents would alter their preferences at different levels of l could be that they pay more attention to attributes (e.g., costs) that are more salient.[22] For instance, when l is very close to home parents could give more importance to how little it would take to bring their child to the center than to other attributes of l; in turn, when l is far, distance becomes less relevant and other costs (e.g., the cost of obtaining a slot easily) could matter more in the childcare decision.

2.3.3 Endogenous Choices of Childcare and Children Development

This section follows the general framework in Section 2.2. To start, I define potential childcare choices and potential development outcomes.

Suppose a parent (decision maker), denoted by i, that can choose between three childcare alternatives: home care (h), small nurseries (s), and large centers

[22]This type of behavior, or preference distortion, fits with the context-dependent model of choice formalized by Bordalo, Gennaioli, and Shleifer (2013). While I do not model distortions in parental preferences, the insights from this literature can help explain some of the choice patterns I observe. A formal model of preference distortion could help predict intervals in the support of the instrument where monotonicity holds locally, but not (globally) across intervals.

(l). Let d denote parental choice of childcare, with $d \in \{h, s, l\}$. Although l provided more services than s, they also provided care for a larger number of children at higher teacher per child ratios. One could assume that l are better than s in only one, or some, dimensions such as cognitive or nutritional development. However, given the holistic nature of children development, fostering investments in one dimension at the expense of others could affect many areas of children development. Since the care offered at l and h differs in many ways from the care offered at s, imposing a ranking between these options would require strong assumptions. Hence, I specify these options as unordered, in the sense that *a priori* s or l are no better than h, and l is no better than s.

The choice of childcare, d, affects children development denoted by Y_i. I focus on three dimensions: cognitive, socio-emotional, and nutritional. X_i contains exogenous covariates, measured at baseline prior to the decision between h, s, and l. These baseline variables include children characteristics such as age in months, sex, and measures of development (cognitive, socio-emotional, and nutritional). It also includes household income and mother's years of education. Furthermore, let $Z_i \in \{Z_1, Z_2\}$ contain instrumental variables where Z_1 is a binary variable and Z_2 is a continuous variable. Specifically, Z_1 is the outcome of the random lottery and Z_2 is the distance to l. Denote z as a realization of Z_i; for instance, z_1 represents values of the lottery which could be either zero or one.

Potential treatment is denoted by $D_i(z_1, z_2)$ and represents the childcare choice that parent i would make when offered instrument value z. To illustrate, for the binary instrument $D_i(1, z_2)$ (resp. $D_i(0, z_2)$) would be the childcare option that the parent selects when the outcome of the lottery is one (resp. zero). Potential outcomes are denoted by $Y_{di}(z_1, z_2)$, which represents children development when parent i chooses childcare d at instrument values (z_1, z_2). That is, $Y_{hi}(z_1, z_2)$ represents child development when parents select home care at value z_1 of the lottery and z_2 of the distance. $Y_{si}(z_1, z_2)$ represents child development when parents select small nurseries, s, at value (z_1, z_2) of the instruments. $Y_{li}(z_1, z_2)$ represents child development when parents select large centers, l, at value (z_1, z_2) of the instruments. Note that I could compare children development across three margins of choice: h *versus* s, h *versus* l, and s *versus* l. The first two margins could be thought of as the extensive margins (i.e., the choice between care at home *versus* care in the public system). The margin of s *versus* l would be commonly referred to as the intensive margin, but in a setting of unordered choices is not clear that l provides more or better care.

The instruments in Z satisfy $(Z_1, Z_2) \perp\!\!\!\perp (Y_h, Y_s, Y_l, \{D(z_1, z_2)\}_{z_1, z_2 \in \mathcal{Z}})$ for all possible values of (Z_1, Z_2), denoted as $\mathcal{Z}$. If this assumption (of independence) holds, $Y_{di}(z_1, z_2) = Y_{di}$ for all possible values of (Z_1, Z_2). Instruments in Z have to satisfy the exclusion restriction: they should only affect outcomes through their effect on choices. The lottery, Z_1, was randomly assigned and was intended to induce enrollment at l. As such, it should not have a direct effect on child's development. However, even if lottery assignment is random, being treated could affect child's development regardless of choices, because it disrupted established childcare (e.g., some children lost their existing small center). On the other hand, distance to l could be correlated with local factors that can affect children's development. In the analysis I present here, my objective is analyzing the variation that stems from random assignment with noncompliance conditional on the variation from cost-shifters that are frequently used for identification, rather than justifying the exogeneity of the instruments. Nonetheless, I control for local conditions that can affect the development of children with fixed effects at the city level.

In contexts where there are multiple unordered alternatives the variation provided by the instruments can stem from changes along many different margins. In the childcare choice context, if the instruments provide any variation they could do so by inducing agents along any of the three possible margins: $s \leftrightarrow l$, $s \leftrightarrow h$, or $h \leftrightarrow l$. I use $\leftrightarrow$ to indicate that without imposing any assumptions on the incentives of each instrument, agents could shift across margins in any direction. To illustrate, suppose there is only a binary instrument (such as the lottery) and parents can only choose between s and l. If the lottery provided incentives to choose l I would assume that when, the lottery outcome goes from zero to one, parents who change their choices do so from s to l. These are counterfactual, unobserved, choices since I cannot observe parents in both states (winning and not winning the lottery). Using the notation above, for parents who comply with the lottery I would assume $D(0, z_2) = s$ and $D(1, z_2) = l$. These are referred to as compliers, while parents who do not change their behavior as a result of the lottery are always-takers (of either s, $D(0, z_2) = s$ and $D(1, z_2) = s$, or l, $D(0, z_2) = l$ and $D(1, z_2) = l$). Always-takers do not provide variation to estimate treatment effects; they would not change their behavior as a response to the instrument.

Now suppose the lottery outcome is still binary but parents can choose between

s, l and h. Depending on the assumptions and incentives provided by the lottery I would have multiple complier groups. With only two childcare options I assumed one complier group, $D(0, z_2) = s$ and $D(1, z_2) = l$. Now, with three options, it could be that $D(0, z_2) = s$ and $D(1, z_2) = h$, or $D(0, z_2) = h$ and $D(1, z_2) = l$, among others. Similar to the setting I analyze, Kline and Walters (2016) exploit an experiment that allowed parents to enroll their children in Head Start and estimate the impact of Head Start in the presence of close substitutes of care. The authors restrict parental behavior by assuming that the Head Start offer should only induced parents to choose Head Start (rather than choose other options of care).[23]

The studies mentioned above examined the case of multiple choices with binary or discrete instruments. One advantage of estimating with binary or discrete IVs is that there are limited states that the researcher has to analyze to define counterfactual choices, and in turn complier groups. In the example of the binary outcome of the lottery, parents would face only two scenarios: winning the lottery or not. Suppose there is a third value of the instrument that provided monetary incentives for parents to enroll their children in l. There could be parents who only choose l if they receive monetary incentives. Or, parents who only choose l if they receive an offer to transfer to l. Or, parents who choose l either if they obtain a slot or obtain financial resources. Those are only a few examples of potential parental behavior, and in practice there can be many complier groups with heterogeneous

[23] Childcare choices and their effects on children development have also been explored using Machine Learning methods. Dean and Jayachandran (2019) analyze the effect of expanding access to kindergarten in India. They employ Machine Learning methods to predict counterfactual options of care and estimate treatment effects for different complier groups. See also Rodriguez and Saltiel (2020) for an application of machine learning methods to the context of choice of preschool.

treatment effects.

In addition to the lottery outcome, I exploit variation in distance to l. Imagine a parent i who was five miles away from l and now is only a mile away from l. If proximity, a smaller distance, from l decreases the costs of bringing the child to l, then parent i could choose l when it is only a mile away. In general, I assume that when the continuous distance decreases parents who change their behavior would do so by choosing l. However, as I discussed in the Data section, parents could react differently to distance when l is near than when it is further away. Imagine that parent i cares for her child at home and lives 50 miles away from l, would we expect her to choose l if the distance to the center decreases in one mile? What if her home was 5 miles away from l and now there is a center 4 miles away? Assuming that parents will value changes in the distance equally, no matter what the distance is, can be restrictive.[24]

Given that both the lottery outcome and distance to l affect the choice of childcare l, this setting is an example of the case of multiple instruments that affect multiple margins of choice. One of the implications mentioned in Section 2.2 is that, unless the researcher conditions on one of the instruments, there is uncontrolled variation that would result in a violation of monotonicity. If parents win the lottery and simultaneously the distance to the l center increases, parents could be induced to switch into and out of choosing l. Thus, I follow the conditional approach in Section 2.2 and define potential parental choices at different outcomes of the lottery

[24] Assuming that agents react to changes in distance to an option equally at all values of the distance could be reasonable in scenarios where agents are willing to travel far, such as college choice. In the case of childcare choices traveling longer distances is less attainable.

conditional on distance to l. To do so, in the section that follows I present a latent utility model of childcare choices where I impose restrictions on the costs of childcare. Then, I present an exploratory analysis to assess those assumptions in the data. Last, I define the combinations of potential parental choices that satisfy those restrictions.

2.3.4 Latent Utility Model of Childcare Choices

Assume that parental choice of childcare can be represented by a discrete choice problem with additively separable errors. Let $U_{id}(z_1, z_2)$ represent latent utility at values $Z_1 = z_1$ and $Z_2 = z_2$ of the instruments, and assume it depends on unobserved preferences and costs:

$$U_{id}(z_1, z_2) = \mu_{id} - V_{id}(z_1, z_2)$$

$V_{id}(z_1, z_2)$ represents the disutility of choosing d at values $Z_1 = z_1$ and $Z_2 = z_2$ of the instruments. μ_{id} represents unobserved preference heterogeneity for choice d. Parents choose the option with the highest latent utility,

$$D_i(z_1, z_2) = \underset{d \in \{h,s,l\}}{\operatorname{argmax}} \left(\mu_{id} - V_{id}(z_1, z_2)\right)$$

V_d depends on two instruments Z_1, the lottery outcome, and Z_2, distance to l. The lottery offered the chance to transfer to l, but it was not enforced. Given that the lottery was intended towards decreasing costs of attendance and information of l, I assume that the cost of choosing l is lower when parents win the lottery: $V_l(1, Z_2) < V_l(0, Z_2)$. In contrast, the outcome of the lottery should not affect the cost of s. That is, $V_s(1, Z_2) = V_s(0, Z_2)$.

I assume that the disutility from home care h is not affected by the lottery outcome, $V_h(1, Z_2) = V_h(0, Z_2)$. Children who do not enroll in any early education center could receive care from their parents and other family members (e.g., siblings, grandparents) at the expense of income from work or hours of education. None of these potential costs should be affected by the outcome of the lottery, and it is reasonable to assume that the absolute value of home care is the same whether or not parents win the lottery. Given the assumptions on changes in costs as the outcome of the lottery changes it follows that $U_l(1, Z_2) \geq U_l(0, Z_2)$, $U_s(1, Z_2) = U_s(0, Z_2)$, and $U_h(1, Z_2) = U_h(0, Z_2)$.

Let $Z_2 \in \mathbb{R}$ denote the distance to l from the child's home. Z_2 is the straight line distance from the child's home to l, measured in kilometers. Measures of proximity to different alternatives have been previously used as instruments for years of education (Card, 1995) or enrollment in two -or four-year college (Mountjoy, 2019). Proximity to an alternative follows the logic of a cost-shifter, e.g. the closer parents are to l the less costly it is to enroll children and bring them to daycare. Parents in near proximity to l have lower costs of obtaining information about the type of care as well as the enrollment criteria and process. In addition, parents with eligible children (age and poverty level) living in the vicinity of l were given priority to obtain a slot. Hence, for those living near to l, the cost of obtaining a slot and sending their child to care is lower.

In contrast to the case of the binary lottery outcome in Z_1, Z_2 can take on many possible values. Since Z_2 follows the logic of cost-shifter I assume $\frac{\partial V_l(Z_1,Z_2)}{\partial Z_2} \geq 0$, or that the cost of choosing l increases as the distance from the child's home to l increases. Since $U_l(Z_1, Z_2)$ only depends on Z_2 through $V_l(Z_1, Z_2)$, then $\frac{\partial U_l(Z_1,Z_2)}{\partial z_2} \leq 0$. For h and l, I assume $\frac{\partial V_h(Z_1,Z_2)}{\partial Z_2} = 0$, which means that the cost of choosing h is not affected by the distance to l. Similarly, for s I assume that z_2 has no direct effect on the cost of s and hence $\frac{\partial V_s(Z_1,Z_2)}{\partial Z_2} = 0$. From these assumptions on how costs vary as the distance to l changes, it follows that $\frac{\partial U_l(Z_1,Z_2)}{\partial Z_2} \leq 0, \frac{\partial U_s(Z_1,Z_2)}{\partial Z_2} = 0$, and $\frac{\partial U_h(Z_1,Z_2)}{\partial Z_2} = 0$.

Table 2.1 shows the assumptions on changes in utility by changes in the different instruments. Changes in the outcome of the lottery increase the relative utility of l w.r.t h and s. It follows that changes in the lottery outcome would induce parents along two margins. In specific, when $z_1 = 0 \rightarrow z_1 = 1$ parents would be induced to change their choices from $h \rightarrow l$, and $s \rightarrow l$. On the other hand, changes in distance to l decrease the relative utility of l w.r.t h and s. Hence, Z_2 induces parents to switch along the margins $l \rightarrow h$ and $l \rightarrow s$. These changes satisfy monotonicity (i.e., there are no two way flows when each instrument, separately, change). Since changes in the lottery outcome would induce parents towards l from s and h, the variation in Z_1 does not allow to identify treatment effects along each of the two margins. At most, it would be possible to identify the treatment effect of l *versus* the next-best (i.e., the preferred option between h and s).

I analyze the joint variation from Z_1 and Z_2 and its potential effect on the utility of each childcare option.[25] First, I assume that $\frac{\partial V_h(1,Z_2)}{\partial Z_2} = \frac{\partial V_h(0,Z_2)}{\partial Z_2} = 0$. This assumption states that changes in distance to l do not affect the cost of choosing h no matter what the outcome of the lottery is. What motivates this assumption is that costs of home care are often in the form of foregone earnings (e.g., the mother does not work to take care of the child) or foregone years of education (e.g., younger siblings take care of the child and assign less hours to studying). Similarly, the cost of choosing s does not depend on the outcome of the lottery or the distance to l. Hence, $\frac{\partial V_s(1,Z_2)}{\partial Z_2} = \frac{\partial V_s(0,Z_2)}{\partial Z_2} = 0$. In contrast, $\frac{\partial V_l(0,z_2)}{\partial z_2} > \frac{\partial V_l(1,z_2)}{\partial z_2}$, which means that increases in the distance to l further increase the costs when not winning the lottery. Table 2.1 summarizes changes in the latent utility as a result of changes in the instruments.

Table 2.1: Changes in the latent utility of each childcare option evaluated at different realizations of the instruments, Z_1, Z_2

d	$U_d(0, Z_2) - U_d(1, Z_2)$	$\frac{\partial U_d(Z_1,Z_2)}{\partial Z_2}$	$\frac{\partial U_d(0,Z_2)}{\partial Z_2} - \frac{\partial U_d(1,Z_2)}{\partial Z_2}$
h	0	0	0
s	0	0	0
l	< 0	≤ 0	≤ 0

The sign of the cross-derivatives for U_l reflects that reductions in the cost of choosing l from winning the lottery are higher when Z_2 is small. In this sense, winning the lottery makes distance to l more salient (l "stands out" more than h and s when $z_1 = 1$ and z_2 is small). One justification for this assumption is that, when winning the lottery, parents were more likely to receive information about l, but also the context of winning could inflate the relative utility of l.[26] The implication

[25] If Z_1 was continuous, this would require analyzing the cross derivative of costs w.r.t Z_1 and Z_2. Given that Z_1 is binary in my setting I compare the derivatives of the costs w.r.t Z_2 evaluated at both values of Z_1.

[26] The concepts of salience and of giving different weights to different attributes of the options

of the combined incentives of the lottery and the distance to l is that some complier groups could disappear once distance to l is relatively large.

The case of childcare choice also has distinct features that justify the assumptions on the sign of the cross-derivative of U_l. The sample consists of low income households, and parents of this income segment can place a high value on convenience of the childcare center. One example is how much time it would take parents to bring their child to the center, or the distance from home to the center. If the center is near to home, parents could easily walk with their children to the center. Once it gets too far, parents would either have to walk or to take a vehicle (e.g., bus or car) to the center. For low income families the latter would represent an added cost. Hence, while winning the lottery is expected to reduce the cost of choosing l for some parents, this reduction would be smaller, or negligible, if the center is too far from their home.

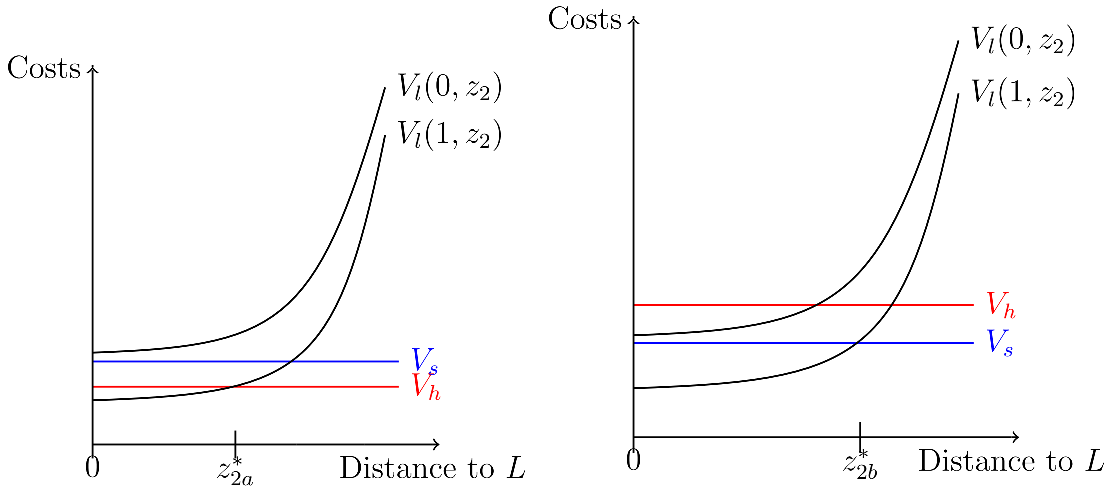
To fully capture the assumptions on costs, I assume that $\frac{\partial^2 V_l(Z_1,Z_2}{\partial {Z_2}^2} \geq 0$ such that the costs of choosing l are convex and increasing in Z_2. Figure 2.2 illustrates two cases of cost structures that imply changes along different childcare margins. Panel (a) displays the case of costs that would induce parents to switch between h and l, when distance to l is below a threshold z^*_{2a}. Above this threshold, parents with this cost structure would behave as h-always-takers. Panel (b) describes the case of costs that would induce parents to switch along the $s-l$ margin, below a threshold z^*_{2b}. Above this threshold, given that V_s is the lower cost, they would

that I use here to justify my assumptions are general interpretations of the model in Bordalo et al. (2013).

behave as s-always-takers.

Figure 2.2: Relative costs that induce parents along the $h-l$ margin, in Panel (a), and $s-l$ margin, in Panel (b)

(a) Lower costs of home care *versus* s (b) Lower costs of s *versus* home care



The restrictions on costs, while untestable with the available data, provide a framework to define potential responses as the instruments change. In addition, the analysis of costs also shows how different realizations of Z_2 could capture distinct complier groups. This in turn highlights the rationale for using the variation provided by the lottery outcome while conditioning on the second instrument, Z_2. In the next section, I present an exploratory analysis with a twofold objective. First, to determine whether there is empirical evidence of joint effects on choices from the lottery outcome and distance to l. Second, to inform the definition of thresholds of distance to l at which the set of complier groups might defer.

2.3.5 Exploratory Analysis

Building on the analysis of parental utility and costs of childcare, I estimate the probability of choosing l, s, and h on the lottery outcome, distance to l, and baseline variables. Moreover, I include the interaction between the lottery outcome and distance to the nearest l center. The role of the interaction between the IVs is twofold. First, it allows me to test the hypothesis of joint incentives of the lottery outcome and distance to the nearest l center. Second, it exactly identifies the model of choices and instruments (i.e., there are three available childcare alternatives and the variation stems from three exogenous variables). Thus, I estimate the following model:

$$D_d = \beta_0^d + \beta_1^d Z_1 + \beta_2^d Z_2 + \beta_{1,2}^d Z_1 \times Z_2 + \beta_X^d X + \epsilon^d, \quad \text{for } d \in \{h, s, l\}$$

Table 2.6 shows that, on average, winning the lottery increases the probability of choosing l by 18 percentage points, and decreases the probability of choosing s in 16 percentage points. These results are in line with the assumptions on the costs of s and l at different lottery outcomes (e.g., winning the lottery decreases the relative costs of choosing l and increases the likelihood of choosing l). That is, due to the variation in the lottery outcome some parents substitute childcare at s for childcare at l. Similarly, the effect of an increase in the distance to l has a sizable effect in shifting parents towards s and away from l. As the distance to l increases the probability of choosing l decreases by almost 28 percentage points, whereas the probability of choosing s increases in about 32 percentage points.

My findings of distance to l as a strong predictor of the probability of choosing l are in line with previous findings for s centers. Attanasio et al. (2013) and Bernal and Fernández (2013) highlight the role of distance to s as one of the determinants of enrollment in s. Specifically, Attanasio et al. (2013) find that participating in s had a positive effect on nutritional outcomes. The authors use an IV strategy with cost-shifters such as distance to s, fees, and local availability of s. They find evidence that high fees and long distances are some of the reasons why children might not attend s. On average, the travel time to s in their sample is of 22 minutes, while the median travel time is of 10 minutes.

The choice of home care, h, is less responsive to the lottery outcome and to the distance to l, on average, than choosing s or l. Solely winning the lottery, or having an l center near, does not directly affect the probability of choosing home care. Both of these results point out that while s and l could be perceived as childcare substitutes by parents, taking care of children at home responds to factors other than the relative cost. If parents who choose h have tighter constraints, preventing them from enrolling their children in public care, they would need more substantial incentives to substitute h for s or l. For instance, winning the lottery on its own could not have an effect on substituting home care for l, if l is too far.

In terms of the probability of choosing h, an increase in the distance to l is seven percentage points higher for those who win the lottery than for those who did not win (see Table 2.6). The positive sign of the interacted term opposes the assumption that winning the lottery would, on average, induce parents away from

h and into choosing l. It seems that at least for some parents winning the lottery once the center is too far from home increases the likelihood of choosing h over l or s. However, I expect to observe this behavior only after some distance to l that is large enough to discourage parents from choosing l.

To further illustrate how the incentives provided by the lottery depend on how close or far l is, I estimate the probability of choosing each childcare choice conditional on the lottery outcome at different points in distance to l. The first row of Figure 2.5 shows the probability of each childcare choice when winning the lottery (the green line) and when parents do not win (the black line), evaluated at different values of distance to l. The second row plots the difference between these probabilities (i.e., the marginal effect of the lottery). Each column refers to a childcare choice (l, s, and h). I observe that the probability of choosing l is always higher for those who win the lottery, than for those who did not win. As a result, the marginal effect of the lottery outcome is positive at all values of the distance. For the probability of choosing s, I observe the opposite pattern, such that winning the lottery is associated with a higher probability of choosing s than not winning the lottery. This result translates into a negative marginal effect, which becomes somewhat larger when l is relatively far. These results are consistent with my assumptions on the costs of choosing s and l and how those translate into the probability of choosing these options of care.

The probability of choosing h conditional on the lottery outcome differs in many ways from the pattern I observe for the probabilities of choosing s and l. While the marginal effects of the lottery for l and s are relatively constant as distance to l increases, I observe an increasing effect of the lottery outcome on the probability of choosing h (column 3, Figure 2.5). Moreover, the sign of the effect changes after l is approximately 0.5km away from home. In particular, winning the lottery is associated with a higher probability of choosing h for higher distances to l, and a lower probability of choosing h when l is relatively near. In addition, the marginal effect of the lottery on the probability of choosing h is only statistically significant when l is at a higher distance than 0.75km. These patterns can be better explained by the literature on parental choices and investments which suggests that parents can give weight to different attributes of the available choices,[27] and base their investment decisions in what they observe for children of similar age and backgrounds as their own children. In Appendix A.4, I present an alternative cost structure consistent with convexity and observed parental behavior along the $s-h$ margin. It requires additional assumptions that exploit the fact that children in the experiment were initially at s. In addition, although the care alternatives are unordered their costs can be ranked. I present alternative cost structures and subjective parental preferences as two potential explanations of the behavior along the $s-h$ margin. Given that I lack detailed information to determine the drivers of switching along the $s-h$ margin, I cannot rule out either argument.

Although proximity has been recognized in the literature as an important determinant of parental choice of childcare in developing countries (Attanasio et al., 2013, Hojman and López Bóo, 2019), its interaction with other variables has received

[27]For example, departures from standard optimization models to incorporate subjective parental beliefs and perceptions can be found in Attanasio, Cunha, and Jervis (2019); Cunha (2015); Cunha, Elo, and Culhane (2013) and Wang, Puentes, Behrman, and Cunha (2020).

less attention. I further exploit the variation of distance to l by defining a threshold at which the incentives of the lottery outcome change. The definition of the threshold is motivated by the analysis of the probability of choosing h conditional on the lottery outcome as seen in column 3, Figure 2.5. As noted above, after a distance of about 0.5km the probability of choosing h is higher conditional on winning the lottery *versus* not winning it.

Why would parents respond differently to winning the lottery depending on how far l is? There are several characteristics of childcare that can explain these patterns. Consider the case of a parent who has a center near enough to walk with her children every morning, compared to a parent who would have to travel a longer distance to bring her children to the center and pick her up in the afternoon. The latter could discourage many parents, particularly from low income levels, from enrolling their children in a center, and could induce them to select home care. As shown in Section 2.3.4, the cost of choosing a center could be flat for shorter distances (such that parents can walk with their children) but would increase rapidly after a threshold (such that parents would have to travel by bus or car with their kids).

I assume that winning the lottery when l is near, below the threshold, implies counterfactual choices that differ from those that parents would make when l is far, above the threshold. These assumptions stem from the exploratory analysis and the assumptions on convex costs of he large centers. I formalize the counterfactual choices in the following section.

2.3.6 Conditional Rules of Parental Behavior

I formalize potential parental behavior following the assumptions on how child-care choices would change as a result of changes in costs (through changes in distance to l and the lottery outcome) as well as the patterns of the conditional probabilities at different values of the distance. To start, I assume that the lottery induces parents along three margins: $h \rightarrow l$, $s \rightarrow l$, and $s \rightarrow h$. The first two follow directly from the analysis of relative costs and utilities of h and s *versus* l. The latter stems from observing a higher average probability of choosing h when winning the lottery *versus* not winning, for relatively high values of distance to l. All together, these patterns of choice imply that parents can be induced towards and away from h. That is, $h \rightarrow l$ implies that parents are induced away from h, while $s \rightarrow h$ implies that parents are induced towards h. This is a violation of monotonicity (Imbens and Angrist, 1994) which requires that there are no two-way flows from changes in the instrument (the lottery outcome, in this case).

One important observation from the exploratory analysis is that below a distance to l of about 0.75km, the differences in the probability of choosing home care when winning *versus* not winning the lottery are negligible. As a result, I expect that the share of $h - l$ compliers is low or almost zero. I define an interval above 0.75km, where the conditional rules of behavior change. Above a distance to l of $0.75 - 0.85$km, I assume that parents choosing h are either h-always-takers or $s - h$ compliers. As a result, conditional on Z_2 the responses induced by the lottery outcome, Z_1, satisfy monotonicity (that is, the responses satisfy conditional/partial monotonicity in *A2*).

Table 2.2: Z_1-responses at different evaluation points of Z_2

Behavior below z_2^*						Behavior above z_2^*					
Z_1	g_1^-	g_2^-	g_3^-	g_4^-	g_5^-	Z_1	g_1^+	g_2^+	g_3^+	g_4^+	g_5^+
0	h	s	l	h	s	0	h	s	l	s	s
1	h	s	l	l	l	1	h	s	l	h	l

The table above presents the different responses induced by variation in the lottery outcome, above and below threshold z_2^*. g_1, g_2, and g_3 are h, s, and l always-takers at either side of the threshold. They don't provide any variation to estimate treatment effects. In turn, g_4^- are $h-l$ compliers and g_5^- are $s-l$ compliers below the threshold z_2^*. Meanwhile, g_4^+ are $s-h$ compliers and g_5^+ are $s-l$ compliers above the threshold z_2^*. These responses imply matrices $B_d(Z_2 < z_2^*)$ and $B_d(Z_2 > z_2^*)$, defined in section 2.2.3, which are binary matrices that take the value of one every time childcare choice d is chosen. Choice specific matrices define $B(Z_2 < z_2^*)$ and $B(Z_2 > z_2^*)$,

$$B(Z_2 < z_2^*) = \begin{vmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{vmatrix}, \qquad B(Z_2 > z_2^*) = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{vmatrix}.$$

I employ these matrices to estimate the probabilities of responses g^- and g^+, their average baseline variables, and counterfactuals following the estimation strategy in Section 2.2.3. For z_2^*, I use an interval of distance to l between $0.75 - 0.85$km.

2.3.7 Implementation

One of the main components of the estimation of complier shares and counterfactuals is the interaction between the lottery outcome and proximity to l. I employ a parametric approach that allows me to implement the conditional restrictions given the dimension of my sample.[28] I estimate propensity scores with the following regression model:

$$D_d = \beta_0^d + \beta_1^d Z_1 + \beta_2^d Z_2 + \beta_{12}^d Z_1 * Z_2 + \beta_X^d \mathbf{X} + \epsilon_d,$$

where $\mathbf{X}$ denotes baseline variables and ϵ_d is an unobserved error component. Baseline variables included in $\mathbf{X}$ are child's age in months, sex, a binary variable to indicate lowest income level, years of education of the mother, number of children 0-5 years of age living at home, binary variables for birth order, cognitive, socio-emotional, and nutritional development. I estimate the equation for each choice separately with a Linear Probability Model. I evaluate the propensity scores for winning *versus* not winning the lottery, at increments of 0.1km in the distance to l.

[28] Generally, these objects can be estimated nonparametrically (or semiparametrically) without imposing restrictions on the functional form. While a local approach could better approximate the relation between the instruments, the data requirements are higher. It is also worth noting that a fully nonparametric approach (such that the instruments and baseline variables enter the decision model nonlinearly) suffers from the curse of dimensionality and would threathen the convexity restriction.

I employ equation (A.1) to estimate the probabilities of always-takers and compliers in Table 2.2.

To estimate the average of baseline variables for each response group and their counterfactuals I estimate choice d-covariates interactions and choice d-outcomes interactions on the instruments and their interactions. For each vector in $Y_{\text{outcome}} \in \{YD_H, YD_S, YD_L\}$, I estimate the following reduced-form equation:

$$Y_{\text{outcome}} = \beta_0^Y + \beta_1^Y Z_1 + \beta_2^Y Z_2 + \beta_{12}^Y Z_1 * Z_2 + \beta_X^Y X + \epsilon_Y,$$

and I compute the counterfactuals for always-takers and compliers with equation (2.4). Similarly, for each vector in $X_{\text{baseline}} \in \{XD_H, XD_S, XD_L\}$ with $X \in \mathbf{X}$, I estimate the following regression:

$$X_{\text{baseline}} = \beta_0^X + \beta_1^X Z_1 + \beta_2^X Z_2 + \beta_{12}^X Z_1 * Z_2 + \epsilon_X,$$

and I compute the average of baseline variables with equation (A.3) in the Appendix A.2. All the regressions include city fixed effects. For inference, I use bootstrapped standard errors.[29]

[29] An important caveat in terms of the bootstrap standard errors is that I lack data to identify small centers in my sample, which is the level at which the lottery was assigned. Given this data limitation, I cannot compute bootstrap standard errors with clusters at the small center level.

2.4 Results

Figure 2.6 shows the shares of always-takers and compliers that result from using the variation in the lottery outcome, without conditioning on the distance to the nearest large center. Almost 18% of children in the sample correspond to $s-l$ compliers, and this fraction is statistically significant. To estimate this fraction, I employ the combinations of counterfactual choices that result from the cost analysis in Section 2.3.3. These combinations imply a response matrix with two complier groups: along the $s-l$ and $h-l$ margins. However, as I discuss in Section 2.3.2, there is evidence that some parents reacted to winning the lottery outcome by choosing home care. The latter contradicts the assumption that parents would only move away from h (and into l) as the lottery outcome varies. Overall, without conditioning on the distance to l, the variation in the lottery outcome would imply a negative share of responses along the $h-l$ margin.

Table 2.7 show the local effects for compliers along the $s-l$ margin, without conditioning on distance to the nearest l center. These results suggest negative effects on cognitive development, and positive effects on socio-emotional and nutritional dimensions. Importantly, they are biased as a result of uncontrolled variation from parents switching into and away from home care. Appendix A.3 shows what IV identifies by attributing all the variation in the lottery outcome to the $s-l$ margin. Two additional effects bias the effects in Table 2.7: the effect for parents switching along the $h-l$ margin, and the effect from parents switching along the $s-h$ margin.

2.4.1 Complier shares

In what follows, I estimate complier shares and their counterfactuals using the conditional rules of parental behavior in Section 2.3.6. Figure 2.7 shows the average prevalence of always-takers and $s - l$ compliers induced by the variation in the lottery outcome, above and below distance to the nearest l. When distance to the nearest l is below 0.75km (Panel (a)), the group with the largest share, of 52%, corresponds to l-always-takers or parents who would have chosen l no matter the outcome of the lottery. In other words, a high fraction of parents would have chosen l regardless of winning the lottery, given that the relative costs of choosing l are low when l is at a small distance from home. Meanwhile, the fraction of parents who are induced away from s and into choosing l due to the variation in the lottery outcome (i.e., compliers along $s - l$) is about 18% and it is statistically significant.

The share of compliers when l is relatively far from the child's home largely differ from the findings described above. Panel (b) of Figure 2.7 shows the prevalence of always-takers and compliers induced by the variation in the lottery outcome, when distance to l is above 0.85km. The fraction of parents that would choose to enroll in h, and away from choosing s, when winning the lottery is 7.3%. In turn, the fraction of compliers along the $s - l$ margin is almost twice as high, at 15.5%. Both shares are statistically significant. Overall, changes in the outcome of the lottery induce about 22.8% of parents to change their childcare choices, when distance to l is above 0.85km. The remaining fraction corresponds to h-always-takers (10%), s-always-takers (39.5%) and l-always-takers (27.8%). The higher share of s-always-takers *versus* l-always-takers is in line with the assumption that (some) parents substitute l with s as the former becomes more costly (i.e., when the distance to l is large).

The fraction of $s-l$ compliers is relatively constant for values of the distance to l below 0.75km. In contrast, Panel (a) in Figure 2.8 shows that, as distance to l increases, the fraction of l-always-takers decreases while the fraction of s-always-takers increases. This is consistent with parents substituting their childcare choices towards the relatively cheaper option; in this case, as distance to l increases it becomes more costly in relative terms to bring children to l. For a distance to l above 0.85km, panel (b) of Figure 2.8 shows that the share of s-always-takers increases, while the share of l-always-takers decreases, as distance to l becomes larger.

As the cost of taking children to l increases (i.e., a higher distance to l), compliers along the $s-h$ margin increase while compliers along the $s-l$ margin decrease (panel (b), Figure 2.9). In other words, when the cost of taking the kids to l is too high, more parents would rather care for their children at home or at an s center, rather than at l. Despite the increase in the share of $s-h$ compliers, and the sizable share of $s-l$ compliers, the compliers' shares lose significance as distance to l increases. This follows from observing the confidence intervals in Panel (b) of Figure 2.9, which become larger the further away l is. Meanwhile, panel (a) shows that the share of compliers is almost flat when the l center is nearby but more precisely estimated as distance to l increases.

2.4.2 Average Baseline Characteristics

Compliers along the $s-l$ margin at a distance to l less than 0.75km are more likely to be older males with lower average cognitive development and higher socio-emotional development at baseline (see Table 2.8). They are also closer to l, on average, than always-takers. In terms of age, compliers are almost 5-6 months older than s-always-takers and l-always-takers, respectively. About 64% of compliers are male children. This fraction is the highest across all groups. Less than half of compliers belong to households in the poorest income levels, while this share is about 53% for h-always-takers. More important, there are stark differences between compliers and always-takers in their baseline development. Compliers are, on average, worst in all dimensions of development. The only exception is average nutritional development, for which home-always-takers are below all groups. In turn, compliers along the $s-l$ margin are well below average baseline development in both cognitive and socio-emotional dimensions.

The majority of compliers along the $s-h$ and $s-l$ margin are relatively younger, female children. Table 2.9 shows the average of baseline characteristics for always-takers and compliers induced by the lottery outcome, when l is further than 0.85km. Overall, the differences in average age across the different groups is not large. Meanwhile, $s-l$ compliers have more educated mothers (by at least five years) than any other group and have the lowest socio-emotional baseline development, on average. One striking difference is that almost the totality (87.7%) of children in the complier groups are girls, while the majority of always-takers are male children.

Similarly, the prevalence of girls is considerably large compared to compliers who exhibit the same behavior, $s - l$, but are below 0.75km in distance to l (Table 2.8). This result stresses that even if compliers along $s - l$ are prevalent above and below 0.75km, these groups may differ in their baseline characteristics and can also differ in terms of their unobservables.

On average, parents who respond to the lottery outcome along the $s - h$ margin, when distance to l is large, have children with the lowest baseline levels of nutrition (Table 2.9). In turn, average cognitive development is the lowest for children of complier parents along the $s - l$ margin. Although I have limited information to determine the mechanisms that induce parents along either margin, these differences in development at baseline could explain the sorting patterns. At an early age (0-5 years), assessing children development is a complex task and parents can form their perceptions based on what they observe for children of a similar age and context (Mulcahy and Savage, 2016). For instance, Wang et al. (2020) present a reference-dependent utility model of parental feeding practices, which is based on evidence that shows how perceptions of normal height and weight can stem from comparisons with other children in the family, or with similar backgrounds. Moreover, parental concern for their children development differs if parents can observe their child performing some tasks (e.g., has trouble with communication but learned to walk on time), *versus* struggling in many different areas (Mulcahy and Savage, 2016).

In light of the insights from the literature regarding parental perceptions and practices, I argue that parents moving along the $s - h$ and $s - l$ margin do so as a response to the relative development of their children at baseline *versus* the average development at s. To start, overall baseline development of children in the $s-h$ complier group is well below the average in terms of nutritional development. Meanwhile, children among the $s-l$ complier group were initially below average in all dimensions of development. This group also has more educated mothers who, on average, could be more concerned with finding alternative forms of care for their children besides s or h. It is also worth noting that the average nutritional development of s-always-takers is above the mean, and way above that of $s-h$ compliers. Thus, below average development and maternal levels of education could account, at least in part, for the behavior of compliers in my sample.

2.4.3 Counterfactuals and Treatment Effects

Given the baseline development levels of compliers, parents who chose to transfer their children from s to l may do so in the hopes of improving their development. To determine if compliers benefit, or not, from choosing l instead of s I estimate their average counterfactuals and local treatment effects for cognitive, socio-emotional, and nutritional development at follow-up. The top row of Figure 2.10 presents average counterfactuals for cognitive, socio-emotional, and nutritional development. The black line represents the development level that children would have attained if they enroll in s. The blue line refers to the development level children would have attained from choosing l.

For cognitive and socio-emotional development, I find that children do worst by attending l than they would have done so by choosing s. This holds for all values of distance below 0.85km. This pattern translates into an upward slopping, but negative, LATE for cognitive development and a downward slopping, and negative, LATE for socio-emotional development (row 2, panel (a) and (b) in Figure 2.10). That is, the difference in the counterfactuals for cognitive development tends to narrow as distance to l increases, while for socio-emotional development it becomes larger. Importantly, these differences are statistically significant at the 10% level for cognitive development.

In terms of nutritional development, the counterfactuals of choosing s and l for compliers decrease as distance to l increases. However, this decrease is faster, or steeper, if children would have chosen to enroll in s than in l. This translates into a modest increase in LATE as distance to l increases, such that for distance to l above 0.45km approximately the effect is positive although it is not statistically significant. Overall, for parents relatively near l, there is little evidence that their children would be better-off by choosing l rather than s. As noted above, compliers are older on average and have relatively low levels of cognitive and nutritional development at baseline, which could explain why there is virtually no effect on those dimensions from choosing l instead of s.

The top row of Figure 2.11 presents average counterfactuals for cognitive, socio-emotional, and nutritional development. The black line represents the development level that children would have attained if they enroll in s. Importantly, this counterfactual is a weighted average of what each complier group would have experienced from choosing s. It cannot be identified separately for each complier group without additional assumptions. The blue line refers to the development level compliers along the $s - l$ margin would have attained from choosing l. The red line refers to the development level compliers along the $s - h$ margin would have attained from choosing h. Column (a) in Figure 2.11 shows that compliers along the $s - l$ margin do worst in cognitive development from choosing l, at low levels of distance to l. In turn, $s - h$ compliers would experience the lowest level of cognitive development from choosing home care. Column (b) shows the opposite pattern, but it translates into higher behavioral problems (i.e., a higher socio-emotional score signals behavioral issues). The pattern is also consistent in column (c), which shows that in nutritional development $s - h$ compliers do worst by choosing h.

The second row of Figure 2.11 presents conditional LATEs of s centers *versus* the next-best childcare alternative. This effect combines the LATE for compliers along the $s - l$ margin and compliers along the $s - h$ margin. The bottom left panel shows positive, although not statistically significant, effects on cognitive development of s centers *versus* the next-best. The next panel in the second row shows positive effects of s centers *versus* the next-best on socio-emotional development. This positive effect translates into more behavioral problems, on average, for those at s centers *versus* the next-best. On their own, the next-best results can be uninformative and mask considerable differences in the effects between the two distinct complier groups. For instance, the counterfactuals for socio-emotional development suggest that compliers along the $s - l$ margin could be driving the positive next-best effects. Meanwhile, compliers along the $s - h$ margin could experiment far worst effects on socio-emotional development.

One approach to separately identify the LATEs for each complier group is to assume that their potential development at s centers would have been the same, on average. This refers to the homogeneity assumption described in Section 2.2.3. Formally, for the case of parental choice of childcare, this assumption states that:

$$E[Y_s|\text{Complier:}s-l, Z_2 > z_2^*] = E[Y_s|\text{Complier:}s-h, Z_2 > z_2^*]$$

This assumption has been used previously by Kline and Walters (2016) and Lee and Salanié (2020). One important drawback is that these complier groups can be systematically different making the homogeneity assumption less likely to hold. Hence, prior to implementing homogeneity in my sample, I analyze the average baseline characteristics of the complier groups above and below the threshold of distance to l. Figure 2.12 shows the average of selected baseline characteristics for compliers along the $s-l$ margin and $s-h$ margin. The grey bar denotes the threshold where the parental rules of behavior change discontinuously. Overall, compliers along each margin differ in their baseline characteristics (i.e., there are considerable differences between the orange and blue lines depicted above the threshold). Given that these differences can threaten the validity of the homogeneity assumption, I control for baseline variables in the estimation of counterfactuals.

Figure 2.12 shows modest differences in the average of baseline variables, for compliers along the $s-l$ margin above and below the threshold where the parental rules of behavior change discontinuously (i.e., the grey bar). This supports the assumption that $s-l$ compliers above and below the threshold only differ in their proximity to l, but are on average similar in their characteristics and treatment effects. While I observe a modest jump in baseline development, I argue that these differences are not considerable to infer that, at the threshold, $s-l$ compliers are distinct groups. Hence, treatment effects for $s-l$ compliers should be smooth above and below the threshold.

Figure 2.13 presents treatment effects for each complier group. Above the threshold, I employ the homogeneity assumption to separately identify LATEs for $s-l$ and $s-h$ compliers. To start, the LATEs for $s-l$ compliers do not display discontinuous jumps at the threshold. The LATE of cognitive development along the $l-s$ margin is negative but upward slopping. In contrast, the LATE for the $h-s$ margin in terms of cognitive development is relatively constant and negative. Column (b) in Figure 2.13 suggest a lower prevalence of behavioral problems, on average, for children along the l centers *versus* s centers margin. In turn, the results suggest a higher prevalence of behavioral problems for children along the h *versus* s margin. Overall, while some children might benefit from switching between s and l centers, the results indicate no benefit for children switching away from s centers and into home care. The results along the latter margin lack predictive power; nonetheless they are indicative of sizable differences in the effects of choosing alternative childcare choices.

Table 2.10 summarizes the results for LATE across dimensions of children development, for the l centers *versus* s centers margin. In contrast to the results in Table 2.7, without conditioning on distance to the nearest l center, I uncover heterogeneous results that suggest that some children might benefit in terms of cognitive development. This is the case for children above the distance threshold switching away from s and into l centers. These children have more educated mothers and are younger, on average. In contrast, I find negative effects on cognitive development for children along the $s-l$ margin who are below the distance threshold. This effect has the same sign as the average unconditional effect in Table 2.7, but is larger in magnitude.

2.4.4 Robustness check: sensitivity to the distance threshold

One of the key components to estimate LATEs with *conditional vectors* is the definition of the thresholds where the rules of behavior would change. In the main specification, I use an interval of the distance to the large centers rather than a specific realization such that parental behavior would change at $0.75 - 0.85$km. In this Section, I evaluate the conditional LATEs when the rules of behavior change discontinuously at the extreme values in that interval: (i) 0.75km, and (ii) 0.85km.

Table 2.11 shows the results of estimating the conditional LATEs with different distance thresholds. First, when distance to l is below the threshold, the negative LATE on cognitive development of large *versus* small centers remains relatively the same in size (0.72 at the lowest threshold of 0.75km, and 0.7 at the highest point of 0.85km). The significance and direction of the LATEs is also the same across thresholds, conditional on distance to l below the threshold.

In turn, the LATE of large *versus* small centers for distance to l above the threshold does seem to depend on its definition. When the threshold is of 0.75km (and distance to l is above this threshold), the effect on cognitive development of large *versus* small centers is 0.8 standard deviations. This LATE is almost half as the estimated effect when the threshold is 0.85km. This difference might be a result of assuming that the responses of $s-l$ and $s-h$ are both prevalent above 0.75km, when it might be the case that responses along $s-h$ can only be uncovered for a higher distance to l centers. One argument that might support this explanation is that the effect of the lottery on choosing home-care is not significant and negligible between 0.75-0.85km. Regardless, across thresholds the LATE on cognitive development remains not statistically significant. Last, it is also worth noting that for socio-emotional and nutritional development the differences in the LATEs are smaller across the different thresholds.

2.5 Concluding Remarks

In this chapter, I study the identification and estimation of treatment effects in contexts where agents can choose between multiple options, and the researcher has access to multiple instrumental variables that can affect the same options. I focus on settings with two main features. First, agents can have preferences over the alternatives and self-select into their preferred choice. The latter motivates employing IVs as an estimation strategy. Second, the options are unordered, meaning that I cannot say that one option is better than the other, *a priori*. As a result, agents can have different fallback options and derive different gains or losses from their choices.

The IV approach has limitations. It estimates local effects for agents who change their behavior as the IV changes (i.e., compliers). With multiple unordered choices, compliers are heterogeneous, and agents can switch along many margins. For instance, if a large childcare center opens close to home, parents could switch between home and large centers, small and large centers, or home and small centers. The set of compliers becomes more complex when multiple instruments are available. For example, in my application of childcare choices I show that, despite winning the lottery, some parents might be discouraged from choosing large centers if the distance from home is relatively large. While this type of joint response to multiple instruments has been recognized in the literature, current methods implicitly assume that the behavior of compliers (or responses) to one instrument is the same across the distribution of other instruments. However, the case of multiple instruments that affect the same choices is relevant for many empirical applications.

My main contribution is to present an empirical strategy that addresses some of these challenges. First, I account for the joint effects of multiple instruments on the probability of choosing an option. That is, I allow for the response to the variation in one instrument (for example, an offer of a slot at a center) to differ depending on other instruments (for example, proximity to the center) that affect the same choice. To do so, I employ a latent utility framework and model responses to the instruments through their effect on the costs of each option. I impose restrictions on the shape of the cost function. In particular, I assume that the cost function is convex. The latter allows me to define conditional rules of behavior that satisfy monotonicity, locally.

I apply these tools to estimate the impact of childcare choice in Colombia on children's development. In this setting, parents can choose between home care, small centers (which serve between 12-15 children), or large centers (which serve 25 children per teacher). My sample consists of low-income families, and care at small and large centers is publicly provided. My findings suggest that the choice of small centers *versus* large centers follows the logic of substitutes of care. I find that $15 - 18\%$ of parents substitute care at small centers for care at large centers when the latter is relatively cheaper. In contrast, substituting home care for large centers requires stronger incentives. In particular, winning the lottery only induces parents to substitute home care for large centers when those are nearby. In turn, for a relatively large distance to large centers, winning the lottery seems to drive parents towards home care. While my results lack predictive power, they are indicative of the importance of uncovering variation along different margins of choice in the estimation of multivalued treatment effects.

2.6 Tables and Figures

Table 2.3: Summary Statistics

Variable	Obs	Mean	SD	Min	Max
Age (months)	1246	36.814	9.615	13	55
Male	1246	0.512	0.500	0	1
Children 0-5 yoa at home	1246	1.528	0.709	1	5
Mother's years of education	1246	8.740	2.985	0	18
Low income household	1246	0.436	0.496	0	1
Type of childcare					
Home care (h)	1246	0.128	0.334	0	1
Small centers (s)	1246	0.323	0.468	0	1
Large centers (l)	1246	0.549	0.498	0	1

Source: subsample from Bernal et al. (2019)
Note: The sample consists of children who were initially at small centers. All socioeconomic variables were collected at baseline.

Table 2.4: Average characteristics, by type of care

Variable	Large centers (l)	Small centers (s)	Home care (h)
Age (months)	36.980	36.241	37.553
Male	0.515	0.526	0.465
Children 0-5 at home	1.477	1.536	1.730
Mother's years of education	8.986	8.697	7.791
Low income household	0.409	0.442	0.535
Distance to small center (s), in km	0.299	0.331	0.290
Children development at baseline			
Cognitive	0.027	-0.094	0.121
Socio-emotional	-0.022	0.091	-0.135
Nutritional	0.035	0.068	-0.323
Outcomes [Obs.]			
Cognitive [1231]	0.166 [674]	-0.021 [401]	-0.662 [156]
Socio-emotional [1238]	-0.064 [678]	-0.035 [403]	0.368 [157]
Nutritional [927]	0.127 [513]	0.007 [299]	-0.584 [115]
%	54.9	32.34	12.76
N	684	403	159

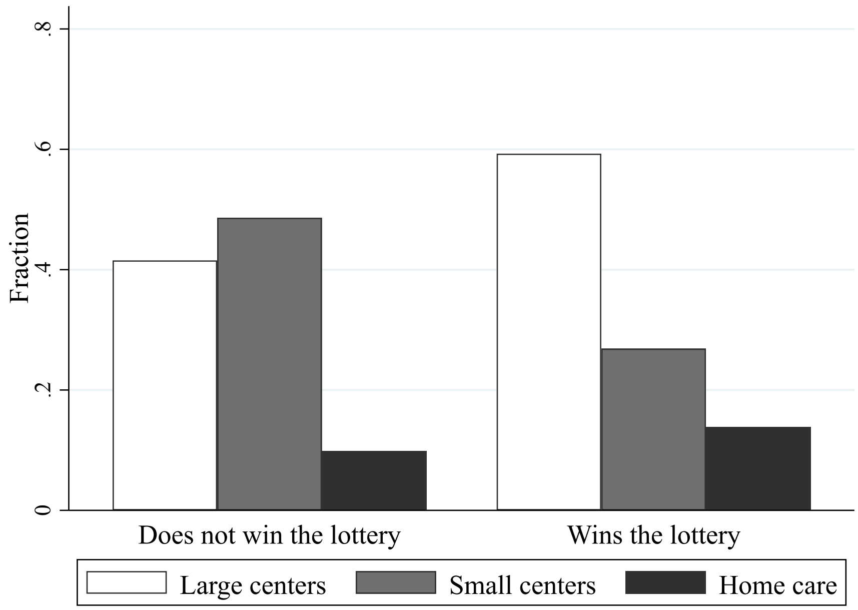
Source: subsample from Bernal et al. (2019)
Note: The sample consists of children who were initially at small centers. All socioeconomic variables were collected at baseline. Cognitive development is a composite of scores from the ASQ. Socio-emotional development is a composite of behavioral components of the ASQ. All scores from the ASQ are age standardized. Nutritional development corresponds to z-scores for weight-for-age, height-for-age, and weight for height.

Table 2.5: Average of instruments, by type of care

Variable	Large centers (l)	Small centers (s)	Home care (h)	All
Wins lottery	0.805	0.615	0.802	0.743
Distance to large center (l), in km	0.668	1.125	0.912	0.848

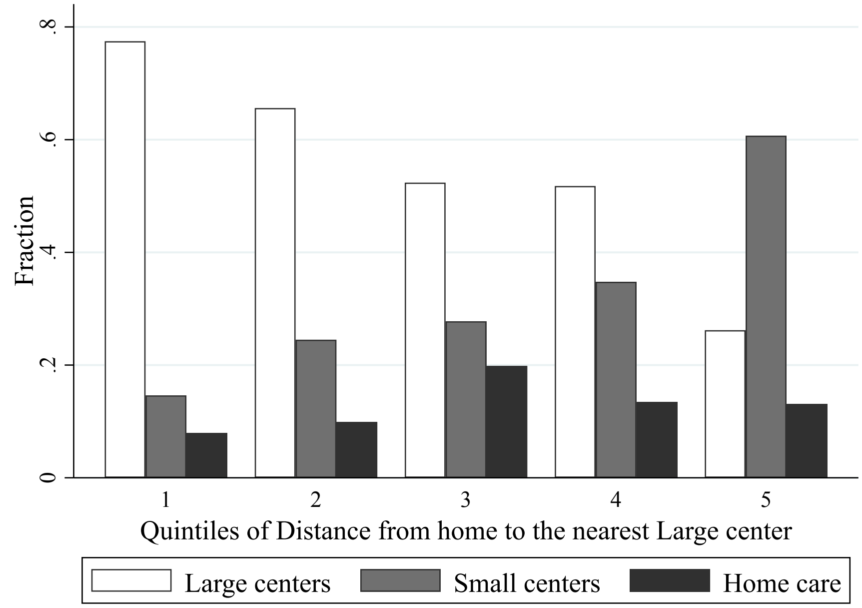
Source: subsample from Bernal et al. (2019)
Note: The sample consists of children who were initially at small centers. Each column represents a childcare choice; the last column presents averages for the full sample for comparison. The row "Wins lottery" refers to the fraction of households that where assigned to received information and the option to transfer to large centers. Distance to large centers is the geographic distance from the child's home to the nearest large center, in km.

Figure 2.3: Type of care selected, by lottery status



Source: subsample from Bernal et al. (2019)
Note: The figure illustrates the distribution of childcare choices by outcome of the lottery. The white bars correspond to large centers, the grey bars to small centers, and the black bars to home care. The bars on the left show the fraction of parents who select each type of care among those who did not win the lottery. The bars on the right show the fraction of parents who select each type of care among lottery winners.

Figure 2.4: Type of care selected, by quintiles of distance to the nearest Large center



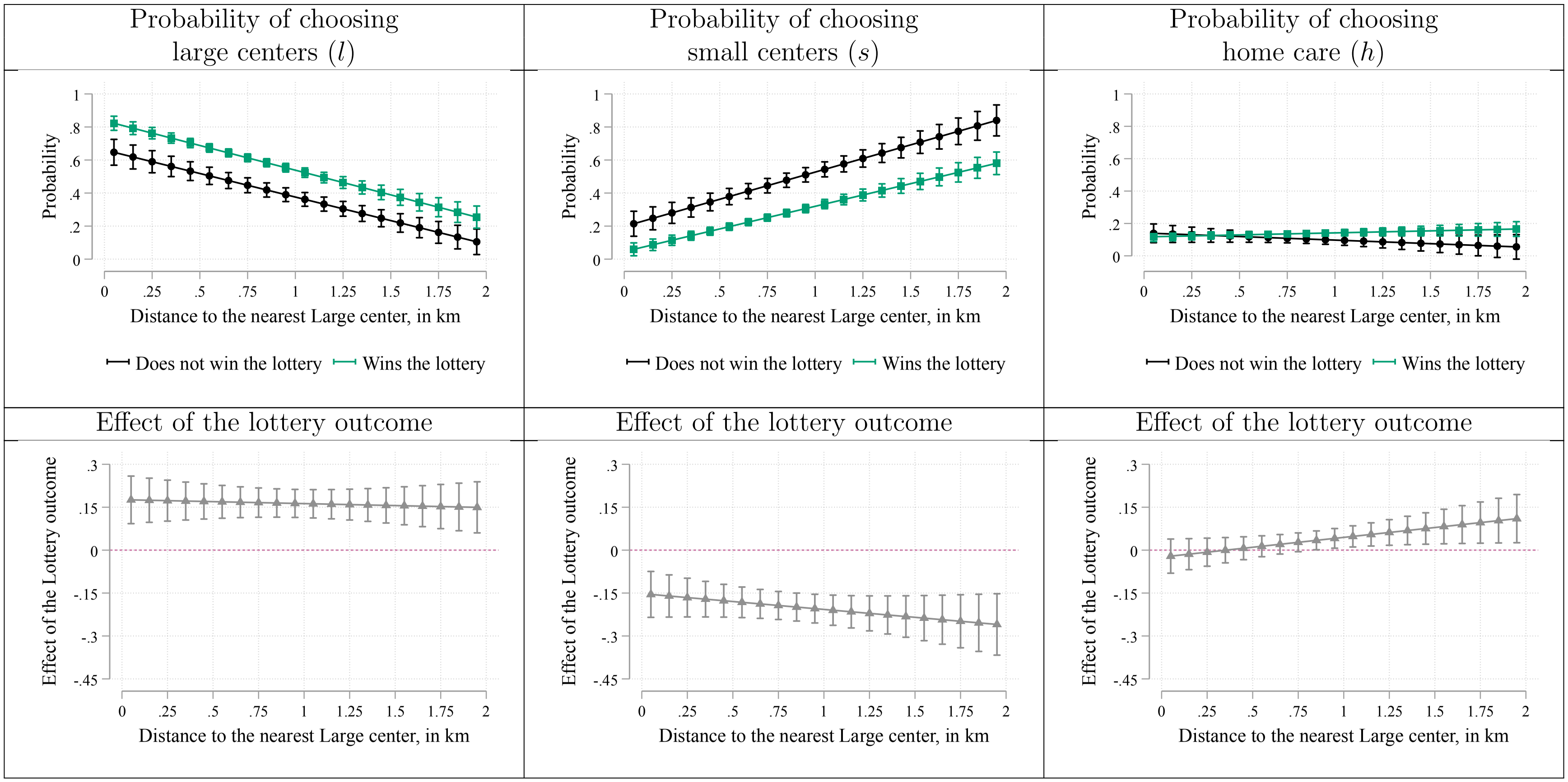
Source: subsample from Bernal et al. (2019)
Note: The figure illustrates the distribution of childcare choices by quintiles of distance to large centers. The white bars correspond to large centers, the grey bars to small centers, and the black bars to home care.

Table 2.6: First-stage results: The effect of winning the lottery varies with distance to l, for the probability of choosing h

Variable	Large centers (l)	Small centers (s)	Home care (h)
Wins Lottery	0.187***	-0.164**	-0.023
	(0.054)	(0.052)	(0.038)
Distance to l, in km	-0.280***	0.322***	-0.041
	(0.043)	(0.048)	(0.038)
Wins Lottery × Distance to l, in km	-0.022	-0.047	0.069*
	(0.045)	(0.050)	(0.041)
Distance to s, in km	0.083**	-0.074**	-0.009
	(0.037)	(0.036)	(0.028)
Age in months	-0.010	0.014	-0.004
	(0.010)	(0.009)	(0.007)
Male	0.022	0.018	-0.040**
	(0.024)	(0.023)	(0.017)
Low income household	-0.006	-0.013	0.018
	(0.027)	(0.026)	(0.019)
Cognitive Development	0.000	-0.009	0.008
	(0.008)	(0.008)	(0.006)
Socio-emotional Development	-0.019*	0.028**	-0.010
	(0.011)	(0.011)	(0.007)
Nutritional Development	-0.004	0.013	-0.009
	(0.009)	(0.008)	(0.006)
Mother's years of education	0.016***	-0.005	-0.010***
	(0.005)	(0.005)	(0.003)
Children 0-5 at home	-0.036	0.000	0.036**
	(0.024)	(0.022)	(0.018)
Birth order=1	0.113	-0.138	0.026
	(0.156)	(0.145)	(0.127)
Birth order=2	0.081	-0.137	0.056
	(0.153)	(0.141)	(0.126)
Constant	0.718**	0.049	0.232
	(0.237)	(0.218)	(0.166)
N	1,238	1,238	1,238
R-squared	0.201	0.187	0.119

Source: subsample from Bernal et al. (2019)
Robust Standard errors are in parentheses. All regressions include city fixed effects. Each choice is estimated separately.
$*p < 0.1, **p < 0.05, ***p < 0.01$

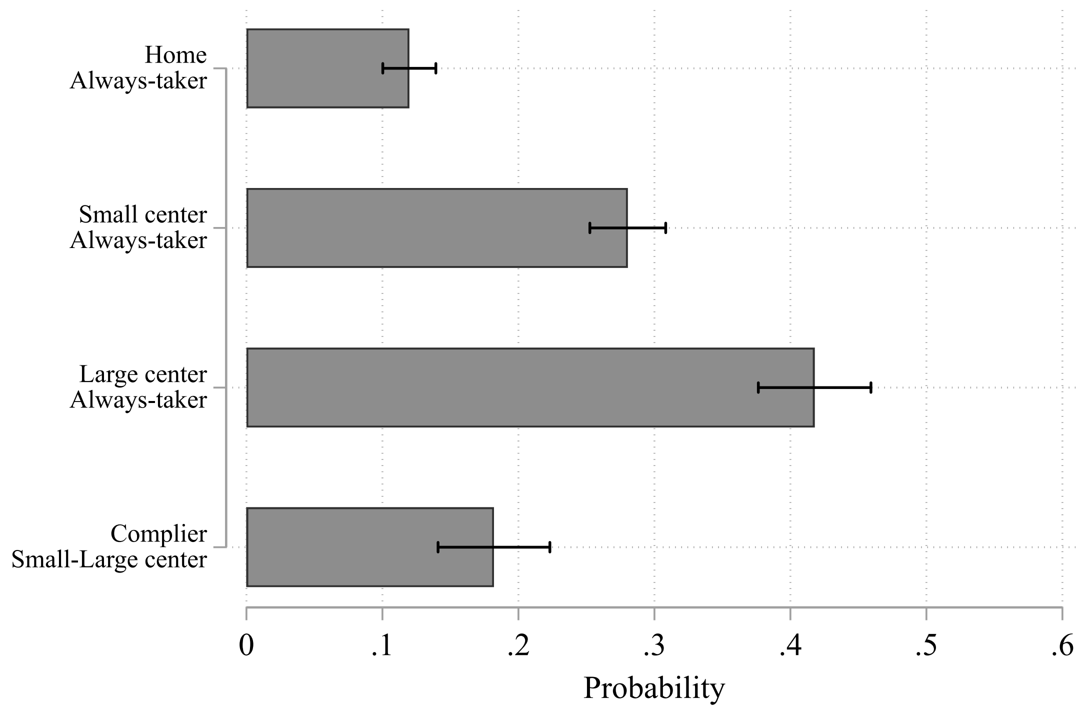
Figure 2.5: Effect of the lottery outcome on the probability of choosing a type of childcare, by distance to the nearest large center



Source: subsample from Bernal et al. (2019)

Note: The figure shows the probability of each choice of childcare conditional on winning and not winning the lottery, at different values of distance to the nearest large center (top panel). Each column denotes a childcare choice. The panel in the second row plots the difference in the probability of each childcare choice of winning *versus* not winning the lottery, by distance to the nearest large center. The probabilities and the marginal effects results from the regression in section 2.3.5, with estimated coefficients in Table 2.6.

Figure 2.6: Unconditional Probability of always-takers and compliers due to variation in the lottery outcome



Source: subsample from Bernal et al. (2019)
Note: The figure shows the estimated probability of always-takers and compliers induced by the variation of the lottery outcome without conditioning on distance to large centers. The results for the margin of home care *versus* large centers are negative, of 3.2%, but not statistically significant. The latter is not presented in the graph.

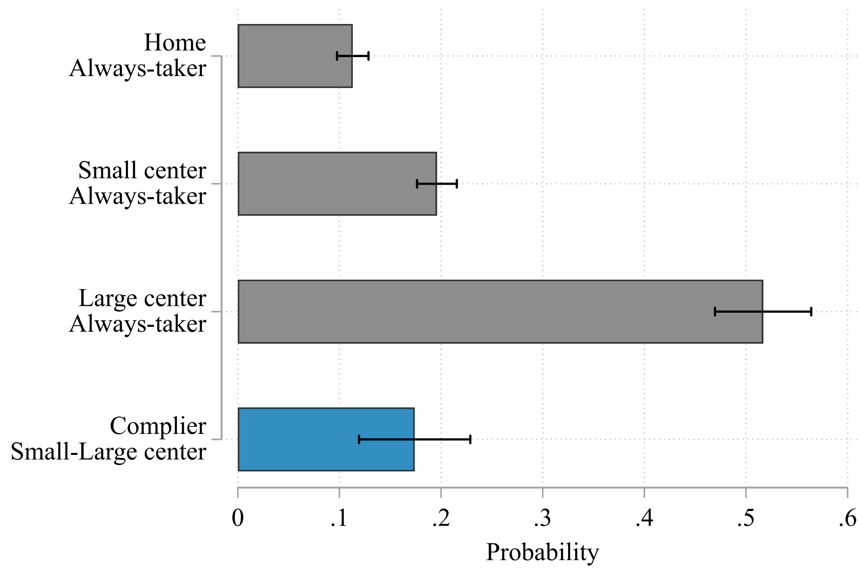
Table 2.7: LATEs, due to the unconditional variation in the lottery outcome

LATE	Cognitive	Socio-emotional	Nutritional
l versus s	-0.359	-0.435	0.203
	(0.306)	(0.351)	(0.326)

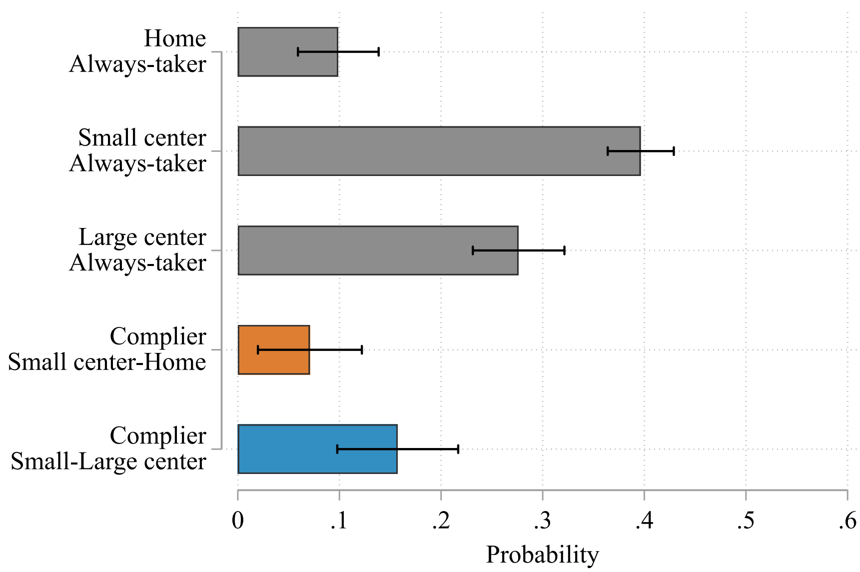
Source: subsample from Bernal et al. (2019)
Bootstrap standard errors are in parentheses. All estimations include city fixed effects.
$*p < 0.1, **p < 0.05, ***p < 0.01$

Figure 2.7: Average Probability of always-takers and compliers due to variation in the lottery outcome, above and below distance to large centers cut-off

(a) Distance to large centers (l) is *less than* 0.75km



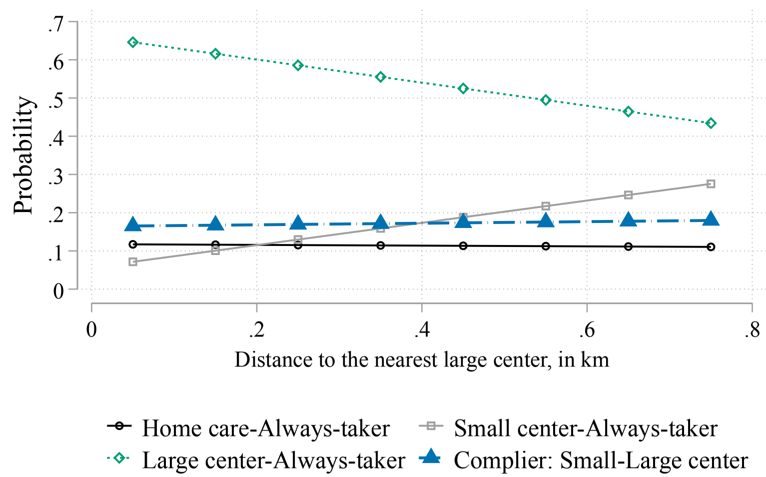
(b) Distance to large centers (l) is *above* 0.85km



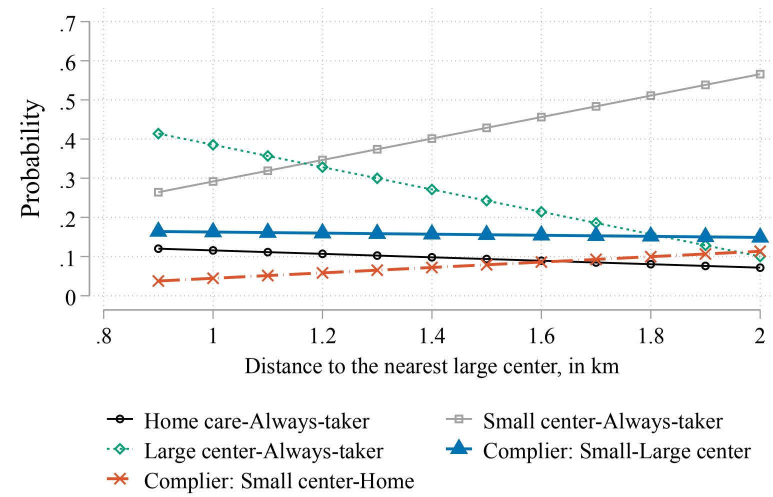
Source: subsample from Bernal et al. (2019)
Note: The figure shows the estimated probability of always-takers and compliers induced by the variation of the lottery outcome conditioning on distance to large centers. The top panel presents the probabilities of always-takers and small-large centers compliers, when distance to the nearest large center is below 0.75km. Panel (b) shows the probability of always-takers and compliers along two margins (small center-home care, and small-large centers). The x-axis shows the estimated probability for each group. The y-axis displays the groups.

Figure 2.8: Probability of always-takers and compliers due to variation in the lottery outcome, above and below the cut-off of distance to large centers

(a) Distance to large centers (l) is *less than* 0.75km



(b) Distance to large centers (l) is *above* 0.85km

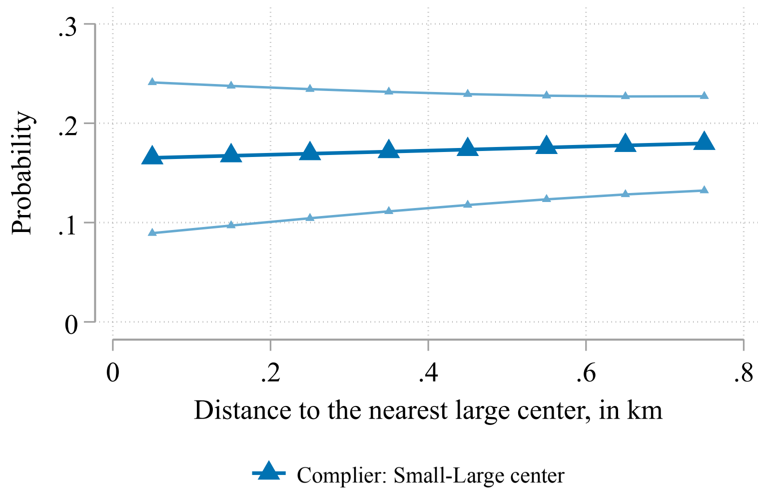


Source: subsample from Bernal et al. (2019)

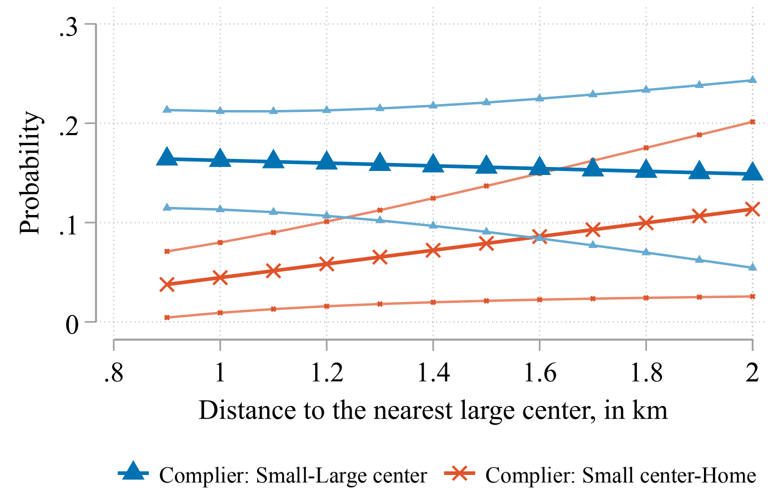
Note: The figure shows the estimated probability of always-takers and compliers induced by the variation of the lottery outcome, by distance to the nearest large centers. Panel (a) presents the probabilities of always-takers and small-large centers compliers, when distance to the nearest large center is below 0.75km. Panel (b) shows the probability of always-takers and compliers along two margins (small center-home care, and small-large centers).

Figure 2.9: Probability of compliers due to variation in the lottery outcome, above and below the cut-off of distance to large centers

(a) Distance to large centers (l) is *less than* 0.75km



(b) Distance to large centers (l) is *above* 0.85km



Source: subsample from Bernal et al. (2019)

Note: The figure shows the estimated probability of compliers induced by the variation of the lottery outcome, by distance to the nearest large centers. Panel (a) presents the probabilities of small-large centers compliers, when distance to the nearest large center is below 0.75km. Panel (b) shows the probability of compliers along two margins (small center-home care, and small-large centers). Both figures show confidence intervals at the 90% level.

Table 2.8: Average Baseline Characteristics of always-takers and compliers due to variation in the lottery outcome, when distance to the nearest large center (l) is ***less than*** 0.75km

Variable	Home (h) always-taker	Small center (s) always-taler	Large center (l) always-taker	Complier: Small-large center
Age in months	37.554	36.186	36.094	42.061
Male (%)	0.494	0.533	0.501	0.606
Children 0-5 at home	1.847	1.543	1.440	1.664
Mother's years of education	7.480	8.593	8.958	8.945
Low income household (%)	0.533	0.383	0.394	0.457
Cognitive development	0.101	0.066	0.085	-0.343
Socio-emotional development	-0.083	-0.008	-0.184	0.647
Nutrition	-0.293	0.062	0.079	-0.037
Distance to s, in km	0.197	0.213	0.262	0.184
Distance to l, in km	0.423	0.543	0.500	0.376

Source: subsample from Bernal et al. (2019)

Note: All socioeconomic variables were collected at baseline. Cognitive development is a composite of scores from the ASQ. Socio-emotional development is a composite of behavioral components of the ASQ. A higher score in socio-emotional development signals behavioral problems. All scores from the ASQ are age standardized. Nutritional development corresponds to z-scores for weight-for-age, height-for-age, and weight for height.

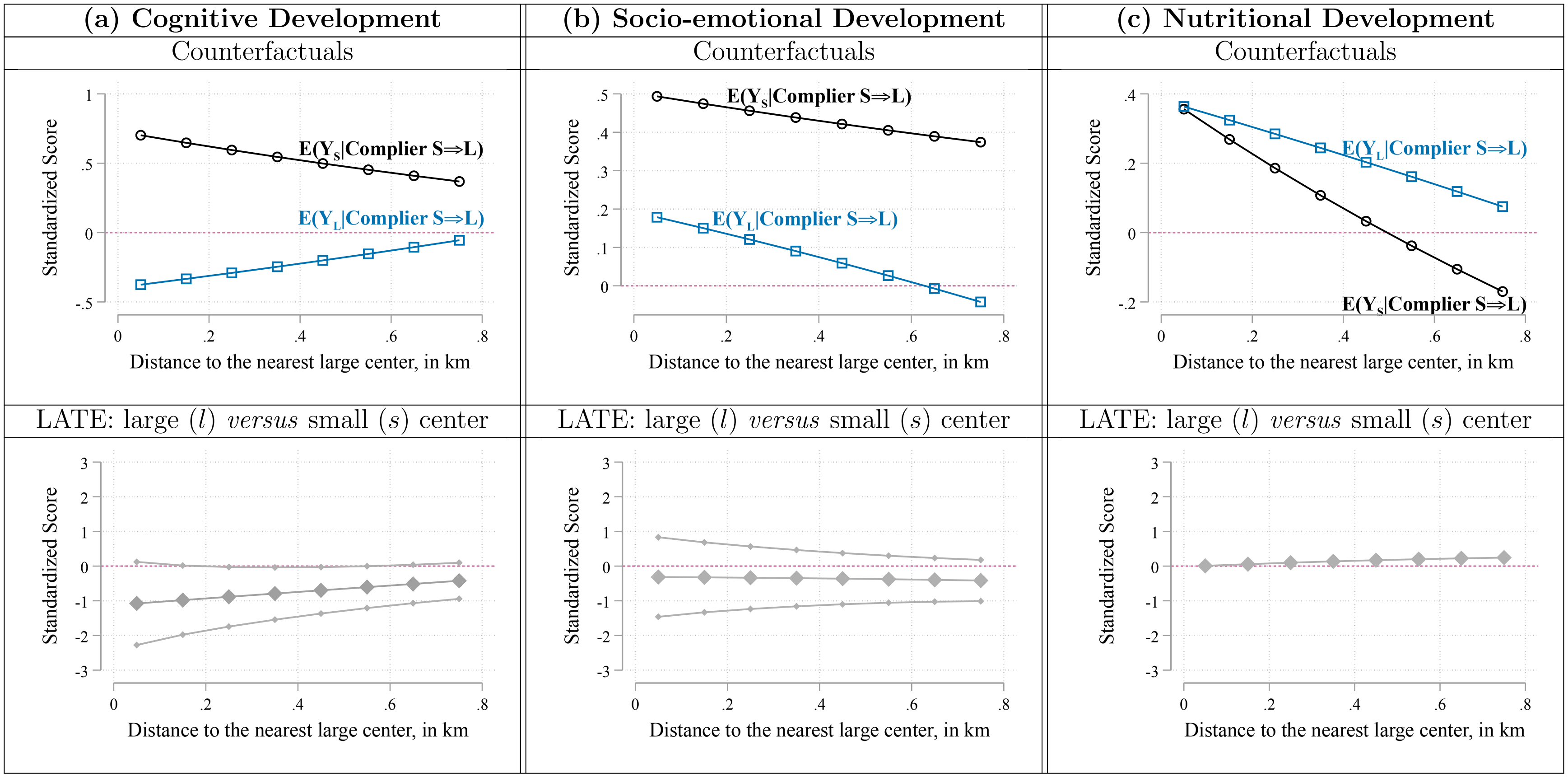
Table 2.9: Average Baseline Characteristics of always-takers and compliers due to variation in the lottery outcome, when distance to the nearest large center (l) is ***above*** 0.85km

Variable	Home (h) always-taker	Small center (s) always-taler	Large center (l) always-taker	Complier: Small center-home care	Complier: Small-large center
Age in months	39.601	36.313	37.759	34.147	33.923
Male (%)	0.506	0.555	0.584	0.257	0.294
Children 0-5 at home	1.651	1.456	1.309	1.431	1.683
Mother's years of education	9.193	8.686	8.468	5.072	10.972
Low income household (%)	0.551	0.457	0.348	0.605	0.491
Cognitive development	-0.113	-0.131	0.197	0.304	-0.448
Socio-emotional development	0.010	0.044	-0.173	-0.490	0.258
Nutrition	0.352	0.117	-0.033	-0.924	-0.068
Distance to s, in km	0.485	0.309	0.298	0.198	0.661
Distance to l, in km	1.693	1.601	1.021	0.847	1.540

Source: subsample from Bernal et al. (2019)

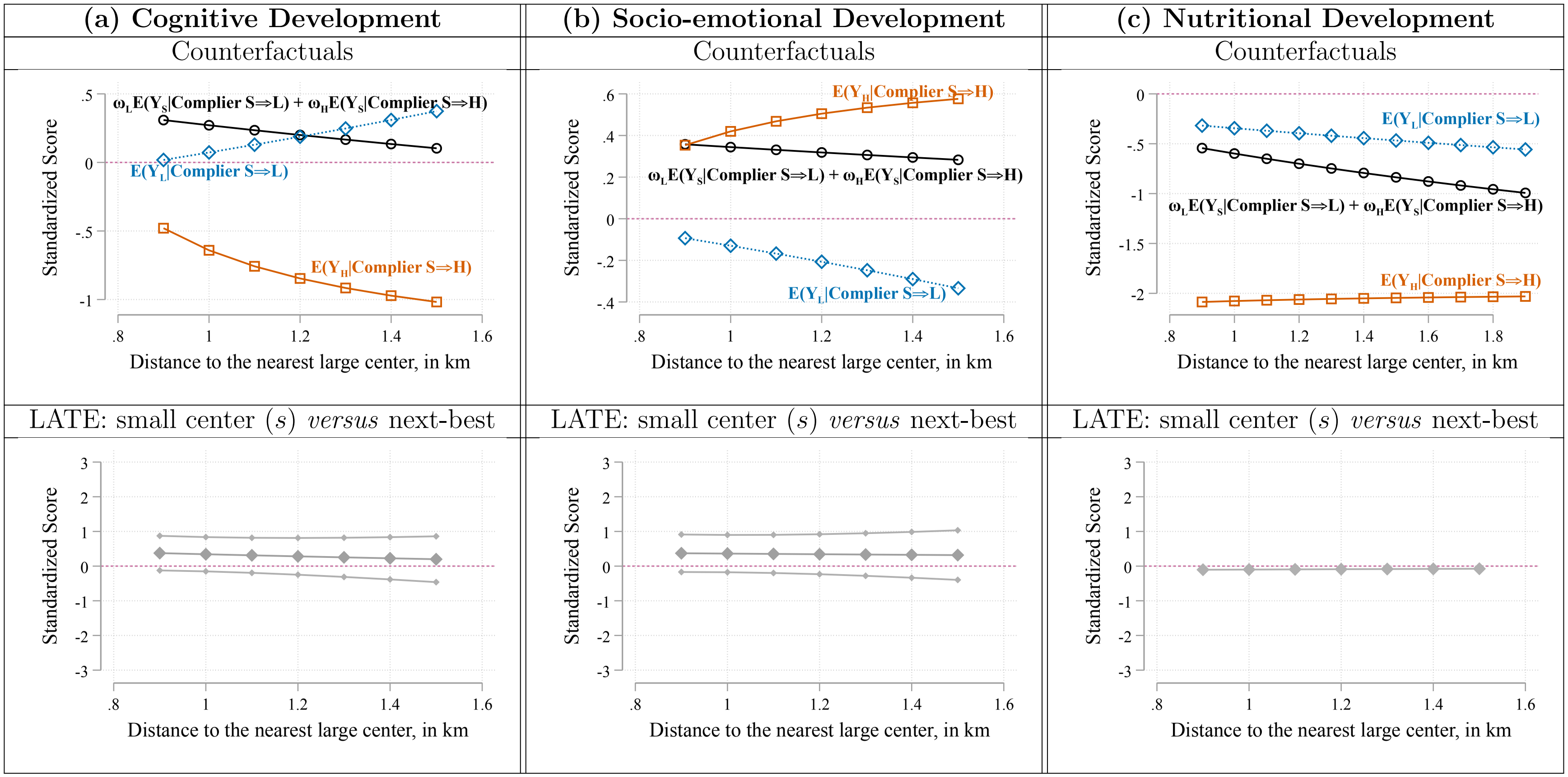
Note: All socioeconomic variables were collected at baseline. Cognitive development is a composite of scores from the ASQ. Socio-emotional development is a composite of behavioral components of the ASQ. A higher score in socio-emotional development signals behavioral problems. All scores from the ASQ are age standardized. Nutritional development corresponds to z-scores for weight-for-age, height-for-age, and weight for height.

Figure 2.10: Counterfactuals and Local Average Treatment Effects, for small-large center $(s-l)$ compliers when distance to the nearest large center (l) is ***less than*** 0.75km



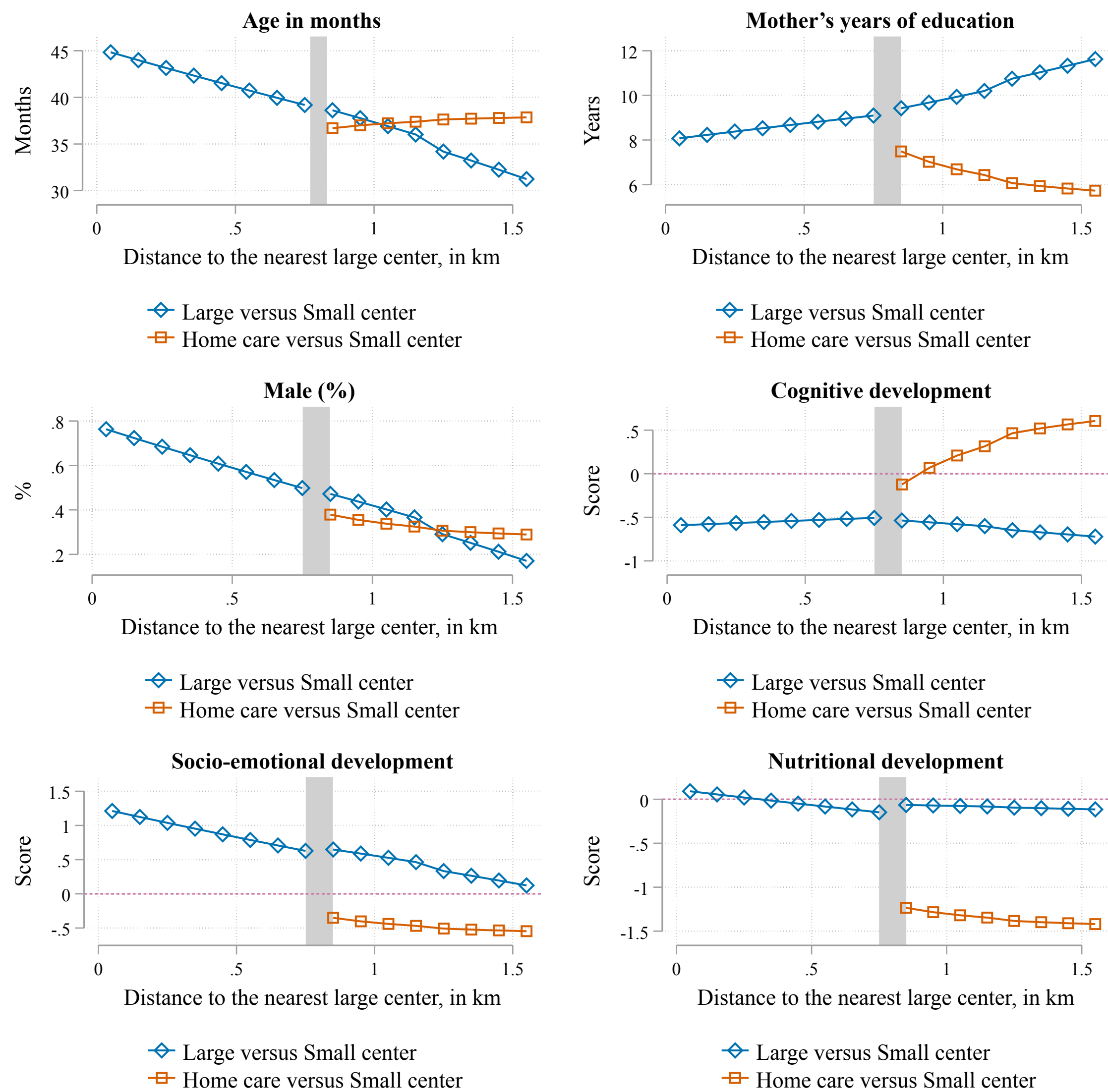
Source: subsample from Bernal et al. (2019). Note: The figure shows counterfactuals (in the first row) and LATE (in the second row) for compliers along the small-large centers margin, for distance to the nearest large center below 0.75km. Column (a) presents results for cognitive development, which is a composite of scores from the ASQ. Column (b) presents results for socio-emotional development, which is a composite of behavioral components of the ASQ. Column (c) presents results for nutritional development. $E(Y_s|\text{Complier S} \rightarrow L)$ (hollow circles in black) denotes the counterfactual outcome of small centers for compliers along the small-large center margin. $E(Y_L|\text{Complier S} \rightarrow L)$ (hollow squares in blue) denotes the counterfactual outcome of large centers for compliers along the small-large center margin. Confidence intervals at the 90% level are constructed with bootstrap standard errors.

Figure 2.11: Counterfactuals and Local Average Treatment Effects, for compliers when distance to the nearest large center (l) is ***above*** 0.85km



Source: subsample from Bernal et al. (2019). Note: The figure shows counterfactuals (in the first row) and LATE (in the second row) for compliers along the small-large centers margin and compliers along the small centers-home care margin , for distance to the nearest large center above 0.85km. Column (a) shows results for cognitive development, column (b) presents results for socio-emotional development, column (c) shows results for nutritional development. $E(Y_L|\text{Complier S})$ is the counterfactual outcome of large centers for compliers along the small-large center margin. $E(Y_H|\text{Complier S} \to H)$ is the counterfactual outcome of home care for compliers along the small centers-home care margin. $\omega_L E(Y_s|\text{Complier S} \to L) + \omega_H E(Y_s|\text{Complier S} \to H)$ is a weighted average of the counterfactual for small centers along the two margins of choice. $\omega_L + \omega_H = 1$. Confidence intervals at the 90% level, with bootstrap standard errors.

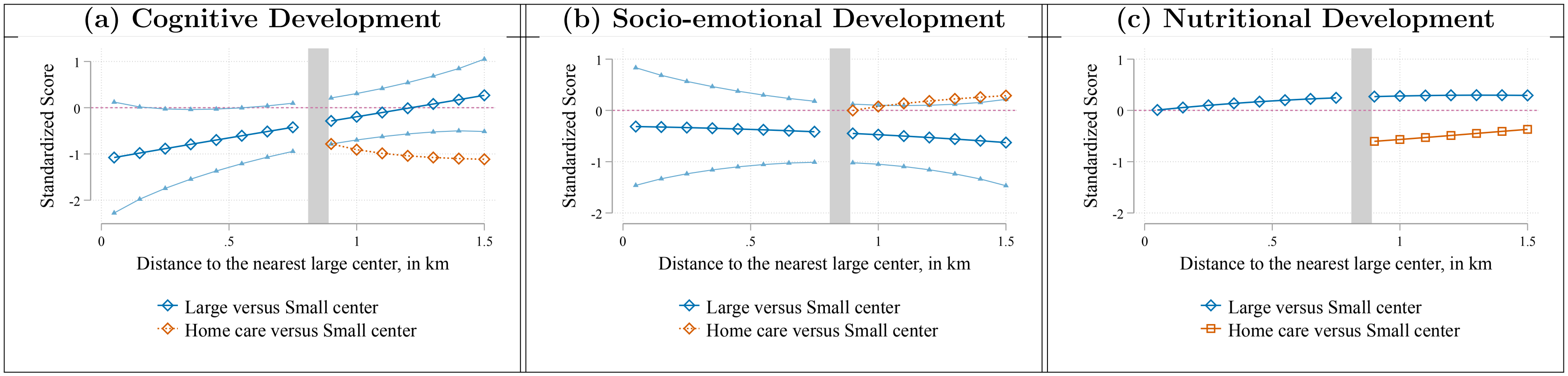
Figure 2.12: Average of baseline characteristics for compliers along the large *versus* small center margin and the home-care *versus* small center margin, by distance to the nearest large center (l)



Source: subsample from Bernal et al. (2019)

Note: The figure shows the average of baseline characteristics for compliers along the small-large centers margin (hollow diamonds in blue) and compliers along the small centers-home care margin (hollow squares in red), by distance to the nearest large center. The grey bar denotes the threshold at which the conditional rules of behavior change.

Figure 2.13: Local Average Treatment Effects for compliers along the large *versus* small center margin and the home-care *versus* small center margin, by distance to the nearest large center (l)



Source: subsample from Bernal et al. (2019)

Note: The figure shows conditional LATEs for compliers along the small-large centers margin (hollow diamonds, in blue) and compliers along the small centers-home care margin (hollow squares, in red), by distance to the nearest large center. Column (a) presents results for cognitive development, column (b) presents results for socio-emotional development, column (c) presents results for nutritional development. Effects below the bar in grey are estimated with the conditional rules of parental behavior when distance to large centers is below the threshold. Above the grey bar, the effects for each margin of choice are separately identified by imposing the assumption of homogeneity of counterfactuals.

Table 2.10: Conditional LATEs, due to variation in the lottery outcome by distance to the nearest large center (l)

LATE	Cognitive	Socio-emotional	Nutritional
Distance to the nearest large center below 0.75km			
l versus s	-0.718*	-0.368	0.06
	(0.407)	(0.450)	(0.492)
Distance to the nearest large center above 0.85km			
s versus next-best	0.229	0.323	0.021
	(0.441)	(0.452)	(0.281)
l versus s	0.168	-0.579	0.367
	(0.399)	(0.442)	(0.594)

Source: subsample from Bernal et al. (2019)

Bootstrap standard errors are in parentheses. All estimations include city fixed effects.

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

Table 2.11: Robustness check: Conditional LATEs, at different thresholds of distance to l

	Cognitive			Socio-emotional			Nutritional		
LATE	0.75km	0.75-0.85km	0.85km	0.75km	0.75-0.85km	0.85km	0.75km	0.75-0.85km	0.85km
Below threshold									
l versus s	-0.718*	-0.718*	-0.683*	-0.368	-0.368	-0.374	0.06	0.06	0.076
	(0.407)	(0.407)	(0.457)	(0.450)	(0.450)	(0.481)	(0.492)	(0.492)	(1.091)
Above threshold									
s versus next-best	0.252	0.229	0.229	0.331	0.323	0.323	0.068	0.021	0.021
	(0.341)	(0.441)	(0.441)	(0.373)	(0.452)	(0.452)	(0.243)	(0.281)	(0.281)
l versus s	0.083	0.168	0.168	-0.553	-0.579	-0.579	0.325	0.367	0.367
	(0.361)	(0.399)	(0.399)	(0.421)	(0.442)	(0.442)	(0.392)	(0.594)	(0.594)

Source: subsample from Bernal et al. (2019)

Bootstrap standard errors are in parentheses. All estimations include city fixed effects.

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

Chapter 3: On Track to Idleness: High School Type Choices and College Enrollment upon Graduation

3.1 Introduction

Vocational training during high school aims at easing school-to-work transitions, providing students with skills that help them succeed in the labor market. Although some countries have adopted a specialization approach (such as Germany and Japan, Chmielewski, 2014), where students follow either an academic or vocational track[1] early on, other countries have promoted efforts to de-track their educational system by postponing the timing of selection into training tracks and extending the years of general learning with somewhat positive effects on labor market outcomes (Bertrand et al., 2021; Canaan, 2020; Ollikainen, 2021; Pekkarinen, Uusitalo, and Kerr, 2009; Zilic, 2018). On the other hand, the evidence on educational trajectories remains mixed: from null- to adverse- effects on dropping out of high school (Bertrand et al., 2021; Zilic, 2018), positive effects on high school completion (Canaan, 2020) or higher education degrees (Canaan, 2020; Pekkarinen

[1] I refer to *tracking* as the educational systems where students must choose either a vocational or an academic track at some point during primary school or high school. This differs from other tracking strategies that assign students to different groups based on their academic performance. For a review of evidence on the effects of tracking by academic *versus* vocational training, and tracking by ability see Betts (2011).

et al., 2009). Despite recent advances in the literature on vocational education, there is no clear consensus on whether flexible tracking or school specialization can improve students' academic and labor market outcomes.

In this chapter, I estimate the effect of choosing a type of high school on student's academic performance and educational trajectories. I study the context of high school choice in Chile, where students can choose from three types of schools: Academic, Vocational, or Hybrid. The latter type is particularly interesting, since it offers both tracks (academic or vocational) allowing for a higher heterogeneity in class composition (at least in terms of track preference) than specialized schools. Nonetheless, the role of Hybrid high schools on students' educational attainment and achievement has not been previously studied. In my analysis, I provide evidence of the differential effect on educational outcomes of attending Hybrid schools rather than specialized schools such as Vocational or Academic.

An additional feature of the Chilean system is that all schools must provide the same core curricula for the first two years of high school, regardless of the type of training offered. As a result, Hybrid schools might also allow some students to delay the choice of track compared to Academic or Vocational schools. All in all, differences in student's outcomes between types of schools are likely to suffer from self-selection and reflect heterogeneity in class composition as well as differences in overall school quality, timing of tracking, and type of track chosen.

The literature on vocational training has dealt with selection and endogeneity challenges with difference-in-difference strategies (Hanushek, Schwerdt, Woessmann, and Zhang, 2017), instrumental variables (Kreisman and Stange, 2020), or regression discontinuity designs (Bertrand et al., 2021; Canaan, 2020; Dougherty, 2018; Elacqua, Prada, Navarro-Palau, and Soares, 2019; Malamud and Pop-Eleches, 2010; Silliman and Virtanen, 2019; Zilic, 2018). An additional issue that has received less attention is that, when making their schooling decisions, students face multiple choices and therefore have different fallback options. For instance, if Academic schools where no longer available, some students who choose Academic schools would have chosen Hybrid schools while other would have chosen Vocational schools. These different fallback options can translate into, for instance, distinct effects of Academic schools across students. While there are several methodological (Angrist, Imbens, and Rubin, 1996; Heckman and Pinto, 2018; Heckman et al., 2006; Heckman and Urzúa, 2010) and empirical (Kirkeboen et al., 2016; Kline and Walters, 2016; Mountjoy, 2019) papers dealing with identification in settings with multiple choices, there is less evidence for contexts of high school choice, and vocational education in specific.

In this chapter, I use choice by outcome interactions and univariate Two-Stage Least Squares (2SLS) (Heckman and Pinto, 2018; Mountjoy, 2019) to estimate effects of type of high school in a context of multiple endogenous choices. I use administrative data for the cohort of high school entrants in Chile in 2012, their choice of type of high school and their educational trajectories between 2015-2018. I also use information on test scores two years after entering high school, to measure short-term effects of school choice on academic achievement. I exploit variation in travel time to each type of school, by georeferencing the students' primary school and the location of each high school establishment. Using the variation on travel time to each type of school, I show that IV identifies Local Average Treatment Effects (LATEs) of each type of school *versus* the next-best (*i.e.* a combinaton of different fallback options) (Heckman et al., 2006; Kirkeboen et al., 2016). I further employ an assumption of homogeneity (Hull, 2018; Kline and Walters, 2016; Lee and Salanié, 2020), to decompose each high school LATE into fallback option subLATEs.

I find that Academic schools have positive local effects on the probability of completing high school and enrolling in Bachelor's programs, *versus* the next-best (*i.e.*, a mixture of Academic and Vocational high schools). I attribute most of this effect to students who would change their choice of school from Vocational and into Academic schools, as the travel time to Academic schools decreases. These effects are estimated using the variation in travel time to Academic schools, and about 8.4% of students in my sample would switch towards Academic schools when travel time to this option decreases.

On the other hand, I find that Vocational schools have a positive LATE on the probability of completing high school, but decrease the probability of enrolling in academic colleges, *versus* Hybrid Schools. Importantly, neither Vocational nor Hybrid schools seem to have an effect on the probability of enrolling in a vocational college. While I observe that Vocational schools might increase the probability of enrolling in a vocational college, this effect is not statistically significant and it is negligible in magnitude. Overall, it seems that Vocational schools can help students complete high school but have no effect in inducing them to continue their studies. These results are in line with previous findings (Brunello and Checchi, 2007; Brunner et al., 2019, and Bassi and Urzua, 2010 for Chile), such that vocational education can decrease the likelihood that students invest on their education later on. Previous studies have also found that de-tracking educational systems (*i.e.*, increasing the age of tracking or mixing academic and vocational training) in countries such as France (Canaan, 2020), Finland (Ollikainen, 2021), and Norway (Bertrand et al., 2021), can increase completion of vocational higher education degrees.

Last, I also uncover positive and significant effects of Hybrid schools on the probability of completing high school and enrolling in an academic college. The Hybrid school LATEs are estimated using the variation in travel time to these schools, which would induce about 14% of students to change their school choices. The decomposition of Hybrid schools LATE into fallback option subLATEs shows larger gains among students who would change their choice from Vocational schools and towards Hybrid schools, as the travel time to Hybrid schools decreases. In particular, I show that students who would change their choice of a Vocational school towards a Hybrid school, as the latter becomes relatively cheaper, would experience an increase in their likelihood of enrolling in an academic college as well as improvements in their average test scores. On the other hand, students switching from Academic schools and towards Hybrid schools have a higher probability of not completing high school.

The results for Academic and Hybrid schools seem to be robust to adding controls for local amenities, as well as to different specifications of the functional form of the instruments. For the former, I use data on proximity to hospitals, police stations, and libraries to control for the provision of services at the local level. The latter refers to including quadratic and cubic terms of the instruments (travel time) to determine how sensitive are the results to the different specifications. Although the results seem to be stable for Academic an Hybrid schools, most of the effect of Vocational schools on high school completion disappears. Only the LATEs of Vocational schools *versus* the next-best on the probability of not completing high school remains unchanged, and it is significant and negative. All in all, the latter indicates that Vocational schools might not improve students' outcomes in the short-to-medium term.

This chapter is organized as follows. The following Section briefly describes the institutional context. Section 3.3 describes the data, analyzes school-type choices and outcomes, and presents results using a standard IV strategy. Section 3.4 presents a simplified framework of high school choice and the school-type effects that I can identify with the variation in the IVs. Section 3.5 discusses the estimation strategy and identification of treatment effects. Section 3.6 presents and discusses empirical results, and Section 3.7 concludes.

3.2 School System Characteristics

The education system in Chile is organized with eight grades of primary school and four of secondary. After completing primary school, students select a type of high school: (i) academic, (ii) vocational, or (iii) hybrid. Academic high schools (*Científico Humanista*, in Spanish) focus on a general curriculum, designed to prepare students for the school-to-college transition. Vocational schools (*Técnico Profesional*, in Spanish) enhance skills useful for the school-to-work transition and offer different specializations; their curriculum is adapted to the diverse requirements of labor sectors. Students in these schools can select from commercial, industrial, technical, agricultural, or maritime specializations. Hybrid schools (*Polivalente*, in Spanish) offer both tracks, academic and vocational, but allow students to choose within vocational tracks two years after entering high school. During the first two years of high school, all schools follow the same curricula and subsequently specialize in academic or vocational training during the last two years of secondary education.

The role of Hybrid schools has mostly remain unstudied. Most papers that analyze vocational training in Chile either focus on the final choice of track or ignore Hybrid schools altogether (Larrañaga et al., 2013). The prevalence of these schools increased around 1992-1993 with an educational program[2] whose objective was to expand the coverage of vocational training, particularly for low-income students. This expansion stems from schools which were academic initially but then started to incorporate vocational tracks. Compared to purely vocational schools, Hybrid schools do not offer as many specializations, with only 40% of these schools offering more than four options (Sevilla and Sepúlveda, 2015). There is heterogeneity regarding how students select a track within Hybrid schools, but it is common that those with highest test scores end up on the academic track (Sevilla and Sepúlveda, 2015).

[2] *Programa de Habilitación de Establecimientos Humanista-Científicos con Especialidades de Educación Técnico- Profesional*, in Spanish.

3.3 Data

This chapter uses data on test scores, schooling trajectories, and college enrollment. It focuses on the universe of students enrolled in the first grade of high school in 2012 in Chile, their choice of type of high school, academic performance, and higher education decisions. For schooling choices and test scores, I use two administrative datasets: the System for Measuring the Quality of Education (SIMCE, in Spanish) for 2011 and 2013, and the Registry of Enrollment in Higher Education for 2015 to 2018 from the Higher Education Information Service (SIES, in Spanish). Also, I use administrative data collected by the Ministry of Education on academic performance and attendance from 2011 to 2018 for the universe of enrolled students in that period (*Rendimiento escolar 2011-2018 por estudiante*, in Spanish).

I restrict my analysis to individuals born between 1993 and 1997, who were enrolled in the last year of primary school (*i.e.*, eighth grade) during 2011. For this class, school-type choices took place in 2012 which the year when they were entering high school. Of those students, around 80% took the National Standardized Exam (SIMCE) in 2011 during the last year of primary school. The data from SIMCE contains information on socioeconomic variables, type of school, test scores, and measures of classroom activities and classroom/school learning environment. I link the student-level data with geographic information for all primary schools and high schools collected by the Ministry of Education; the data includes geographic coordinates for all schools in the educational system. With this detailed information, I construct measures of travel time and distance from each student's primary school to each type of high school and use them as instruments in the estimation strategy.[3]

Almost the totality (about 80%) of the high school entering class that presents the SIMCE exam in 2011 took the SIMCE exam in 2013 when they attended the second year of high school (two years after selecting a school type). The data from SIMCE 2013 includes test scores for math and reading; I use these test scores to measure short-term academic achievement for students in my sample. On the other hand, I use the data from SIMCE 2011 to control for baseline characteristics at the student level (before the decision on high school type takes place). SIMCE 2011 includes information on maternal education level, household income, household size, student's sex and age, as well as information on classroom activities and violence, bullying, and overall school and classroom environment.[4] Moreover, the SIMCE dataset contains information at the school level to match with the school registry from the Ministry of Education. I use the registry to restrict my sample to students who effectively switch schools between the end of primary and start of high school,

[3]The information on georeferenced location of student's primary school comes from the 2014 registry of schools from the Ministry of Education, two years after the high school-type choice takes place. To serve the instrumental variables' argument, ideally, I should use distance to each school-type before entering high school. Nonetheless, internal migration in Chile is minimal, at about 0.6% per year (Soto and Torche (2004) and Rau, Urzúa, and Reyes (2015)). This low migration rate supports the argument that the fraction of students and schools changing residence between 2012 and 2014 is not considerably large to threaten the strategy.

[4]SIMCE collects data on classroom activities and environment using a questionnaire for math, reading, and science teachers. All questions are identical, regardless of the academic subject. I focus on math teachers' responses, which contains information for a higher share of students than responses from other teachers. Classroom learning activities is a composite of frequency of in-class group or individual activities, students' presentations, review of homework and assignments, review of curricula and contents, solving problems and questions; as well as teacher confidence in motivating students with low interest and those with disruptive behavior to learn the class materials. Classroom learning environment is a composite of frequency of high school level incidents such as fights, insults, threats and bullying between students, use of weapons (knives, guns, etc.), destruction of school tools and facilities, insults from students to teachers; as well as of frequency of classroom level incidents such as disrespectful behavior between students and from students to teachers, fights during class, lack of care of the classroom and facilities.

which corresponds to 70% of those in the last grade of primary in 2011. Notably, the transition from primary school to high school also entails a reshuffling of student's peers since most students do not transfer to high school with other students in their primary school.[5]

Students in my sample are evenly distributed between Academic (A), Vocational (V), and Hybrid (H) high schools (see Table 3.2). About 33.2% of students attended an academic high school, 33.9% attended a vocational high school, and 32.9% were in a hybrid high school. Despite the even distribution across high school-type, most students select an academic track instead of vocational, with about 60% of students among the former path. To understand the effect of the type of high school on educational trajectories (*e.g.*, high school completion and enrollment in higher education), I match students in my sample to the records of academic performance and attendance during high school from the Ministry of Education, and enrollment in higher education from SIES. I use a three-year window from 2015-2018 to define the student's educational status, where 2015 refers to on-time high school graduation. In 2018, 62.6% of the sample was attending higher education, with 35% in academic colleges and 26% in vocational colleges. Moreover, 17% of students do not graduate from high school, and about 20% complete high school but do not enroll in higher education.

Among those in the estimating sample, a higher fraction of disadvantaged male students attended hybrid or vocational high schools (see Table 3.3). While the share

[5]See Appendix B. On average, the share of students who transfer to the same high school is about 7%.

of female students among A high schools is 55.6%, this share is only 45.6% among V high schools and 50% among H high schools. In terms of household income, half of the students in V and H high schools belong to households in the bottom two levels of income, compared to only 35% of students in A high schools in those income levels. Figure 3.1 shows school participation by income level. More than half of students in the highest income level attended an A high school; among those in the lowest income level almost 30% (30%) participated in a hybrid (vocational) high school. Figure 3.2 shows mothers' level of education in 2011, by type of school. Mother's education seems to correlate strongly with the choice of an A high school: about 60% of students with mothers who enrolled or completed a higher education degree choose an A high school; in contrast, among those with mothers with less than primary the share of A students is between 20-25%. Last, the type of high school that the student's mother attended (Academic or Vocational) does not seem to correlate with the type of school students choose. Figure 3.2 shows that the distribution of school-type choice is even among those with mothers who attended Academic or Vocational High Schools (as shown by the horizontal bars for "Secondary Academic" and "Secondary Vocational" in Figure 3.2).

Students who enrolled in academic (A) high schools had higher average test scores, measured with SIMCE 2011, than those at V or H schools (Table 3.3). Figure 3.3 shows the distribution of test scores for Reading, Math, Biology, and Social Science by type of school. Regardless of the academic subject, the distribution of test scores of those in A schools is always to the right of the distribution of test scores for V and H schools. Moreover, there are no visible differences in the distribution of test scores between those at H and V schools. In other words, students in H and V schools started from similar levels of academic achievement at baseline, before choosing the type of high school. On average, the most considerable difference in test scores in the last grade of primary between H and V schools corresponds to reading: the average reading test score for students in V schools is -0.129 compared to -0.095 for students in H schools.

The share of students who do not complete high school and do not enroll in higher education at H or V schools is twice as high as that of A schools (Figure 3.4). While about 10% of students in A high schools do not complete high school, this share is around 18.6% for V school students and 21.2% for H students. The percentage of high school graduates among A and H schools is lower than among V schools, which in practice translates into higher shares of students enrolling in higher education at the former (A and H) schools with shares of 78.1% and 55.2%, respectively.

While the differences in educational attainment could indicate whether or not certain types of schools successfully foster students' academic trajectories, they can penalize vocational education. The objective of vocational training is to ease the school-to-work transition and promoting task-specific abilities, perhaps at the expense of enhancing students' academic skills. Given that I do not have data on labor market participation, I focus on enrollment in vocational programs in higher education in this paper. Figure 3.4 shows that the share of students who enroll in vocational Higher education is higher among V than among A schools. On the other hand, differences in vocational higher education enrollment between V and H schools are negligible (30% versus 29%).

Academic schools are predominantly private and represent about 64.6% of the total number of schools (Table 3.8). In contrast, less than half of V and H schools are either private/subsidized or only private. The difference between these two categories is that the former can be publicly and privately funded and charge lower tuition than private schools. Moreover, less than 6.2% of all schools in my sample are rural. This share is higher among V schools (at 18%) followed by H schools (at 6.1%) and last by A schools with just 3.6% of rural schools. In terms of the average number of students enrolled in the first grade of high schools, the size of the class is three times higher among H schools *versus* A schools, and 4.6 times higher among V than A schools. Overall, while there are fewer V and H schools their class size tends to be larger.

3.3.1 Exploratory Analysis

The type of school students choose depends on socioeconomic characteristics and constraints, as well as unobserved factors such as preferences and abilities. In this sense, students self-select into academic, vocational, or hybrid schools based on observed and unobserved variables. Moreover, they choose among alternatives for which no inherent rank exists, which can be (weakly) substituted for each other. The estimation of causal effects of high school choices on test scores and college enrollment requires dealing with: (i) endogeneity from unobservables and self-selection, and (ii) the unordered nature of choice. Otherwise, the estimated effects will be biased and inconsistent.

Table 3.5 shows results of running the following equation by OLS,

$$Y_i^j = \beta_0 + \beta_1 D_{Ai} + \beta_2 D_{Vi} + \beta_3 X_i + u_i \tag{3.1}$$

where $D_{Ai} = \mathbb{1}[\text{Academic high school}]$ is a binary variable that takes the value of one if the student chooses an Academic (A) school; $D_{Vi} = \mathbb{1}[\text{Vocational high school}]$ is a binary variable that takes the value of one if the student chooses a Vocational (V) school. Hence, the omitted category is H (hybrid) high schools. β_1 compares academic schools versus hybrid schools, and β_2 compares vocational schools with hybrid schools. Y_i^j represents outcomes (educational attainment at 2018),

$$Y^j = \begin{cases} 1, & \text{if the student reaches educational status} = j, \\ 0, & \text{else} \end{cases}$$

where $j \in \{$High School Incomplete, High School Graduate, Technical/Technological Higher Education, Bachelor's Program$\}$. X contains baseline variables before the choice of type of school, such as student's age, sex, standardized test scores in reading, math, biology, and social science from SIMCE in 2011. It also includes household income level and maternal level of education. In addition, I include indexes of learning activities and environment at the classroom and school level. I also control for local characteristics in 2011 of the municipalities where the student attended primary school such as the fraction of urban population, the unemployment rate, population density, and poverty level. All the regressions include controls for rural and private schools. I also estimate regressions where Y represents the standardized math and reading test scores in SIMCE 2013 (two years after starting high school). The results of these regressions are in Table 3.6.

The OLS estimates are biased due to self-selection into high school-type, omitted variables, and unobserved heterogeneity in u_i (the unobserved component in equation (3.1)). For example, cognitive and noncognitive factors can make students more likely to choose A, ultimately affecting their decision to enroll in college. Also, this framework only allows for pairwise comparisons between school types; none of these coefficients capture treatment effects of specific choices. Endogeneity combined with the unordered nature of choice results in students having different fallback options: some students who choose A would have chosen V (if A was not available), while other students would have selected H instead. Even if the researcher selects different omitted categories in equation (3.1), it would impose a general fallback option that may, or may not, match with the fallback option of the students.

The results in Table 3.5 suggest that relative to hybrid schools, A and V decrease the probability of not completing high school. In terms of higher education enrollment, the results show that students who attended A schools are more likely to enroll in a Bachelor's program than those at H schools, and students from V schools are more likely to enroll in a technical or technological program than H students. In terms of test scores, Table 3.6 shows that, on average, students at H schools have a lower performance in all subjects than students at either A or V.

3.3.1.1 Travel time from Primary Schools to High School Types

To estimate causal effects of school-type choice on educational achievement, I use an Instrumental Variables strategy and exploit variation in travel time (in hours) from the school attended by the students during the last grade of primary to the closest Academic, Vocational, and Hybrid high school.[6] The argument is that travel time, or proximity, to a type of high school decreases the costs of choosing that option but has no direct effect on high school completion or higher education enrollment (*i.e.*, the exclusion restriction). Among the arguments that could threaten the exclusion restriction are: (i) students follow the same high school choice as their peers and those in the same primary school choose the same high school, (ii) areas with shorter travel times between schools are more developed, have better services, and offer more educational opportunities. In terms of the former, I show evidence in Appendix B of a considerable change in peer composition from primary to high school: on average, 7% of students go to the same high school as their peers in primary school. For the latter, I control for local characteristics of the municipalities where the students attended primary school.[7] Furthermore, in Section 3.6.1 I present several robustness checks including controls for proximity to local amenities (such as hospitals, police stations, and libraries).

[6]I use travel time from primary schools to each type of high school to proxy for the travel time from the student's residence, given that I only have available information on schools location rather than student's place of residence. Nonetheless, it is likely that travel time from the student's primary school is a good proxy of actual travel time from the student's home since a large fraction of students in Chile attend schools located in the same municipality where they reside (see, for instance, Kutscher, Nath, and Urzúa (2020) for students enrolled between 2015-2019).

[7]Moreover, in the context of Chile, residential decisions are not determined by school supply. Sanchez (2018) shows that students do not face residential restrictions and can attend any school depending on their willingness to travel and pay.

For my sample, I compute the travel time to all high schools by using georeferenced data from the Ministry of Education for each educational establishment. I use Open Street Maps and the Open Source Routing Machine (OSRM) to compute travel time over public roads by car and construct a variable of minimum travel time to each type of school, by each primary school in my sample. Table 3.7 shows the average hours of travel by car to the nearest A, V, and H high schools by type of school chosen. The average travel time to Academic schools is lower for students who effectively choose an A school than those at V or H. In particular, the average travel time to the nearest A school is almost twice as high for students at H schools *versus* those at A schools. Travel time to the nearest V school follows the same pattern (*i.e.*, shorter, on average, for those who choose V), and in particular is about 1.5 times higher among those in H schools than those in V schools. Similarly, students of H schools have the lowest average travel time to H, and the difference is higher between H and V schools than between H and A schools. It is also worth noting that the average travel time for all schools is below 12 minutes, which indicates that proximity to school is an important predictor of school choice.

Regardless of the type of school, the distribution of high school enrollment along the minimum travel time seems to follow the logic of school substitutes. Figure 3.6 shows school-type enrollment by quartiles of travel time to A schools (first column), V schools (middle) and H schools (last column). To start, across the quartiles of travel time to the nearest A school, students seem to substitute towards relatively cheaper options: as travel time to A increases, enrollment in A schools decreases, and enrollment in V and H increases. There are similar patterns for travel time to V and H schools. Importantly, some of these patterns can capture other constraints, characteristics, and preferences of the students.

Table 3.8 shows that, after controlling for other factors such as student and local characteristics, travel time is an important predictor of school-type choice. The results displayed in Table 3.8 correspond to the first-stage:

$$D_{si} = \alpha_0^s + \alpha_1^s Z_A + \alpha_2^s Z_V + \alpha_3^s Z_H + \alpha_X^s X + \epsilon_{si} \tag{3.2}$$

where $D_{si} = \mathbb{1}[\text{Student chooses type s}]$ for $s \in \{A, V, H\}$, are binary variables that denote the high school-type choice of the student. Z_A refers to the travel time to the nearest Academic high school, Z_V to the nearest Vocational high school, and Z_H to the nearest Hybrid high school. X is a matrix of baseline variables detailed in Section 3.3.1.

Table 3.8 shows that an additional hour in travel time to the nearest A school decreases the probability of choosing A in 29 percentage points and increases the likelihood of selecting V and H in 23 and six percentage points, respectively. An additional hour of travel time to the nearest V schools has no significant effect on the probability of choosing A schools but decreases (increases) the probability of choosing V (H) in 47 (48) percentage points. Last, an additional hour in travel time to the nearest H school decreases the probability of choosing H by 68 percentage points and increases the likelihood of choosing A and V in 35 and 32 percentage points. While these effects seem large in magnitude, this results from the low average of hours of travel time to all schools in the sample, as I discussed in the previous section (the average travel time is below 12 minutes).

The first-stage results also present important selection patterns, particularly in terms of mother level of education and standardized test scores at baseline. In terms of the former, having a mother who studied in an Academic high school increases the probability of attending an A high school in four percentage points and decreases the likelihood of enrolling in V and H schools in 1.4 and 2.5 percentage points, respectively. In terms of test scores, an increase in a standard deviation increases (decreases) the probability of choosing A (V and H), regardless of the academic subject. At the local level, increases in the poverty index increase the probability of selecting H. Last, improvements in the classroom/school climate of primary schools increase the probability of choosing A and decrease the probability of choosing V and H.

Tables 3.9 and 3.10 present results of 2SLS estimation using as instruments the travel time to the nearest A, V, and H schools. I estimated the outcome equations and the first-stage jointly. The interpretation of these effects warrants caution: in settings with multiple unordered choices, standard 2SLS does not recover well-defined effects of one option *versus* another. Rather, multivariate 2SLS with multiple IVs estimates effects that combine Local Average Treatment Effects across many margins of choice (Mountjoy, 2019). Table 3.9 shows no statistically significant effect of A schools on any educational outcome. In turn, attending V schools seems to decrease the probability of graduating from high school and enrolling in a Bachelors program. Moreover, attending V schools appears to have no significant effect on the probability of enrolling in a technical or technological program.

While indicative, these effects can mask important heterogeneity between different fallback options. For instance, although there seems to be no effect of A schools, it is not necessarily the case that the effect A schools *versus* V is not statistically significant (or A *versus* H schools). Nonetheless, standard 2SLS does not allow the researcher to separately identify effects for these different combinations of the available school-type options. Last, Tables 3.9 and 3.10 also report F-statistics for diagnosing weak instruments, which are well above critical values for both regressions.

To illustrate the shortcomings of standard 2SLS in settings with multiple endogenous choices, suppose that there are only two possible types of schools (A and V) instead of three, and that we also have available different instruments of travel time to each option. When the instrument changes, there are four possible types of individuals: A-always-takers, V-always-takers, compliers, and defiers. For example, A-always-takers will choose to go to an A school regardless of travel time; in turn, compliers will respond to proximity changes by choosing A schools when they are relatively near. Defiers, in turn, will go to the furthest school. Both compliers and defiers provide variation to estimate treatment effects: they are types of students who would be induced to change their behavior by changes in the instrument. To identify treatment effects there is one additional assumption: monotonicity (Imbens and Angrist, 1994), which states that changes in the IVs must induce all agents to move in the same direction. This assumption would rule out defiers. 2SLS in this setup will provide estimates of LATEs.

With three school-types, and three continuous instruments, students induced to choose a particular school when the instruments change come from different initial states.[8] There are many possible combinations of students changing from different schools when the instrument changes. For example, students induced to choose an academic school when the instrument changes can stem from initial states of choosing vocational or hybrid schools. That is, students choosing A schools stem from different fallback options and define distinct complier groups (*e.g.*, switching from V towards A, or switching from H towards A). 2SLS does not directly identify different types of compliers, and the source of the variation that 2SLS uses to estimate the coefficients of interest is not straightforward.

The following section presents a theoretical framework as the basis for identifying complier populations, or more generally, individuals that are induced to change their behavior by changes in the instruments. Identification of counterfactuals and mean treatment effects under this framework are formalized in Section 3.4.1.

3.4 Framework of High School-type Choices

This section presents a utility maximization model in the context of schooling choices. The goal, using choice theory, is to determine school-type decisions at marginal changes of the instruments. It is a simplified framework that ignores potential wages associated with each option, peer effects, and general equilibrium effects. The main contribution of this framework is that it helps understand the potential behavior of students who would change their schooling choices as the

[8] The IV literature has recognized this challenge (Angrist et al., 1996; Heckman and Urzúa, 2010), such that in contexts of multiple unordered choices standard IV estimation provides effects that combine multiple margins of choice.

instruments change.

Let s be the school-type choice such that $s \in \{\text{A}, \text{V}, \text{H}\}$. Formally, I assume that students choose the school-type option that provides the highest latent utility, which I denote as U_{is}. The utility of each school-type s depends on unobservable preferences for schooling and training tracks, denoted as μ_{is}, and the cost of choosing option s, denoted as $\Lambda_{is}(Z_s)$. Hence, I assume that the cost of option s depends on travel time in hours and that the travel time to other options has no direct effect on the utility of school type s. In practice, this assumes that each option is only directly affected by one instrument.[9] School proximity has the role of cost-shifter, which means that living near a school-type reduces the cost of attendance. This assumption follows the literature of college choice in the U.S., as in Card (1995) and Mountjoy (2019). I focus here on the case with $\mathbf{Z} = \{Z_A, Z_V, Z_H\}$, where each Z_s is a continuous variable measuring the travel time to the closest school of type s.

$$U_{is} = \mu_{is} - \Lambda_{is}(Z_s)$$

The optimal choice of high school when instrument $Z_s = z_s$ can be defined as:

$$D_i(z_s) = \underset{s \in \{A,V,H\}}{\text{argmax}} \left(\mu_{is} - \Lambda_{is}(z_s)\right),$$

which means that the optimal schooling choice might change as the instrument Z_s

[9]This is a standard assumption in context with multiple options. It also has important implications for the estimation of treatment effects. Heckman and Urzúa (2010) show that when each instrument only affects one option, it is possible to identify Local Average Treatment Effects of each option *versus* the next-best.

changes. That is, $D_i(z_s)$ is a categorical variable that is equal to school choice $s \in \{A, V, H\}$ when student i faces instrument value $Z_s = z_s$. To formalize the assumptions on costs and utility:

$$\frac{\partial U_{is}}{\partial Z_s} \leq 0 \text{ and } \frac{\partial U_{is}}{\partial Z_{s_-}} = 0, \text{ for } t \in \{A, V, H\} \text{ and } s_- \neq s.$$

which states that as the travel time to s increases the utility of choice s decreases, given that the cost of option s increases. Also, changes in travel time to options other than s do not affect the utility of option s directly. Overall, the assumptions on how changes in travel time affect the school-type choice can be translated into students switching towards the relatively cheaper options. For instance, if travel time to option A increases, the utility of option A decreases while the utility of V and H remains unchanged. Hence, students who would change their schooling choices should do so towards options V and H and away from option A. Importantly, these are partial changes such that the travel time to other school-types remains fixed.

Building on the notion of partial changes in the instruments and their effect on the utility of the different options, I impose the assumption of *Partial Monotonicity* as proposed by Mountjoy (2019) and Mogstad et al. (2020a,b). Let $D_{is}(z, Z_{s^-})$ be a binary variable that takes the value of one if the student chooses option s when $Z_s = z$ and other instruments, Z_{s^-}, are fixed. In the context of high school-type choices, partial monotonicity states the following:

Partial Monotonicity - Let z, z' be two values in $\text{supp}(Z_s)$, with $Z_s, Z_{s^-} \in \mathbf{Z}$. For all $s \in \{A, V, H\}$ either $D_{is}(z, Z_{s^-}) \geq D_{is}(z', Z_{s^-})$ for all i, or $D_{is}(z, Z_{s^-}) \leq$

$D_{is}(z', Z_{s^-})$ for all i.

This assumption states that changes in one instrument, while the remaining instruments stay fixed, should induce agents towards (or against from) the same options. The following combinations of potential choices stem from the assumptions on changes in relative utility as the instruments change and partial monotonicity:

Table 3.1: Potential responses to a marginal decrease (*i.e.*, $z > z'$) in travel-time to:

Academic High Schools (Z_A)

Z$_A$	g_1^A	g_2^A	g_3^A	g_4^A	g_5^A
z_A	A	V	H	V	H
z'_A	A	V	H	A	A

Vocational High Schools (Z_V)

Z$_V$	g_1^V	g_2^V	g_3^V	g_4^V	g_5^V
z_V	A	V	H	A	H
z'_V	A	V	H	V	V

Hybrid High Schools (Z_H)

Z$_H$	g_1^H	g_2^H	g_3^H	g_4^H	g_5^H
z_H	A	V	H	A	V
z'_H	A	V	H	H	H

Regardless of the type of high school, g_1 denotes A high schools always-takers, g_2 denotes V high schools always-takers, and g_3 denotes H high schools always-takers. These three groups represent students who would not change their choice of school when travel time decreases. They do not provide variation to estimate the effects of school choice on educational attainment. On the other hand, groups four and five represent *compliers*, or students who would change their school choice towards the relatively cheaper option. That is,

g_4^A: students who would switch from V schools towards A schools, when the latter become relatively cheaper (due to lower travel time).

g_5^A: students who would switch from H schools towards A schools, when the latter become relatively cheaper (due to lower travel time).

g_4^V: students who would switch from A schools towards V schools, when the latter become relatively cheaper (due to lower travel time).

g_5^V: students who would switch from H schools towards V schools, when the latter become relatively cheaper (due to lower travel time).

g_4^H: students who would switch from A schools towards H schools, when the latter become relatively cheaper (due to lower travel time).

g_5^H: students who would switch from V schools towards H schools, when the latter become relatively cheaper (due to lower travel time).

These types of students provide variation to estimate the effects of school-type choice on educational outcomes. However, since each instrument induces two distinct complier groups, the researcher cannot uncover effects of pairwise comparisons across schools. Instead, at most (and without further assumptions), it is only possible to identify the effect of each option *versus* the next-best.

3.4.1 School-type effects:

A critical feature of Hybrid high schools is that they offer Academic and Vocational tracks. Rather than specializing in one type of training, Hybrid schools can offer either track. The lack of specialization implies that within H schools peers' composition largely differs from that in purely A or V schools. Therefore, comparisons between Hybrid schools and A or V schools would capture school-type effects, track effects, plus changes in class composition. To illustrate, in this subsection I present the set of effects that can be identified with the variation in each instrument. To start, let Y_s denote potential outcomes when students choose option $s \in \{A, V, H\}$. Observed outcomes, Y, result from a switching regression model:

$$Y = Y_A D_A + Y_V D_V + Y_H D_H,$$

where D_s are binary variables that denote the observed choice of school. Following the combinations of potential choices that result from shifts in travel time to A, V, and H in Table 3.1, the following effects can be identified:

$$\frac{E[Y|Z_A = z'] - E[Y|Z_A = z]}{E[D_A|Z_H = z'] - E[D_A|Z_H = z]} = \underbrace{\frac{E[Y_A - Y_V|D_V(z) = 1, D_A(z') = 1]\Pr(g_4^A)}{\Pr(g_4^A) + \Pr(g_5^A)}}_{\text{School-type}\times\text{ track}} + \underbrace{\frac{E[Y_A - Y_H|D_H(z) = 1, D_A(z') = 1]\Pr(g_5^A)}{\Pr(g_4^A) + \Pr(g_5^A)}}_{\text{School-type}\times\text{ track}\times\text{ class composition}},$$

which commonly denotes the effect of A high schools *versus* the next-best (*i.e.*, a mixture of V and H schools). $\Pr(g_4^A)$ and $\Pr(g_5^A)$ denote the share of compliers along the $V - A$ margin and $H - A$ margin, respectively. It is worth noting that while high school-type and choice of track are equivalent at A or V schools, within H schools there can be students in Academic or Vocational tracks. Appendix C shows how comparisons between A schools and H schools (the second term on the right of the equation above) would capture a combination of differences in school overall quality, type of track, plus changes in the composition of students. Given the lack of variation to identify effects for tracks within H schools, I interpret these effects as a combination of school quality, tracks, and composition.

Similar effects can be decomposed for changes in Z_V,

$$\begin{aligned}\frac{E[Y|Z_V = z'] - E[Y|Z_V = z]}{E[D_V|Z_V = z'] - E[D_V|Z_V = z]} = \\ &\frac{E[Y_V - Y_A|D_A(z) = 1, D_V(z') = 1]\Pr(g_4^V)}{\Pr(g_4^V) + \Pr(g_5^V)} \\ &+ \frac{E[Y_V - Y_H|D_V(z) = 1, D_H(z') = 1]\Pr(g_5^V)}{\Pr(g_4^V) + \Pr(g_5^V)},\end{aligned} \tag{3.3}$$

which captures a combination of local effects of V high schools *versus* A high schools, plus differences in school composition, track, and quality between V and H schools. Last,

$$\begin{aligned}\frac{E[Y|Z_H = z'] - E[Y|Z_H = z]}{E[D_H|Z_H = z'] - E[D_H|Z_H = z]} = \\ &\frac{E[Y_H - Y_A|D_A(z) = 1, D_H(z') = 1]\Pr(g_4^H)}{\Pr(g_4^H) + \Pr(g_5^H)} \\ &+ \frac{E[Y_H - Y_V|D_V(z) = 1, D_H(z') = 1]\Pr(g_5^H)}{\Pr(g_4^H) + \Pr(g_5^H)}\end{aligned}$$

which, rather than capturing changes in school-track, combines overall differences across all types of schools.

3.5 Estimation Strategy

This section presents the econometric strategy to identify complier groups and treatment effects with univariate 2SLS, using propensity scores and outcome by choice interactions (Mountjoy, 2019). This approach differs from a multivariate 2SLS (*i.e.*, where the first-stage consists of multiple endogenous choices) in that it allows the researcher to identify effects of one option *versus* the next-best alternative. I also employ the framework in Heckman and Pinto (2018) to identify complier shares and their average baseline (prior to the choice of high school) characteristics.

Let s be the school-type choice such that $s \in \{\mathrm{A}, \mathrm{V}, \mathrm{H}\}$. I estimate LATEs of option s *versus* the next-best with 2SLS where the outcome equation is defined as follows:

$$Y = \beta_0^s + \beta_s D_s + \beta_{s_-}^Z Z_{s_-} + \beta_X^Y X + \epsilon_Y,$$

where D_s is a binary variable that takes the value of one if the student chooses school s, Z_{s_-} is a vector of travel time in hours to schools other than s, and X contains baseline student characteristics, classroom and school characteristics, and local conditions. I analyze two types of outcomes: educational attainment in 2018 (three years after on-time high school graduation) and educational performance using SIMCE test scores for math and reading in 2013 (two years after starting high school). The latter are continuous standardized variables, and I define the former as binary variables: $Y_l = \mathbb{1}[\text{if the student reaches educational level } l \text{ by 2018}]$, where l refers to one of four mutually exclusive educational levels: High School Incomplete, High School Graduate but no Higher Education, Technical or Technological college, or Bachelors' program.

The first-stage is defined as follows:

$$D_s = \alpha_0^s + \alpha_A^s Z_A + \alpha_V^s Z_V + \alpha_H^s Z_H + \alpha_x^s X + \epsilon_s,$$

where Z_s denotes travel time in hours to school type $s \in \{A, V, H\}$. I assume that the instruments and unobserved components are independent, conditional on $\mathbf{X}$, that is $\mathbf{Z} \perp\!\!\!\perp (\epsilon_Y, \{\epsilon_s\}_{s \in \{A,V,H\}}) | \mathbf{X}$ and $\mathbf{Z} \perp\!\!\!\perp (\{Y_s\}_{s \in \{A,V,H\}}, \{D(z)\}_{z \in \text{supp}(\mathbf{Z})})$. Hence, β_s^{2SLS} denotes the LATE of s *versus* the next-best school and it is identified with the variation in instrument Z_s while the other instruments in $Z_{s\_}$ remain fixed.

3.5.1 Counterfactuals and subLATEs

The set of school-type effects in Section 3.4.1 are fully identified with the 2SLS approach in the previous section. For instance, β_A^{2SLS} can be expressed as:

$$\beta_A^{2SLS} = \underbrace{\frac{E[Y_A - Y_V | g_4^A] \Pr(g_4^A)}{\Pr(g_4^A) + \Pr(g_5^A)}}_{\text{subLATE}_{A-V}} + \underbrace{\frac{E[Y_A - Y_H | g_5^A] \Pr(g_5^A)}{\Pr(g_4^A) + \Pr(g_5^A)}}_{\text{subLATE}_{A-H}},$$

where g_4^A (g_5^A) denotes compliers who would be induced to change their school choice away from A and towards V (H) high schools, as travel time to A schools increases. Similar expressions of next-best LATEs using 2SLS can be found for V and H high schools. With one IV per endogenous choice it is not feasible to decompose β_A^{2SLS} into different subLATEs, without imposing additional assumptions.

One approach to disentangle LATEs for distinct complier groups is to assume that the average potential outcomes from their common choice would have been similar (Hull, 2018; Kline and Walters, 2016; Lee and Salanié, 2020, and Chapter 2 of this dissertation). For instance, in the context of compliers induced by the variation in Z_A homogeneity states that $E[Y_A|g_4^A] = E[Y_A|g_5^A]$. Although homogeneity is easy to implement, it can impose strong assumptions on agent's behavior. Nonetheless it is useful to, at least, bound LATEs for the different complier groups.

With the homogeneity assumption and 2SLS estimation, I use a choice-outcome interaction approach to identify choice-specific average counterfactuals for the different types of compliers induced by instruments in $\mathbf{Z}$. Using the average counterfactuals, I can recover subLATEs for different complier groups. Formally, for each Z_l, with $l \in \{A, V, H\}$, I estimate the following system of equations with a 2SLS regression:

$$YD_s = \beta_0^s + \beta_{sl}D_s + \beta_{l_-}^Z Z_{l_-} + \beta_X^Y X + \epsilon_Y^s \tag{3.4}$$

$$D_s = \alpha_0^s + \alpha_A^s Z_A + \alpha_V^s Z_V + \alpha_H^s Z_H + \alpha_x^s X + \epsilon_s, \tag{3.5}$$

for all $s \in \{A, V, H\}$. Hence, β_{sl}^{2SLS} denotes the average counterfactual of option s, for compliers induced to choose s due to changes in instrument Z_l. Next, I employ the 2SLS estimators to compute complier-specific subLATEs,

$$\text{subLATE}_{l-s} = \beta_{ll}^{2SLS} - \beta_{sl}^{2SLS},$$

for each $s, l \in \{A, V, H\}$. These subLATEs must be interpreted with caution. If, for instance, $E[Y_A|g_4^A] > E[Y_A|g_5^A]$, then subLATE_{A-V} would underestimate the effect of A *versus* V schools.

Last, I also identify the share of compliers induced by the variation in each instrument and their average baseline characteristics. Let ω_{l-s} denote the share of $l - s$ compliers induced byt the variation in Z_l. To identify ω_{l-s}, I use the first-stage results such that $\omega_{l-s} = \frac{\alpha_l^s}{\alpha_l^l}$ for each $l \neq s \in \{A, V, H\}$. The average baseline characteristics of each complier group can be estimated with 2SLS by substituting Y for X in equation 3.4.

All regressions include fixed effects of the region where the primary school of the student was located. For inference, I use cluster standard errors at the primary school level.

3.6 Results

The share of compliers, average of baseline characteristics, and their treatment effects are estimated using average changes in travel time of 0.2 hours (or about 12 minutes) for all school-types. This change would correspond to almost doubling the average travel time of students in my sample. Figure 3.7 present the shares of the different groups (*i.e.*, always-takers and compliers) induced by changes in Z_A (Panel (a)), Z_V (Panel (b)), and Z_H (Panel (c)). While the majority of students in my sample would not change their schooling choices if travel time to the nearest A school decreases, there is about 8.4% of students who would switch towards A schools if those where relatively closer. In specific, 5.8% of students would switch away from V schools and into A, while 2.6% would do so by substituting H for A schools. An interesting result emerges in Panel (b): the variation in travel time to the nearest V school only induces 10% of students to switch between H and into V schools, but there is no evidence of students switching between A and V schools when the latter are relatively cheaper. Last, Panel (c) shows that changes in travel time to H schools induce an almost equivalent share of students to change their choices from $A - H$ and $V - H$. The former represents 7.6% of the sample, the latter is at 6.8%. Overall, the total variation to estimate LATEs using travel time is between 8.4% (for Z_A) and 14.4% (for Z_H). All the shares are statistically significant at the 5% level.

Compliers along the $V - A$ and $H - A$ margin, induced by changes in Z_A, largely differ in their average characteristics at baseline. Figure 3.8 presents the average of baseline characteristics for always-takers and compliers associated with the variation in Z_A. Panel (a) shows differences in socio-economic characteristics. Compliers along the $V - A$ margin have a higher share of older male students, from low income households and with less educated mothers, than compliers along the $H - A$ margin. Panel (b) also shows that these complier groups differ in their baseline academic performance and classroom/school overall learning quality. On average, $H - A$ compliers seem to outperform $A - V$ compliers in reading, biology, and social science. However, $H - A$ compliers also show the highest average on the classroom activities index and on the learning environment index, compared to $V - A$ compliers.

The majority of $H - V$ compliers induced by the variation in Z_V, are young females, from low income households (Panel (a), Figure 3.9). Moreover, Panel (b) of Figure 3.9 shows that their average performance in SIMCE before entering high school is well below average in all academic subjects. They also have a low average in the classroom activities index, and worst learning environments than, for instance, A-always-takers.

In terms of changes in travel time to the nearest H schools, Panel (a) of Figure 3.10 shows that $A - H$ compliers have higher shares of older males, are more likely to stem from middle-to-high income households, and have more educated mothers, than compliers along the $V - H$ margin. Moreover, there are large difference in the baseline academic performance of these complier groups. Panel (b) shows that $V - H$ compliers under performed in all academic subjects, while $A - H$ compliers are near average in their test scores. It is also the case that the classroom learning activities and learning environment during primary were, on average, better for $V - H$ compliers than for $A - H$ compliers.

Table 3.11 present the average baseline characteristics of the municipalities where the students went to primary school. Panel (a) shows the average baseline local characteristics for groups induced by the variation in Z_A. On average, $H - A$ compliers stem from more urbanized settings and with lower poverty levels, than $V - A$ compliers. Panel (b) shows the average baseline local characteristics for groups induced by the variation in Z_V. These descriptive statistics show that $H - V$ compliers are the least urbanized, compared to always-takers. Last, Panel (c) shows the average baseline local characteristics for groups induced by the variation in Z_H. The average baseline urbanization level is lower among $A - H$ than among $V - H$ compliers; however, the former also have a lower average of unemployment rate and lower poverty levels.

While differences in baseline variables can to some extent explain the sorting patterns across different schools, they can also translate into differential gains or losses in terms of educational outcomes. Table 3.12 shows the results for LATEs of the different instruments. The first row of Panel (a) shows the LATE of A schools *versus* the next-best. On average, A schools decrease the probability of students completing primary but not enrolling in college by 22 percentage points compared to the next-best school, for compliers along $V - A$ and $H - A$ margins. After imposing the homogeneity assumption in Section 3.5.1, I recover the LATEs for $V - A$ and $H - A$ compliers. I observe that the negative effect of A schools on the probability of graduating high school but not enrolling in college is concentrated among $V - A$ compliers.[10] Among compliers, A schools also increase the probability of enrolling in an academic (*i.e.*, bachelor's) program (in 19 percentage points) and this effect seems to stem as well from $V - A$ compliers. The last two columns of Panel (a) show that there is no evidence of gains or losses in academic performance in SIMCE in 2013 from choosing A *versus* the next-best. It is worth noting that these are short-term effects on academic test-scores, since SIMCE 2013 took place two years after starting high school.

Given that I find no evidence of students switching along the $A - V$ margin when travel time to the nearest V school changes, the LATE of H schools *versus* the next-best might be largely driven by students who would switch out of H and into V schools. The second row of Panel (b) in Table 3.12 shows that, among $H - V$

[10] From the analysis on baseline characteristics of $V - A$ and $H - A$ compliers it is likely that the homogeneity assumption is overestimating the effect of A *versus* V schools.

compliers, choosing V schools increases the probability of completing high school without ever enrolling in college in six percentage points *versus* choosing H schools. The LATE of enrollment in academic programs for $H - V$ compliers is negative such that, compared to H schools, V schools decrease the probability of enrolling in academic programs in about 13 percentage points. In contrast, the results for reading and math test scores suggest that switching from H schools and into V schools when the latter are relatively cheaper might improve student's academic performance.

Last, Panel (c) shows that choosing an H school to study either an academic or vocational track, *versus* following those tracks in specialized schools (A or V) might improve the probability of completing high school and enrolling in academic programs in 2018. Importantly, this improvement seems to stem from $V - H$ compliers rather than from $A - H$ compliers. Among the former, on average, H schools *versus* V schools decrease the probability of completing high school and not entering college in 21 percentage points, while increasing the probability of enrolling in a bachelor's program in 28 percentage points. Given that $V - H$ compliers were below average in their baseline academic performance it is likely that the homogeneity assumption over estimates the effect of H schools.

In terms of short-term academic performance, the last two columns in Panel (c) show an average decrease among $A - H$ compliers in both reading and math test scores. In contrast, among $V - H$ compliers there are positive and large local effects in reading and math test scores. Differences in baseline characteristics across the $A - H$ and $V - H$ complier groups can help explain the effects I observe on academic achievement. To start, Figure 3.9 shows that academic performance of $A-H$ was close to the average while $V-H$ had overall lower average test scores for all academic subjects. The literature on peer effects has found some evidence on how high ability students can benefit from other high ability students, but low ability students could be harmed by having high- instead of low- ability peers (Sacerdote, 2011). Although I do not provide estimates of peer effects, I see a somewhat consistent pattern with previous literature: $A-H$ compliers are harmed by switching to a school with lower average test scores, and one of the mechanisms driving this effect could be the change in the composition of peers (along with overall school quality).

3.6.1 Robustness checks

In this Section, I present two robustness checks for the main results (LATEs of each high school choice *versus* the next-best). First, one of the potential threats to identification using travel time as instruments is that better connected areas (with shorter time of travel) might also have better services. Therefore, the exclusion restriction would not hold since fewer hours of travel could be correlated with services that can in turn determine educational outcomes for students. Although I cannot test the exclusion restriction, I present a robustness check in which I control for the proximity to amenities (following the approach in Kutscher et al. (2020)) such as hospitals, police stations, and libraries. Second, I also present a robustness check to different functional forms of the instruments. Given that travel time is a continuous variable I show how the results change when the instruments enter the regressions linearly, compared to the results with quadratic and cubic polynomials of the instruments.

3.6.1.1 Proximity to Amenities

The results I present in the previous section control for local (municipal) characteristics, such as the fraction of urban population, the unemployment rate, population density, and poverty level. Nonetheless, there can be other characteristics within municipalities that could be correlated with the instruments (travel time from primary school to high schools of each type) and the educational outcomes. This would represent a violation of the exclusion restriction and threatens the identification of the LATEs. Although I cannot test this restriction, I show here results controlling for proximity to amenities, such as hospitals, police stations, and libraries.

To compute the proximity to amenities I use data from the Open Street Maps (OSM) and the Open Source Routing Machine (OSRM) which contains geocoded information on police stations, hospitals, clinics, and libraries. I use the georeferenced information for these amenities and the coordinates of primary schools to compute the minimum distance from the latter to the different amenities. Rather than using the proximity to amenities as controls in the main specification, I use them to check the robustness of the results because the earliest OSM data for Chile corresponds to 2014 (three years after the high school choice takes place for my sample).

Table 3.13 compares the results of the LATEs for each option *versus* the next-best for the main specification and controlling for proximity to amenities. Overall, the direction of the effects remains the same as well as the statistical significance (except for the LATE of V *versus* the next-best on math test scores). Panel (a) shows a difference of three percentage points between the LATEs estimated with the main specification and those including amenities. Panel (b) shows that for high school incomplete the LATE with the main specification and the LATE controlling for amenities are virtually the same. In turn, for high school completion and enrolling in a bachelor's program, the results for Vocational schools from the main specification are almost half as those with the controls for amenities. The opposite pattern emerges in Panel (c) for Hybrid schools, where the amenities controls reduce the LATEs by almost half.

Overall, while there are differences in magnitude for some of the LATEs, the significance and direction of the effects do not change. Importantly, the differences in the size of the effects could reflect measurement errors as well as differences in the timing of the school choice and the data on amenities.

3.6.1.2 Functional form of the IVs

The main specification in Section 3.5 imposes a linear relation between travel time and the choice of high school. However, the negative effect of an increase in travel time on choosing a particular school can increase or become less negative as travel time increases. To account for nonlinearities in the relation between travel time and high school choices, I estimate the LATEs with quadratic and cubic functional forms of the instruments.

Table 3.14 shows the first-stage results with different functional forms of travel time to each option. First, most of the terms of he quadratic and cubic functional forms are statistically significant which supports the argument of a nonlinear relation between travel time and choices. For instance, the second column shows a negative and decreasing effect of travel time to Academic schools and the probability of choosing Academic high schools. This would imply a convex probability of choosing A schools as a function of travel time. The same pattern holds for the choice of Vocational and Hybrid high schools: increases in travel time to these options decrease the likelihood of choosing them, but this effect decreases as travel time becomes larger.

Table 3.15 presents the estimates of LATEs of each option *versus* the next-best using different functional forms of the instruments. Panel (a) shows small differences in the LATEs estimates for Academic schools; the direction and significance of the effects also remains relatively the same. Hence, the LATEs of Academic schools seem to not be sensitive to the specification of the functional form of the instruments. Similarly, the LATEs for Hybrid schools also remain largely robust to quadratic and cubic specifications of travel time (Panel (c)). The results in Panel (c) show virtually no differences in terms of direction, size, and significance.

Panel (b) in turn shows more drastic changes in the estimated LATEs for Vocational high schools, except for the outcome of High School Incomplete. The effects on High School completion and on enrollment in a bachelor's program lose significance and become negligible in magnitude, for quadratic and cubic functional forms of travel time. As such, the effect of Vocational schools on educational attainment seems to be more detrimental than with the main specification.

3.7 Concluding Remarks

In this chapter, I estimate the effect of choosing a type of high school on student's academic performance and educational trajectories. I study the context of high school choice in Chile, where students can choose from three types of schools: Academic, Vocational, or Hybrid. The latter type is particularly interesting, since it offers both tracks (academic or vocational) allowing for a higher heterogeneity in class composition (at least in terms of track preference) than specialized schools. Hence, differences in student's outcomes between types of schools are likely to reflect heterogeneity in class composition as well as differences in school quality and type of track. In addition, students self-select into their preferred training track and school type which complicates the identification of school-type effects.

I use administrative data at the student level that links academic trajectories, high school choices, and college enrollment, and exploit variation in the travel time to each type of school. Using an Instrumental Variables (IVs) approach and a latent utility framework I show how, under some assumptions, it is possible to identify effects of Academic schools *versus* the next-best (*i.e.*, a mixture of Vocational and Hybrid), and Vocational *versus* Hybrid schools. Furthermore, I show that the interpretation of these effects warrants caution: it is a combination of type of training (vocational or academic) plus changes in other school characteristics and class composition.

My main findings show that Academic schools have positive local effects on the probability of completing high school and enrolling in Bachelor's programs, *versus* the next-best. These effects are estimated using the variation in travel time to Academic schools, and about 8.4% of students in my sample would switch towards Academic schools when travel time to this option decreases. Also, I find that Vocational schools have a positive LATE on the probability of completing high school, but decrease the probability of enrolling in academic colleges, than Hybrid Schools. To estimate these effects I use the variation in travel time to Vocational schools, which induces about 10% of students to change their school choice from Hybrid and towards Vocational schools, as travel time decreases.

Last, I uncover positive and significant effects of Hybrid schools on the probability of completing high school and enrolling in an academic college. Importantly, most of this effect is driven by students who would change their choice from Vocational schools and towards Hybrid schools, as the travel time to Hybrid schools decreases. Neither Vocational nor Hybrid schools seem to have an effect in the probability of enrolling in a vocational college. While I observe that Vocational schools might increase the probability of enrolling in vocational schools, this effect is not statistically significant and it is small in magnitude (of only 2.2 percentage points). Overall, it seems that Vocational schools can help students complete high school but have no effect in inducing them to continue their studies.

These results can place an excessive penalty on vocational training since its objective is fostering the school-to-work transition rather than the school-to college transition. In addition, a limitation of my analysis is that I lack data on labor market participation and earnings which might differ across students in each type of school. Nonetheless, my results also suggest that some students might benefit from choosing a vocational track in a Hybrid school rather than in a Vocational school, in particular students who might have a preference for pursuing higher education studies.

3.8 Tables and Figures

Table 3.2: Summary Statistics

Variable	**Mean**	**SD**
Female	0.505	0.500
Age in 2011	14.287	0.588
Household Characteristics		
Size	3.547	1.064
Income level=1	0.125	0.331
Income level=2	0.359	0.480
Income level=3	0.223	0.416
Income level=4	0.293	0.455
Mother's highest level of education		
No education	0.056	0.230
Primary	0.302	0.459
Incomplete Secondary	0.170	0.376
High School - Vocational	0.117	0.321
High School - Academic	0.210	0.407
Higher Education	0.145	0.352
Local (comuna) characteristics, 2011		
Fraction of urban population	0.834	0.263
Unemployment rate	0.081	0.036
Population Density	2110.485	3517.109
Poverty rate	0.153	0.066
Educational status, from 2015-2018		
High School Incomplete	0.170	0.375
High School Graduate, No Higher Education	0.205	0.403
Higher Education: Technical/Technological	0.268	0.443
Higher Education: Bachelors	0.358	0.479
School-type choice, 2012		
Academic (A)	0.332	0.471
Vocational (V)	0.339	0.474
Hybrid (H)	0.329	0.470
Academic Track	0.599	0.490
N	89,147	

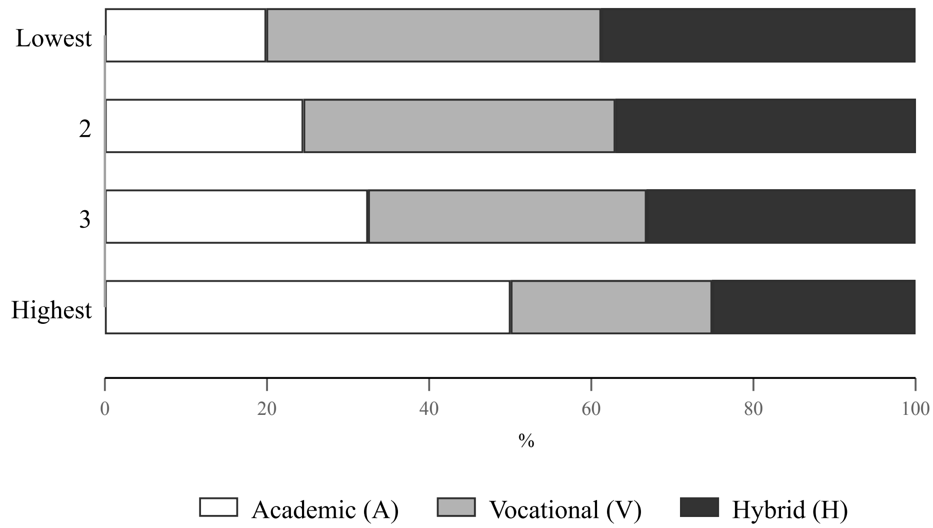
Note: The data corresponds to the universe of students entering first grade of High School in 2012, who took the National Standardized Exam (SIMCE, in Spanish) in 2011 (during the last grade of Primary School). The battery of socioeconomic characteristics and standardized test scores come from the National Standardized Exam (SIMCE, in Spanish) in 2011. Local (comuna, in Spanish) characteristics come from the National Household Survey (CASEN, in Spanish) in 2011.

Table 3.3: Summary Statistics, by type of school

Variable	**Academic (A)**	**Vocational (V)**	**Hybrid (H)**
Female	0.556	0.456	0.504
Age in 2011	14.224	14.304	14.333
Household Characteristics			
Size	3.600	3.521	3.522
Income level=1	0.075	0.152	0.147
Income level=2	0.265	0.408	0.404
Income level=3	0.218	0.226	0.225
Income level=4	0.442	0.215	0.223
Mother's highest level of education			
No education	0.047	0.059	0.062
Primary	0.191	0.352	0.363
Incomplete Secondary	0.133	0.193	0.184
High School - Vocational	0.133	0.115	0.101
High School - Academic	0.231	0.199	0.200
Higher Education	0.265	0.082	0.089
Standardized Test Scores - last grade of Primary			
Reading	0.336	-0.129	-0.095
Math	0.320	-0.130	-0.126
Biology	0.338	-0.131	-0.112
Social science	0.337	-0.141	-0.142
Local (comuna) characteristics, 2011			
Fraction of urban population	0.862	0.839	0.802
Unemployment rate	0.082	0.086	0.076
Population Density	2039.364	1890.093	2409.688
Poverty rate	0.150	0.159	0.151
Clasroom/School learning			
Activities Index (best to worst)	-.0022	.001	0.001
Environment Index (worst to best)	0.158	-0.057	-0.101
Standardized Test Scores - 2nd grade of High School [Obs.]			
Reading [61,629]	0.361	-0.164	-0.184
Math [61,619]	0.427	-0.194	-0.215
Educational Attainment			
HS Incomplete	0.112	0.186	0.212
HS Graduate, No Higher Education	0.106	0.270	0.236
HE Enrolled - Technical/Technological	0.212	0.300	0.290
HE Enrolled - Bachelors degree	0.569	0.245	0.262
N	29,576	30,255	29,316
%	0.332	0.339	0.329

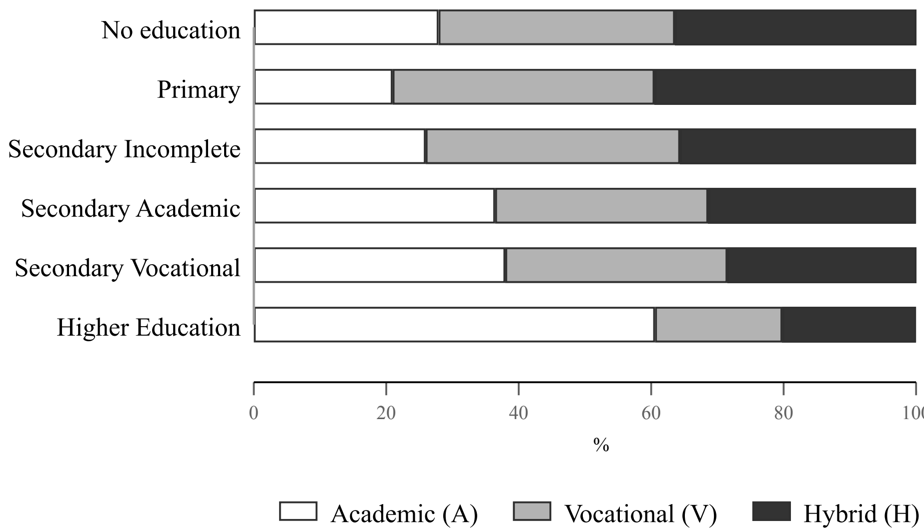
Note: The data corresponds to the universe of students entering first grade of High School in 2012, who took the National Standardized Exam (SIMCE, in Spanish) in 2011 (during the last grade of Primary School). High School choices take place between 2011 and 2012. Clasroom/School Activities Index is a composite of learning activities performed by Math Teachers during the last grade of primary school; Clasroom/School Environment Index is a composite of violence, bullying, care for facilities reported by Math teachers during the last grade of primary school. Educational Attainment corresponds to the student status in a window of three years after on time high school graduation (*i.e.*, from 2015-2018).

Figure 3.1: Household Income Level, at Baseline



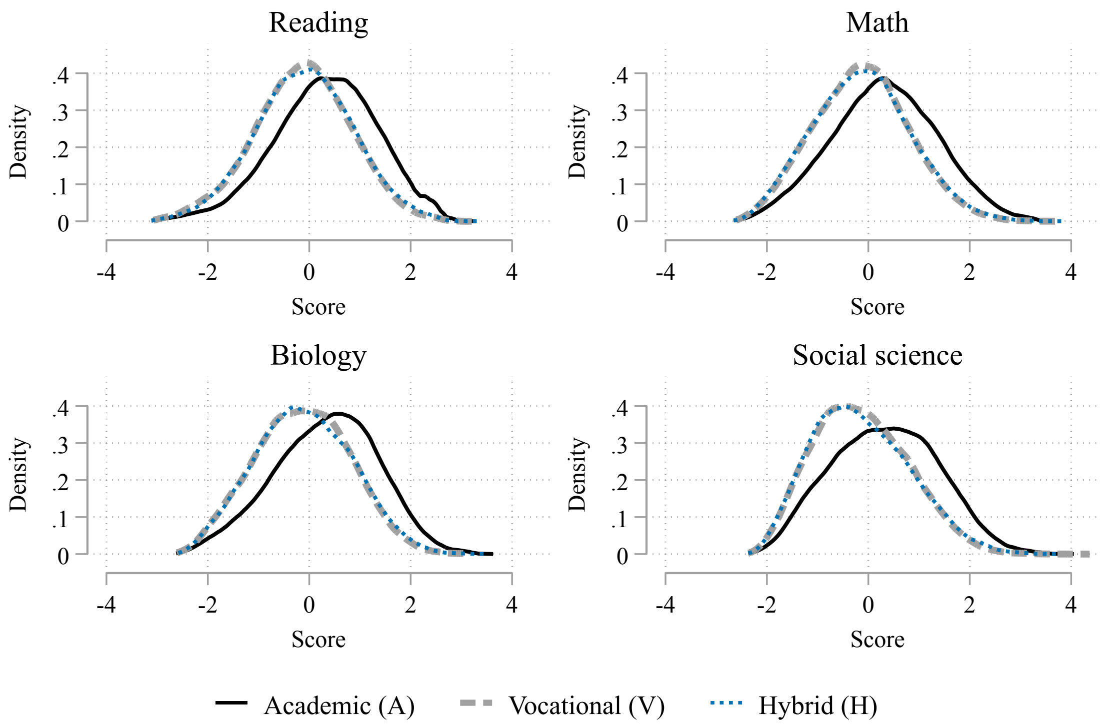
Note: The figure illustrates the distribution of high school choices in 2012 at different levels of household income (measured in 2011). The white bars correspond to Academic (A) High Schools, the grey bars to Vocational (V) High Schools, and the black bars to Hybrid (H) High Schools.

Figure 3.2: Mother's Level of Education, at Baseline



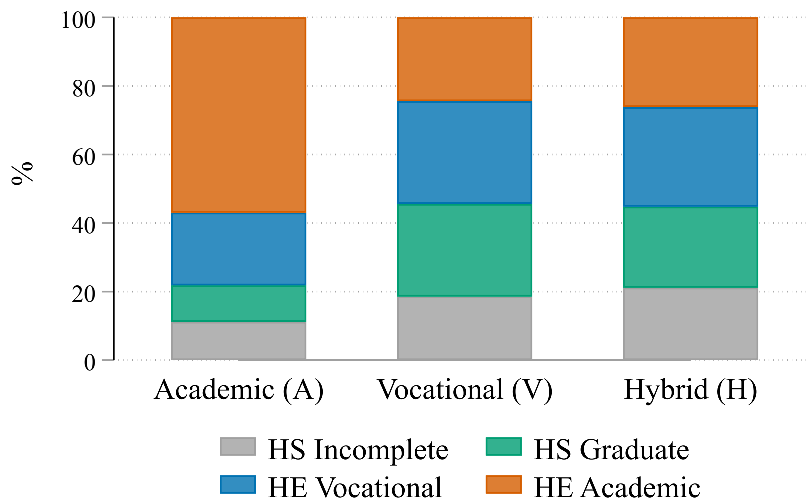
Note: The figure illustrates the distribution of high school choices in 2012 at different levels of education of the student's mother (measured in 2011). The white bars correspond to Academic (A) High Schools, the grey bars to Vocational (V) High Schools, and the black bars to Hybrid (H) High Schools.

Figure 3.3: Test Scores, at Baseline



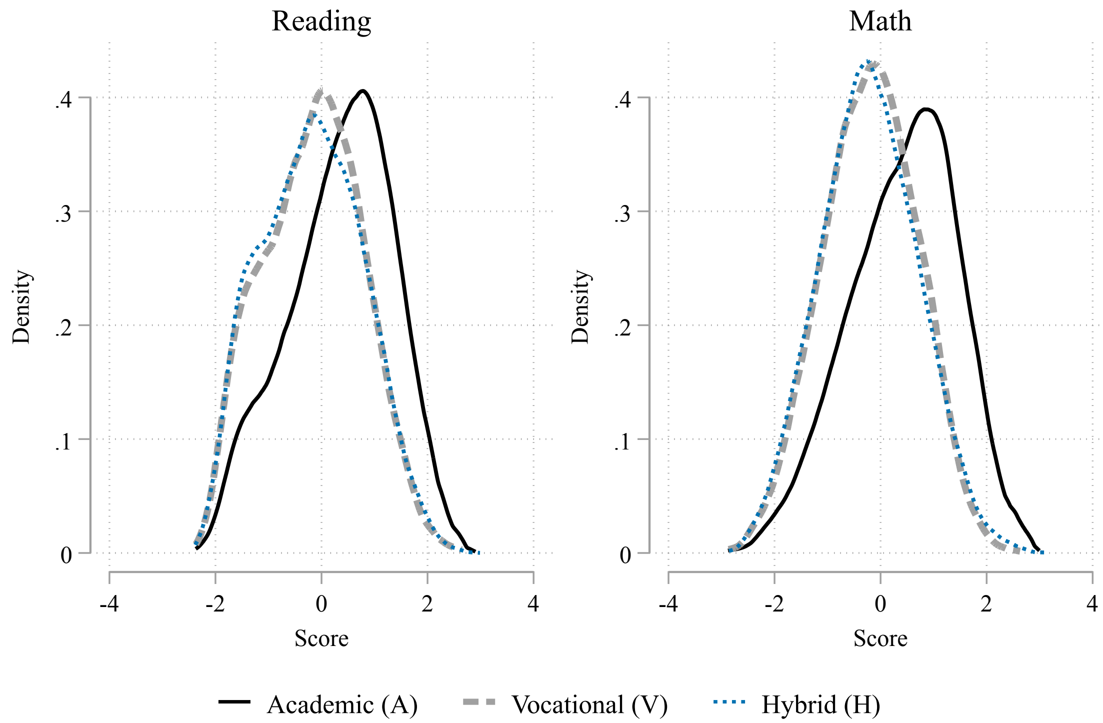
Note: The figure illustrates the distribution of standardized test scores on the SIMCE exam in 2011, during the last year of primary school, by school-type choice. The black solid lines correspond to Academic (A) High Schools, the grey dash lines to Vocational (V) High Schools, and the blue dots to Hybrid (H) High Schools.

Figure 3.4: Educational Trajectory after High School, by type of school



Note: The figure illustrates educational attainment after high school graduation, by school-type choice. Each bar represents a school-type choice. The grey (bottom) area corresponds to High School incomplete, which refers to students who did not graduate High School between 2015 and 2018 or in a window of three years after on-time high school graduation. The green area (second from the bottom to the top) represents the share of students who graduate from High School. The blue area (third from the bottom to the top) is the share of students who enroll in a Higher Education Institution in a Technical or Technological program The last bar (in red) represents the share of students who enroll in Bachelors' programs.

Figure 3.5: Test Scores, at 2nd grade of High School (2013)



Note: The figure illustrates the distribution of test scores in the SIMCE two years after starting High School, by school-type choice. The black solid lines correspond to Academic (A) High Schools, the grey dash lines to Vocational (V) High Schools, and the blue dots to Hybrid (H) High Schools.

Table 3.4: School Characteristics, by type of school

Variable	**Academic (A)**	**Vocational (V)**	**Hybrid (H)**	**All**
Rural	0.036	0.180	0.061	0.062
Private - subsidized	0.663	0.493	0.443	0.592
Private	0.156	0.005	0	0.101
Average number of students	17.626	82.439	53.013	34.314
N	1,678	367	553	2,598
%	64.58	14.13	21.29	

Note: school level data from the sample of students entering first grade of High School in 2012, who took the National Standardized Exam (SIMCE, in Spanish) in 2011 (during the last grade of Primary School). Private - subsidized *versus* private schools differ in the amount of tuition they can charge and their financing; the former can be publicly and privately financed and charge lower tuition than fully private schools.

Table 3.5: OLS Results: Educational Trajectory

Variable	**HS Incomplete**	**HS Graduate No Higher Educ.**	**HE: Enrolled in Technical/Technological**	**HE: Enrolled in Bachelor's program**
Choice: Academic (A)	-0.0140***	-0.0774***	-0.0575***	0.1490***
	(0.0034)	(0.0037)	(0.0043)	(0.0044)
Choice: Vocational (V)	-0.0212***	0.0351***	0.0102**	-0.0241***
	(0.0033)	(0.0038)	(0.0042)	(0.0039)
Female	-0.0561***	-0.0419***	0.0199***	0.0781***
	(0.0025)	(0.0029)	(0.0033)	(0.0031)
Age	0.1797***	-0.0239***	-0.0302***	-0.1256***
	(0.0027)	(0.0023)	(0.0026)	(0.0021)
Household size	-0.0045***	0.0048***	0.0010	-0.0013
	(0.0012)	(0.0013)	(0.0015)	(0.0013)
Income level=2	-0.0336***	-0.0195***	0.0250***	0.0280***
	(0.0045)	(0.0051)	(0.0052)	(0.0042)
Income level=3	-0.0590***	-0.0449***	0.0394***	0.0645***
	(0.0050)	(0.0054)	(0.0058)	(0.0049)
Income level=4	-0.0716***	-0.0591***	0.0213***	0.1094***
	(0.0050)	(0.0054)	(0.0059)	(0.0052)
Mother's highest level of education				
Primary	-0.0009	0.0257***	-0.0084	-0.0164***
	(0.0059)	(0.0064)	(0.0068)	(0.0061)
Secondary incomplete	-0.0132**	0.0024	0.0060	0.0048
	(0.0062)	(0.0069)	(0.0072)	(0.0066)
High School - Vocational	-0.0545***	-0.0123*	0.0129*	0.0539***
	(0.0063)	(0.0071)	(0.0077)	(0.0072)
High School - Academic	-0.0386***	-0.0215***	0.0183***	0.0418***
	(0.0060)	(0.0066)	(0.0071)	(0.0066)
Higher Education	-0.0509***	-0.0337***	-0.0293***	0.1138***
	(0.0062)	(0.0068)	(0.0076)	(0.0073)
Reading, SD test score	-0.0164***	-0.0096***	-0.0068***	0.0328***
	(0.0020)	(0.0021)	(0.0024)	(0.0023)
Math, SD test score	-0.0139***	-0.0234***	-0.0275***	0.0649***
	(0.0018)	(0.0020)	(0.0022)	(0.0021)
Biology, SD test score	-0.0321***	-0.0012	-0.0059**	0.0391***
	(0.0021)	(0.0023)	(0.0025)	(0.0024)
Social Science, SD test score	-0.0208***	-0.0226***	-0.0151***	0.0585***
	(0.0018)	(0.0020)	(0.0022)	(0.0023)
Local characteristics				
Fraction of urban pop.	0.0192***	-0.0403***	-0.0042	0.0253***
	(0.0067)	(0.0064)	(0.0078)	(0.0081)
Unemployment rate	0.0939*	-0.0452	0.0126	-0.0613
	(0.0544)	(0.0571)	(0.0690)	(0.0700)
Population Density	0.0000	-0.0000***	-0.0000	0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Poverty level	-0.0105	0.0212	0.0530	-0.0636
	(0.0319)	(0.0347)	(0.0397)	(0.0395)
Classroom/School learning				
Activities Index	-0.0058***	-0.0008	-0.0001	0.0066***
(best to worst)	(0.0015)	(0.0015)	(0.0017)	(0.0019)
Environment Index	-0.0073***	-0.0007	0.0046***	0.0033*
(worst to best)	(0.0017)	(0.0016)	(0.0018)	(0.0019)
Teacher's sex=female	-0.0072**	-0.0049	0.0079**	0.0042
	(0.0030)	(0.0032)	(0.0035)	(0.0037)
Constant	-2.2722***	0.6576***	0.5923***	2.0222***
	(0.0421)	(0.0380)	(0.0429)	(0.0363)
N	89,147	89,147	89,147	89,147
R^2	0.1555	0.0641	0.0289	0.2828

Table 3.6: OLS Results: Test Scores in 2nd grade of High School

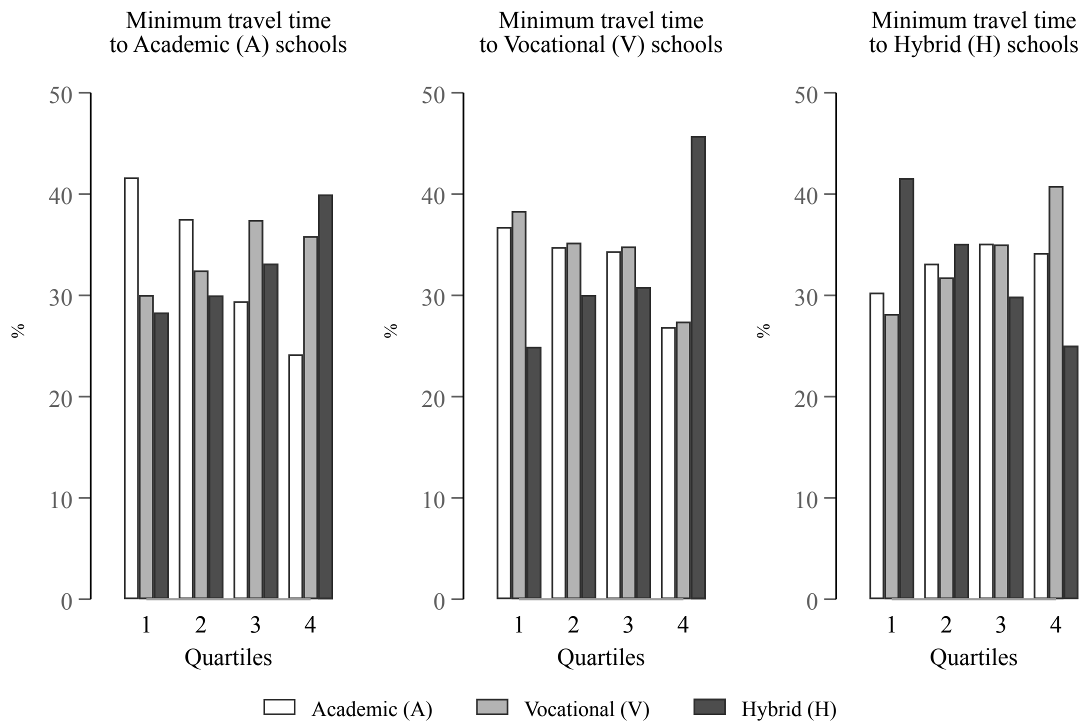
Variable	Reading	Math
Choice: Academic (A)	0.1666***	0.2326***
	(0.0098)	(0.0101)
Choice: Vocational (V)	0.0663***	0.0423***
	(0.0097)	(0.0091)
Female	0.1849***	-0.0742***
	(0.0069)	(0.0064)
Age	-0.0751***	-0.0674***
	(0.0063)	(0.0060)
Household size	0.0069**	0.0109***
	(0.0029)	(0.0027)
Income level=2	0.0083	0.0283***
	(0.0103)	(0.0095)
Income level=3	0.0220*	0.0588***
	(0.0115)	(0.0108)
Income level=4	0.0454***	0.1166***
	(0.0118)	(0.0109)
Mother's highest level of education		
Primary	-0.0039	-0.0325**
	(0.0137)	(0.0127)
Secondary incomplete	-0.0022	-0.0036
	(0.0147)	(0.0136)
High School - Vocational	0.0250	0.0316**
	(0.0159)	(0.0142)
High School - Academic	0.0175	0.0110
	(0.0144)	(0.0132)
Higher Education	0.0721***	0.0884***
	(0.0156)	(0.0143)
Reading, SD test score	0.3604***	0.0593***
	(0.0057)	(0.0049)
Math, SD test score	0.0828***	0.4723***
	(0.0049)	(0.0047)
Biology, SD test score	0.1534***	0.1566***
	(0.0056)	(0.0055)
Social Science, SD test score	0.1441***	0.0665***
	(0.0049)	(0.0045)
Local characteristics		
Fraction of urban pop.	-0.0444**	0.0587***
	(0.0190)	(0.0183)
Unemployment rate	-0.0599	-0.3251**
	(0.1646)	(0.1600)
Population Density	0.0000	0.0000**
	(0.0000)	(0.0000)
Poverty level	-0.0022	-0.1952**
	(0.1008)	(0.0975)
Classroom/School learning		
Activities Index (best to worst)	0.0003	0.0023
	(0.0046)	(0.0045)
Environment Index (worst to best)	-0.0322***	-0.0055
	(0.0049)	(0.0044)
Teacher's sex=female	-0.0179*	0.0032
	(0.0094)	(0.0088)
Constant	0.7470***	0.6850***
	(0.1034)	(0.0985)
N	61,629	61,619
R^2	0.4638	0.5575

Table 3.7: Average travel time (in hours) to school-types, by school choice

Variable	Academic (A)	Vocational (V)	Hybrid (H)	All
Academic (A)	0.057	0.081	0.103	0.079
Vocational (V)	0.123	0.115	0.194	0.144
Hybrid (H)	0.104	0.114	0.077	0.099
N	29,576	30,255	29,316	104,773

Note: Average travel time in hours and by car from the student's primary school to the nearest high school by type. The travel time is computed using OpenStreetMap which contains historical information on roads since 2014, only two years after the choice of high school takes place for the cohort of interest.

Figure 3.6: School-type enrollment, at quartiles of travel time (hours) to each school



Note: The figure illustrates the distribution of school-type choices by quartiles of travel time (in hours) to each type of school. The white bars correspond to Academic (A) high schools, the grey bars to Vocational (V) high schools, and the black bars to Hybrid (H) high schools.

Table 3.8: First-stage: School-type choice

Variable	Academic (A)	Vocational (V)	Hybrid (H)
Hours to closest Academic (A) high school	-0.2972***	0.2337***	0.0635
	(0.0316)	(0.0291)	(0.0399)
Hours to closest Vocational (V) high school	-0.0097	-0.4711***	0.4807***
	(0.0255)	(0.0223)	(0.0299)
Hours to closest Hybrid (H) high school	0.3557***	0.3236***	-0.6793***
	(0.0342)	(0.0266)	(0.0380)
Baseline Characteristics			
Female	0.0832***	-0.0725***	-0.0107***
	(0.0038)	(0.0044)	(0.0041)
Age	-0.0353***	0.0072**	0.0281***
	(0.0027)	(0.0031)	(0.0031)
Household size	-0.0036**	0.0026*	0.0010
	(0.0015)	(0.0015)	(0.0016)
Income level=2	0.0187***	-0.0134**	-0.0052
	(0.0048)	(0.0056)	(0.0056)
Income level=3	0.0682***	-0.0385***	-0.0297***
	(0.0058)	(0.0063)	(0.0063)
Income level=4	0.1729***	-0.1010***	-0.0719***
	(0.0062)	(0.0067)	(0.0065)
Mother's educ=Primary	-0.0527***	0.0377***	0.0150**
	(0.0067)	(0.0075)	(0.0072)
Mother's educ=Secondary Incomplete	-0.0276***	0.0282***	-0.0006
	(0.0073)	(0.0079)	(0.0076)
Mother's educ= High School Vocational	0.0273***	0.0088	-0.0361***
	(0.0080)	(0.0085)	(0.0083)
Mother's educ= High School Academic	0.0395***	-0.0140*	-0.0255***
	(0.0072)	(0.0076)	(0.0075)
Mother's educ= Higher Education	0.1873***	-0.0963***	-0.0910***
	(0.0085)	(0.0085)	(0.0081)
SD Test Scores: Reading	0.0157***	-0.0141***	-0.0016
	(0.0026)	(0.0027)	(0.0028)
SD Test Scores: Math	0.0265***	-0.0109***	-0.0156***
	(0.0026)	(0.0026)	(0.0026)
SD Test Scores: Biology	0.0195***	-0.0140***	-0.0055*
	(0.0029)	(0.0032)	(0.0032)
SD Test Scores: Social Science	0.0387***	-0.0157***	-0.0229***
	(0.0025)	(0.0025)	(0.0027)
Fraction of urban pop.	0.0946***	0.0503***	-0.1449***
	(0.0154)	(0.0150)	(0.0168)
Unemployment rate	-0.0542	0.4289***	-0.3747**
	(0.1403)	(0.1433)	(0.1848)
Population Density	-0.0000***	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
Poverty level	-0.2843***	-0.1816**	0.4659***
	(0.0816)	(0.0790)	(0.0878)
Classroom/School: Activities Index	0.0054	-0.0013	-0.0041
	(0.0035)	(0.0033)	(0.0039)
Classroom/School: Environment Index	0.0199***	-0.0075**	-0.0124***
	(0.0030)	(0.0034)	(0.0037)
Teacher's sex=female	0.0024	-0.0078	0.0055
	(0.0067)	(0.0068)	(0.0077)
Constant	0.6263***	0.1987***	0.1750***
	(0.0513)	(0.0545)	(0.0595)
N	89,147	89,147	89,147
R^2	0.1670	0.0896	0.1239

Note: first-stage regressions of high school choices on the instruments (*i.e.*, travel time) and baseline characteristics. Each choice is estimated separately.

Table 3.9: 2SLS: Educational Trajectory

Variable	**HS Incomplete**	**HS Graduate No Higher Educ.**	**HE: Enrolled in Technical/Technological**	**HE: Enrolled in Bachelor's program**
Choice: Academic (A)	-0.0049	0.0044	0.0253	-0.0248
	(0.0258)	(0.0354)	(0.0339)	(0.0361)
Choice: Vocational (V)	0.0428**	0.0568**	0.0030	-0.1027***
	(0.0189)	(0.0221)	(0.0265)	(0.0265)
N	89,147	89,147	89,147	89,147
R^2	0.1501	0.0579	0.0217	0.2639
Cragg-Donald F-stat	253.92			
Kleibergen-Paap F-stat	48.485			

Cluster standard errors at the primary school level are in parentheses. All estimations include regional fixed effects. Each outcome is estimated separately.
$*p < 0.1, **p < 0.05, ***p < 0.01$

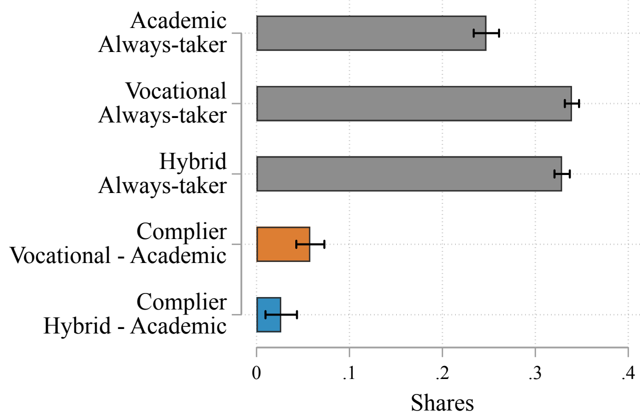
Table 3.10: 2SLS: Educational Achievement

Variable	**Reading**	**Math**
Choice: Academic (A)	0.0382	-0.1519
	(0.1034)	(0.1131)
Choice: Vocational (V)	0.1159*	0.1269*
	(0.0649)	(0.0663)
N	55,006	55,012
R^2	0.4675	0.5291
Cragg-Donald F-stat	185.118	184.234
Kleibergen-Paap F-stat	39.537	38.999

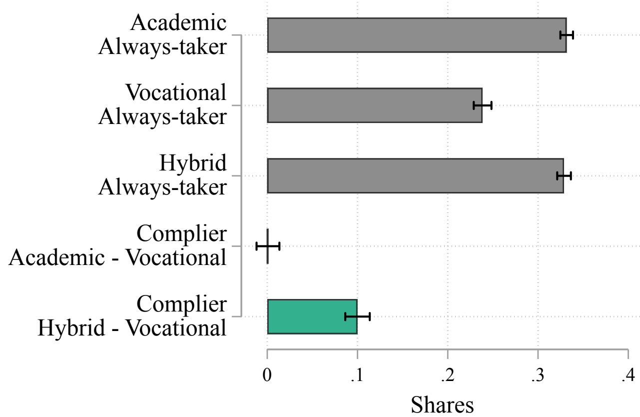
Cluster standard errors at the primary school level are in parentheses. All estimations include regional fixed effects. Each outcome is estimated separately.
$*p < 0.1, **p < 0.05, ***p < 0.01$

Figure 3.7: Response groups to the variation in travel time to each type of school

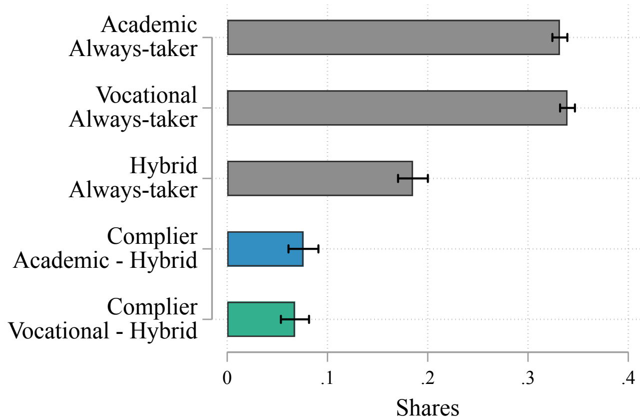
(a) Travel time to Academic (A) High Schools



(b) Travel time to Vocational (V) High Schools



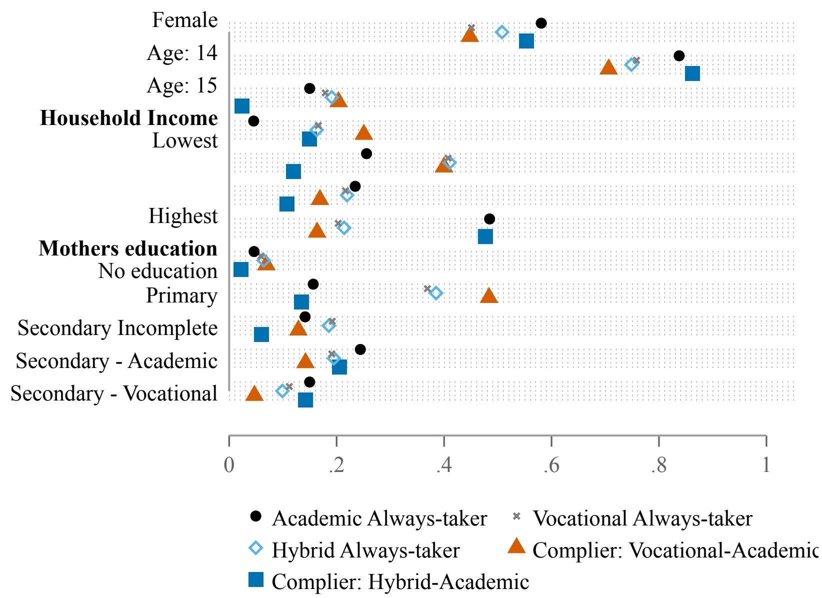
(c) Travel time to Hybrid (H) High Schools



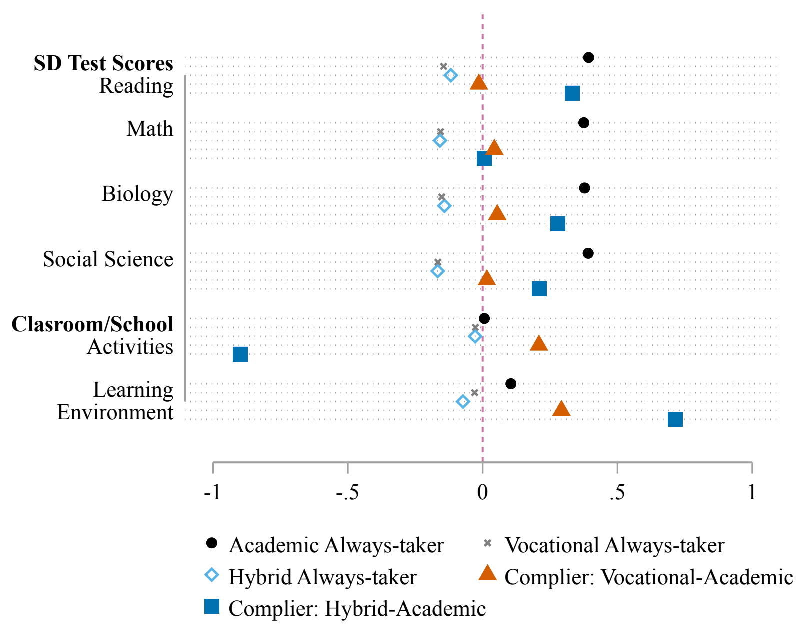
Note: The figure shows the estimated probability of always-takers and compliers induced by the variation of travel time to each school-type. Panel (a) presents the probabilities of groups induced by the travel time to the nearest Academic (A) high school. Panel (b) presents the probabilities of groups induced by the travel time to the nearest Vocational (V) high school. Panel (c) presents the probabilities of groups induced by the travel time to the nearest Hybrid (H) high school. All figures show confidence intervals at the 95% level.

Figure 3.8: Average characteristics of response groups due to the variation in travel time to Academic (A) High Schools, in 2011 (one year before choosing a school-type)

(a) Average of socio-economic characteristics



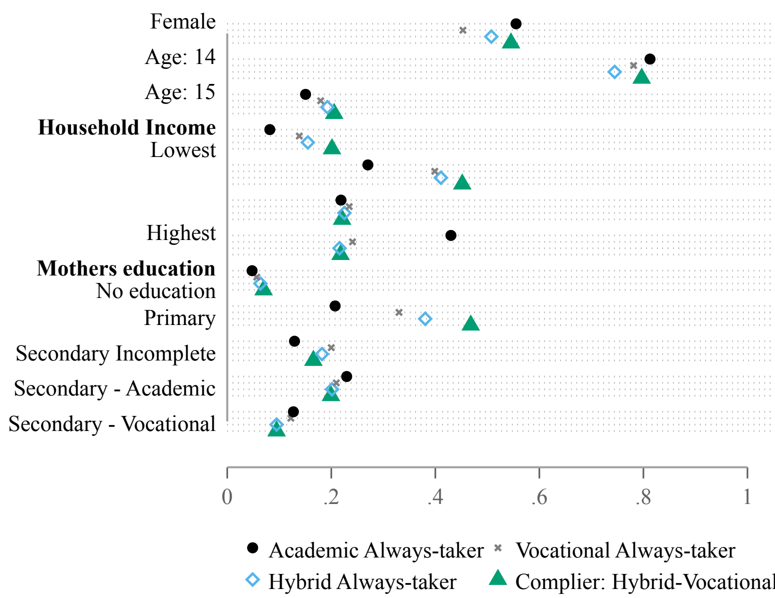
(b) Average of baseline test scores on SIMCE in 2011 and classroom/school quality



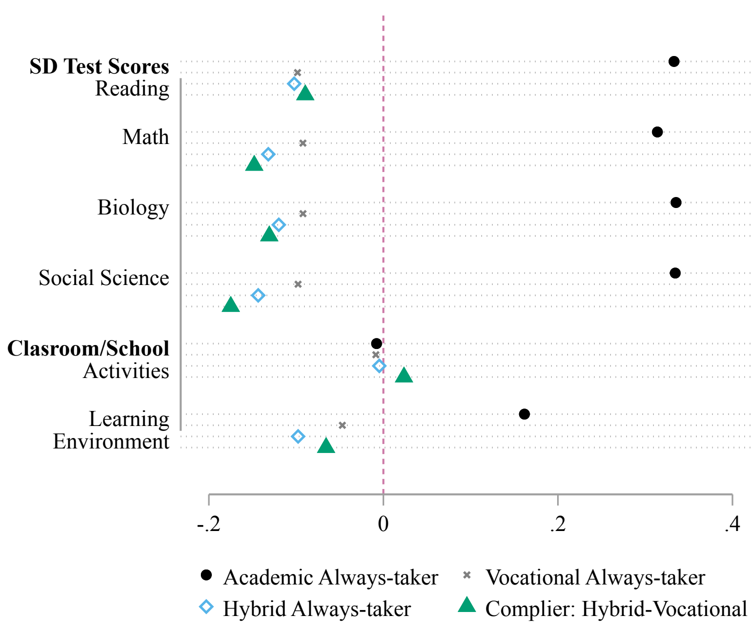
Note: The figure shows the average of baseline characteristics, for the groups of always-takers and compliers induced by the variation in travel time (hours) to the nearest Academic (A) high school. Panel (a) presents the estimated averages of socio-economic characteristics, while Panel (b) focuses on baseline academic achievement and clasroom/school learning activities and environment.

Figure 3.9: Average characteristics of response groups due to the variation in travel time to Vocational (V) High Schools, in 2011 (one year before choosing a school-type)

(a) Average of socio-economic characteristics



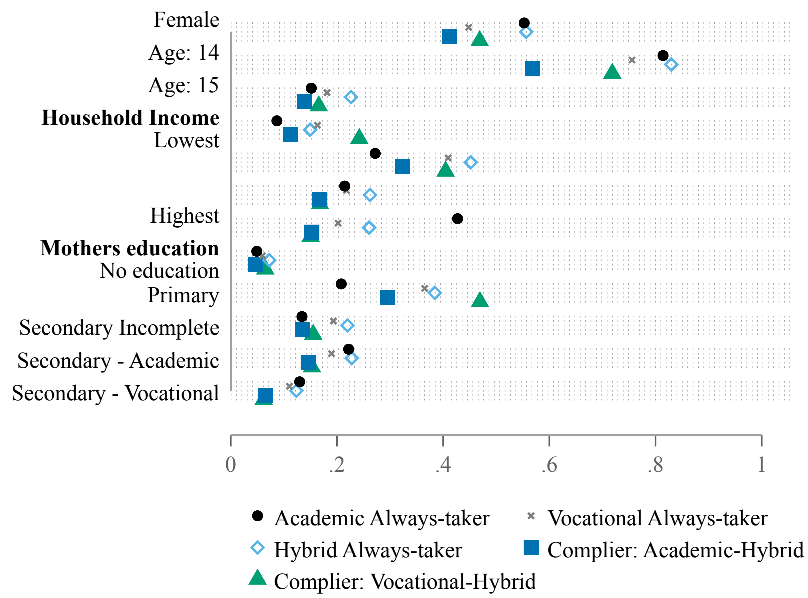
(b) Average of baseline test scores on SIMCE in 2011 and classroom/school quality



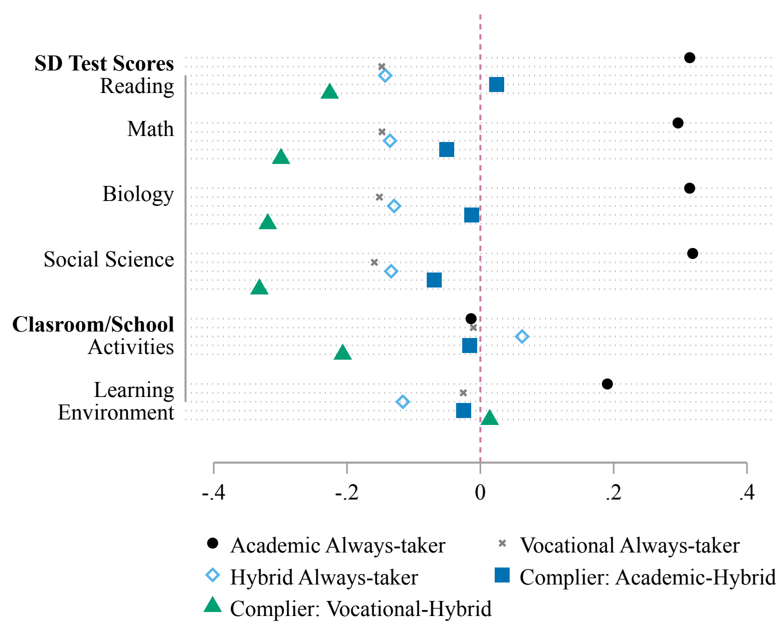
Note: The figure shows the average of baseline characteristics, for the groups of always-takers and compliers induced by the variation in travel time (hours) to the nearest Vocational (V) high school. Panel (a) presents the estimated averages of socio-economic characteristics, while Panel (b) focuses on baseline academic achievement and clasroom/school learning activities and environment.

Figure 3.10: Average characteristics of response groups due to the variation in travel time to Hybrid (H) High Schools, in 2011 (one year before choosing a school-type)

(a) Average of socio-economic characteristics



(b) Average of baseline test scores on SIMCE in 2011 and classroom/school quality



Note: The figure shows the average of baseline characteristics, for the groups of always-takers and compliers induced by the variation in travel time (hours) to the nearest Hybrid (H) high school. Panel (a) presents the estimated averages of socio-economic characteristics, while Panel (b) focuses on baseline academic achievement and classroom/school learning activities and environment.

Table 3.11: Average Local (comuna) Characteristics in 2011, by response groups of the different instruments

Response Group	**Fraction of urban pop.**	**Unemployment rate**	**Population Density**	**Poverty level**
(a) Travel time (in hours) to Academic (A) High Schools				
Academic Always-taker	0.939	0.080	2663.819	0.144
Vocational Always-taker	0.822	0.085	1997.610	0.158
Hybrid Always-taker	0.796	0.077	2572.268	0.151
Complier: Vocational-Academic	0.539	0.086	460.317	0.187
Complier: Hybrid-Academic	0.720	0.088	902.180	0.127
(b) Travel time (in hours) to Vocational (V) High Schools				
Academic Always-taker	0.838	0.079	1922.085	0.150
Vocational Always-taker	0.893	0.085	2459.774	0.157
Hybrid Always-taker	0.782	0.074	2354.162	0.151
Complier: Hybrid-Vocational	0.723	0.086	7.981	0.184
(c) Travel time (in hours) to Hybrid (H) High Schools				
Academic Always-taker	0.850	0.082	2013.525	0.152
Vocational Always-taker	0.821	0.085	1875.660	0.160
Hybrid Always-taker	0.994	0.081	4241.950	0.157
Complier: Academic-Hybrid	0.483	0.064	108.587	0.135
Complier: Vocational- Hybrid	0.599	0.080	154.920	0.162

Note: average of local characteristics, by groups induced by the variation in travel time to each high school-type. Panel (a) presents average of local characteristics of groups induced by the travel time to the nearest Academic (A) high school. Panel (b) presents average of local characteristics of groups induced by the travel time to the nearest Vocational (V) high school. Panel (c) presents average of local characteristics of groups induced by the travel time to the nearest Hybrid (H) high school.

Table 3.12: Local Average Effects for Educational Attainment and Achievement, by type of instrument and complier groups

	Educational Attainment				***Standardized Test Scores***	
	High School	**High School Graduate**	**Higher Ed:**	**Higher Ed:**	***Second grade of High School***	
Average Effect	**Incomplete**	**No Higher Ed.**	**Vocational**	**Academic**	**Reading**	**Math**
(a) Academic High Schools						
versus the next-best	0.0245	-0.2166***	-0.0029	0.1950***	0.0446	-0.1347
	(0.0390)	(0.0489)	(0.0508)	(0.0528)	(0.1350)	(0.1373)
versus Vocational	-0.0332	-0.2636***	0.0364	0.2604***	0.1925*	-0.1001
	(0.0315)	(0.0453)	(0.0410)	(0.0482)	(0.1165)	(0.1358)
versus Hybrid	0.2369	-0.0435	-0.1475	-0.0458	-0.4981	-0.2651
	(0.2188)	(0.1670)	(0.1995)	(0.2187)	(0.5813)	(0.4929)
(b) Vocational High Schools						
versus the next-best	0.0480**	0.0422*	0.0014	-0.0916***	0.1018	0.1415**
	(0.0187)	(0.0229)	(0.0263)	(0.0267)	(0.0635)	(0.0639)
versus Hybrid	0.0356**	0.0596***	0.0312	-0.1263***	0.0016	0.0156
	(0.0172)	(0.0206)	(0.0227)	(0.0246)	(0.0576)	(0.0586)
(c) Hybrid High Schools						
versus the next-best	0.0030	-0.0876***	-0.0232	0.1078***	-0.0197	0.0425
	(0.0146)	(0.0226)	(0.0209)	(0.0242)	(0.0661)	(0.0702)
versus Academic	0.0442**	0.0290	-0.0256	-0.0476	-0.2990***	-0.3302***
	(0.0175)	(0.0239)	(0.0243)	(0.0296)	(0.0681)	(0.0795)
versus Vocational	-0.0423*	-0.2157***	-0.0206	0.2786***	0.2893***	0.4477***
	(0.0239)	(0.0318)	(0.0309)	(0.0311)	(0.0824)	(0.0859)
Observations	89147	89147	89147	89147	89147	89147

Note: LATEs for educational attainment and achievement, using as instruments the travel time to each type of school. Panel (a) presents the LATEs of Academic (A) schools *versus* the next-best; the effects for different fallback options (*e.g.*, vocational or hybrid high schools) are computed assuming homogeneity of potential outcomes among those two choices. Panel (b) presents LATEs of Vocational (V) schools *versus* Hybrid high schools. Panel (c) presents the LATEs of Hybrid (H) schools *versus* the next-best; the effects for different fallback options (*e.g.*, academic or vocational high schools) are computed assuming homogeneity of potential outcomes among those two choices.
Cluster standard errors at the primary school level are in parentheses. All estimations include regional fixed effects. $*p < 0.1, **p < 0.05, ***p < 0.01$

Table 3.13: Robustness check: Main specification and controlling for proximity to amenities

	Educational Attainment				***Standardized Test Scores***	
	High School	**High School Graduate**	**Higher Ed:**	**Higher Ed:**	***Second grade of High School***	
Average Effect	**Incomplete**	**No Higher Ed.**	**Vocational**	**Academic**	**Reading**	**Math**
(a) Academic High Schools versus the next-best						
Main	0.0245	-0.2166***	-0.0029	0.1950***	0.0446	-0.1347
	(0.0390)	(0.0489)	(0.0508)	(0.0528)	(0.1350)	(0.1373)
Amenities	0.0371	-0.1836***	0.0241	0.1223**	0.1182	-0.1323
	(0.0417)	(0.0514)	(0.0531)	(0.0561)	(0.1519)	(0.1465)
(b) Vocational High Schools versus the next-best						
Main	0.0480**	0.0422*	0.0014	-0.0916***	0.1018	0.1415**
	(0.0187)	(0.0229)	(0.0263)	(0.0267)	(0.0635)	(0.0639)
Amenities	0.0425**	0.0817***	0.0205	-0.1447***	0.0979	0.0688
	(0.0212)	(0.0250)	(0.0288)	(0.0297)	(0.0747)	(0.0740)
(c) Hybrid High Schools versus the next-best						
Main	0.0030	-0.0876***	-0.0232	0.1078***	-0.0197	0.0425
	(0.0146)	(0.0226)	(0.0209)	(0.0242)	(0.0661)	(0.0702)
Amenities	-0.0045	-0.0473**	-0.0068	0.0586***	-0.0488	-0.0555
	(0.0149)	(0.0208)	(0.0202)	(0.0222)	(0.0509)	(0.0498)
Observations	89147	89147	89147	89147	61629	61619

Cluster standard errors at the primary school level are in parentheses. All estimations include regional fixed effects. Each outcome is estimated separately.

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

Table 3.14: First-stage: School-type choice, by functional form of travel time to each type of high school

Variable	Academic (A)			Vocational (V)			Hybrid (H)		
Academic (Z_A)	-0.0594***	-0.1282***	-0.2090***	0.0467***	0.0805***	0.1202***	0.0127	0.0478***	0.0889***
	(0.0063)	(0.0137)	(0.0216)	(0.0058)	(0.0127)	(0.0205)	(0.0080)	(0.0162)	(0.0247)
Vocational (Z_V)	-0.0019	-0.0095	-0.0077	-0.0942***	-0.1593***	-0.2310***	0.0961***	0.1688***	0.2388***
	(0.0051)	(0.0110)	(0.0193)	(0.0045)	(0.0095)	(0.0163)	(0.0060)	(0.0132)	(0.0213)
Hybrid (Z_H)	0.0711***	0.1328***	0.1844***	0.0647***	0.0905***	0.1573***	-0.1359***	-0.2233***	-0.3417***
	(0.0068)	(0.0123)	(0.0239)	(0.0053)	(0.0134)	(0.0244)	(0.0076)	(0.0130)	(0.0275)
Z_A^2		0.0139***	0.0551***		-0.0102***	-0.0316***		-0.0037	-0.0234**
		(0.0027)	(0.0084)		(0.0024)	(0.0078)		(0.0031)	(0.0096)
Z_V^2		0.0021	0.0028		0.0143***	0.0487***		-0.0164***	-0.0515***
		(0.0020)	(0.0075)		(0.0017)	(0.0063)		(0.0024)	(0.0079)
Z_H^2		-0.0140***	-0.0420***		-0.0076***	-0.0497***		0.0215***	0.0917***
		(0.0027)	(0.0127)		(0.0029)	(0.0124)		(0.0026)	(0.0148)
Z_A^3			-0.0044***			0.0028***			0.0016*
			(0.0008)			(0.0007)			(0.0009)
Z_V^3			-0.0001			-0.0038***			0.0038***
			(0.0008)			(0.0006)			(0.0008)
Z_H^3			0.0033**			0.0056***			-0.0088***
			(0.0016)			(0.0016)			(0.0018)
Constant	0.6263***	0.6368***	0.6371***	0.1987***	0.2189***	0.2074***	0.1750***	0.1443**	0.1555***
	(0.0513)	(0.0511)	(0.0514)	(0.0545)	(0.0549)	(0.0548)	(0.0595)	(0.0590)	(0.0592)
N	89,147	89,147	89,147	89,147	89,147	89,147	89,147	89,147	89,147
R^2	0.1670	0.1703	0.1716	0.0896	0.0928	0.0947	0.1239	0.1326	0.1359

Note: first-stage regressions of high school choices on the instruments (*i.e.*, travel time) and baseline characteristics. Each choice is estimated separately.

Table 3.15: Local Average Effects for Educational Attainment and Achievement, by functional form of travel time to each type of high school

	Educational Attainment				***Standardized Test Scores***	
	High School	**High School Graduate**	**Higher Ed:**	**Higher Ed:**	***Second grade of High School***	
Average Effect	**Incomplete**	**No Higher Ed.**	**Vocational**	**Academic**	**Reading**	**Math**
(a) Academic High Schools versus the next-best						
Linear	0.0245	-0.2166***	-0.0029	0.1950***	0.0446	-0.1347
	(0.0390)	(0.0489)	(0.0508)	(0.0528)	(0.1350)	(0.1373)
Quadratic	-0.0111	-0.1571***	-0.0210	0.1892***	0.1082	0.0404
	(0.0317)	(0.0393)	(0.0410)	(0.0415)	(0.1079)	(0.1000)
Cubic	-0.0296	-0.1630***	-0.0139	0.2064***	0.1612*	0.1492
	(0.0312)	(0.0377)	(0.0389)	(0.0396)	(0.0978)	(0.0913)
(b) Vocational High Schools versus the next-best						
Linear	0.0480**	0.0422*	0.0014	-0.0916***	0.1018	0.1415**
	(0.0187)	(0.0229)	(0.0263)	(0.0267)	(0.0635)	(0.0639)
Quadratic	0.0372**	0.0046	0.0014	-0.0431*	0.1155*	0.1314**
	(0.0190)	(0.0225)	(0.0251)	(0.0258)	(0.0649)	(0.0637)
Cubic	0.0374**	0.0095	-0.0066	-0.0403	0.0924	0.0899
	(0.0188)	(0.0221)	(0.0246)	(0.0255)	(0.0625)	(0.0614)
(c) Hybrid High Schools versus the next-best						
Linear	0.0030	-0.0876***	-0.0232	0.1078***	-0.0197	0.0425
	(0.0146)	(0.0226)	(0.0209)	(0.0242)	(0.0661)	(0.0702)
Quadratic	0.0027	-0.0898***	-0.0143	0.1014***	-0.0551	-0.0185
	(0.0146)	(0.0201)	(0.0201)	(0.0219)	(0.0536)	(0.0574)
Cubic	-0.0001	-0.0932***	-0.0067	0.1000***	-0.0323	0.0016
	(0.0145)	(0.0194)	(0.0194)	(0.0211)	(0.0543)	(0.0580)
Observations	89147	89147	89147	89147	61629	61619

Cluster standard errors at the primary school level are in parentheses. All estimations include regional fixed effects. $*p < 0.1, **p < 0.05, ***p < 0.01$

Chapter 4: Labor Market Effects of Short-Cycle Higher Education Programs: Challenges and Evidence from Colombia

Note: This chapter of the dissertation is coauthored with Maria Marta Ferreyra and Sergio Urzúa

4.1 Introduction

The effects of higher education expansions and the role of degrees of short duration (such as two-year college degrees or short-cycle degrees) as pathways to provide skills fast for a larger share of students have received increased attention: for the U.S., recent evidence shows that community, or two-year, colleges can benefit students in terms of educational attainment and earnings (Acton, 2020; Bettinger and Soliz, 2016; Denning, 2017; Jepsen, Troske, and Coomes, 2014; Marcotte, 2019; Minaya and Scott-Clayton, 2020; Mountjoy, 2019). Moreover, these effects are mostly driven by students who would not have enrolled in higher education in the absence of short-term degrees. Latin America has also experienced important expansions of the supply of higher education with heterogeneous effects across countries and by type of program (Camacho, Messina, and Uribe Barrera, 2017; Carranza and Ferreyra, 2019; Ferreyra, Avitabile, Botero Álvarez, Haimovich Paz, and Urzúa, 2017;

Ferreyra, Franco Hernandez, Melguizo, and Sanchez Diaz, 2020; González-Velosa, Rucci, Sarzosa, and Urzúa, 2015).

In this chapter, we estimate labor market effects of enrolling in a Short-Cycle program using local variation in the supply of higher education for the universe of high school graduates in 2005 in Colombia. Following long-standing literature on two-year college enrollment for the U.S. (Acton, 2020; Denning, 2017; Leigh and Gill, 2003; Mountjoy, 2019; Rouse, 1995), we argue that access to Short-Cycle degrees might attract some students who would have not enrolled in higher education otherwise (i.e., the *expansion* or *democratization* margin), while also inducing other students to divert from Bachelor's- and into short-cycle- degrees (i.e., the *diversion* margin). Students along these margins likely differ both in their characteristics and in the gains, or losses, from choosing a Short-Cycle program.

Given recent efforts in Colombia, and other Latin-American countries, to expand access to short-cycle degrees (Ferreyra et al., 2017), uncovering effects along these two margins of choice warrants attention. Nonetheless, the identification of effects at each margin comes at the cost of overcoming several identification and estimation challenges. First, students self-select into their preferred choice, and hence OLS estimates of Short-Cycle programs on labor market outcomes would be biased. Second, while Instrumental Variables (IVs) can help overcome the endogeneity problem, the researcher can only identify the effects of Short-Cycle programs *versus* the next-best (which is a mixture of not enrolling in higher education and enrolling in Bachelor's programs). Only additional (but often restrictive) assumptions can decompose the next-best effect into margin-specific effects (Hull, 2018; Kline and

Walters, 2016; Lee and Salanié, 2020).

Our main contribution lies in providing evidence of expansion and diversion effects for a context such as the Colombian higher education system. While the economic returns to higher education in Colombia have been previously studied, the evidence either focuses on the effects of new *versus* existing programs Camacho et al. (2017), selective colleges Barrera-Osorio and Bayona-Rodríguez (2019), or fields within programs Ferreyra et al. (2020) and González-Velosa et al. (2015). Unlike previous studies, and for simplification, we abstract from within-program heterogeneity but provide Local Average Treatment Effects (LATEs) for students along two well-defined and policy-relevant margins of choice: expansion and diversion into short-cycle degrees.

We use an Instrumental Variables (IVs) approach that uses outcomes by choice interactions and propensity scores (as in Mountjoy, 2019) to estimate LATEs of Short-Cycle programs *versus* the next-best. A critical advantage of this strategy is that it allows us to isolate the effect of a single instrument and, in this way, improve over standard Two-Stage Least Square (2SLS) methods with multiple endogenous choices and multiple IVs.[1] Using administrative data for the universe of high school graduates in Colombia in 2005, we can track students from the High School Exit Exam to their enrollment choice and match them to labor market participation

[1] Heckman and Urzúa (2010) show that if one instrument affects only one margin of choice IVs estimate the effect of one option *versus* the next-best. Also, Mountjoy (2019) shows that multivariate 2SLS (multiple choices and IVs) estimate effects that are a combination of many margins of choice and thus do not necessarily identify well-defined effects. For instance, if there are three margins of choice (not enrollment to short-cycle, bachelor's to short-cycle, and not enrolling to bachelor's) and multiple instruments, then 2SLS would estimate a combination of effects for those three margins.

from 2008-2013. We link this dataset to Higher Education data at the municipality level to exploit local variation in a binary variable that takes the value of one if the student went to high school in a municipality with a Higher Education Institution specialized in Short-Cycle degrees, in a 10km radius.

Our main findings are as follows: only 2.9% of students react to the variation in the local supply of short-cycle degrees by enrolling in short-cycle programs. Moreover, almost 97% of this variation corresponds to students who would divert from Bachelor's programs and into short-cycle degrees (*i.e.*, the diversion margin). The remaining 3% corresponds to students along the expansion margin. We also observe a higher diversion share among men and women, of about 85% and 75%. Along these lines, Carranza and Ferreyra (2019) and Ferreyra et al. (2017) also find evidence for Colombia of diversion from Bachelor's degrees and into short-cycle degrees as a result of a Higher Education expansion. Nonetheless, the authors also find important effects of new higher education programs on inducing students to enter the higher education system. For students in our cohort, we also observe a decrease in the share of students who choose not to enroll in higher education. Still, this decrease is driven by higher education institutions that do not specialize in short-cycle degrees. Last, we uncover positive effects of choosing a short-cycle program on labor market participation rates and experience. These effects are higher and statistically significant for women, and we find no significant labor market effects among men.

Our findings are at odds with the evidence for two-year college enrollment in the U.S. Several studies (Acton, 2020; Denning, 2017; Leigh and Gill, 2003; Mountjoy, 2019) found that the share of students along the expansion margin dominates the share of students along the diversion margin. For instance, for the state of Texas, Mountjoy (2019) finds that about two-thirds of students would react to changes in proximity to two-year college by switching along the expansion margin; the remaining 30% would divert from four-and into two-year college. Denning (2017) finds that lower two-year college tuition would induce students who would not have enrolled otherwise to attend a two-year college, with no evidence that students substituting two- for four-year college degrees as a result of decreases in tuition. In terms of labor market effects, the literature has also found that students along the expansion margin benefit from two-year college enrollment in terms of average earnings and years of education, as well as adverse effects for students diverting from a four-year college.

Differences in the role of Short-Cycle degrees *versus* two-year college degrees might explain our findings in contrast with the literature. We argue that in Colombia, and most likely other Latin American countries, Short-Cycle degrees and Bachelor's degrees follow a logic of substitutes. Meanwhile, two-year college might still provide a "transfer function" (Kane and Rouse, 1999) towards earning a four-year college degree for a large share of students in the U.S. In turn, short-cycle degrees are mainly concerned with the college-two-work transition, and we observe that less than 20% of students who choose short-cycle degrees in our sample end up transferring to a Bachelor's degree. Furthermore, we also present descriptive evidence to show that local employment prospects, rather than changes in higher education supply, can be an important driver to induce students to enroll in higher education via short-cycle degrees. All in all, we see our results as complementing the existing literature on expansion and diversion effects with evidence that can be informative for settings similar to the Colombian system.

This chapter is organized as follows. The following Section describes the higher education system in Colombia. Section 4.3 describes the data, presents an exploratory analysis of enrollment choices and labor market participation, and results using a standard Instrumental Variables strategy. Section 4.4 discusses how to identify shares of students along the expansion and diversion margins, as well as their LATEs. Last, we present conclusions and discussion in Section 4.5.

4.2 Higher Education System

The Higher Education System in Colombia offers different types of programs that can be classified in Bachelor Programs (BP) which last four to five years, and Short-Cycle (SC) Programs which last two to three years (similar to US two-year or community college programs). The latter are divided in technical and technological degrees. In terms of Higher Education Institutions (HEI), there are Universities, University Institutes, Technological Institutes, and Technical Professional Institutes. Universities and University Institutes can offer either BP or SC programs, Technological Institutes and Technical Professional Institutes can only offer SC degrees. In our estimation strategy, we exploit variation in the supply of SCP in two ways: (i) at the local level (in the municipality where the student attended high school), (ii) at the type of HEI that is available (by distinguishing between HEI that can only offer SCP and HEI that can offer either BP or SCP).

4.3 Data

We use administrative data on test scores, enrollment choices, higher education trajectories, and wages in the formal labor market, for the universe of high school graduates in 2005. First, individual records from the High School exit exam (Saber 11, in Spanish) contain information on test scores for math, language, science, physics, history, chemistry, geography, and philosophy. These records also include data on student's age and gender, as well as socioeconomic characteristics prior to enrollment in higher education such as mother's education, household income level, and number of siblings. Higher education enrollment, trajectories, and completion come from the System for Dropout Prevention of Higher Education (SPADIES, in Spanish).

We combine the administrative datasets with information on higher education supply from the National System of Information on Higher Education (SNIES, in Spanish), which contains the number of institutions and programs, their year of creation, field and academic level, and enrollment from 1998-2013. We also use municipal level data on labor market participation from the 2005 Census of Population, from IPUMS International (Minnesota Population Center, 2019). We link our data to the Municipal Panel of the Center of Studies on Economic Development, (CEDE, in Spanish), which contains information on population (urban, rural, and total), distance to the main food market, distance to the main city in the department, area in squared km, poverty level, and the fraction of unemployment among households.

Wages and occupation in the formal labor market for those who graduate from a higher education program are available in the dataset from the Labor Market Observatory for Education (OLE, in Spanish) from 2008-2013. For high school graduates and those who did not complete a higher education program we use data for Colombia from the Socio-economic Database for Latin America and the Caribbean (SEDLAC), to estimate labor market participation and experience between 2008-2013. In particular, we follow the labor market trajectories of students within that period, 2008-2013, using data from OLE and from the Integrated Household Survey (GEIH, in Spanish) in SEDLAC. We use the latter to estimate labor market participation of high school graduates and higher education dropouts with a linear probability model for the cohort of individuals who were 14-24 years of age in 2005, on age, household size, household income, and regional dummies. We estimated each equation separately for students who did not enroll in higher education and for higher education dropouts.[2]

The estimating sample consists of high school graduates in 2005, who were between 14-24 years of age at the time of the high school exit exam. About 53.8% of high school graduates in the sample are women, 27.2% come from low income households, and 42% have mothers with less than primary education. More than half of students in the sample (61.2%) never enrolled in a higher education program.[3]

[2]For more details on our imputation strategy see Appendix C. We imputed both labor market participation and wages, but due to data limitations and self-selection into working we restrict our analysis to years of experience and the probability of working in the formal labor market between 2008-2013.

[3]These are students who do not match to SPADIES, the dataset of higher education enrollment, or OLE, the dataset of labor market earnings. It is worth noting that neither SPADIES nor OLE contain information for students who enroll in SENA, which is a public institution that provides post-secondary technical and technological training. In this sense, the share of our sample classified as not enrolled in higher education also contains students who attend SENA. Nonetheless, we do not have information available to separately identify these two groups of students.

Of the remaining students, 38.8% enrolled in higher education with 9.5% in SCP while 29.3% selected a BP.[4] In terms of higher education completion, about 4.7% of students enroll in SCP but do not complete it, and only 3.6% of the sample has a SC degree. In contrast, 14.9% of students obtain a Bachelor's degree. Less than half (41.8%) of the students in our sample did not participate in the formal labor market between 2008-2013. Moreover, the average years of experience[5] within that period are 2.6. (see Table 4.1).

Students who enrolled in Short-Cycle (SC) programs are older, belong to larger low income households, and have less educated mothers, on average, than students who enrolled in Bachelor's programs (BP). Moreover, not enrolled (NE) students stem from the most disadvantaged backgrounds and had the lowest average performance on the High School Exit Exam (Table 4.2). In turn, BP students have higher test scores, on average, than SC students across all academic subjects of Saber 11. The largest differences between SC and BP students correspond to scores in Chemistry, Biology, and Reading; on average, BP students outperformed SC students in at least 0.3 standard deviations.

About 16.4% of students who enrolled in SC programs transfer to a BP, and less than half of those who transfer complete the program. In contrast, only 1.2% of students who enrolled in BP completed a SCP. While the fraction of dropouts is similar across SC programs and BP (about 50%), completion rates are higher

[4] We restrict the first enrollment to be no later than five years after high school graduation.

[5] We measure labor market experience as the number of years the student appears in the OLE dataset between 2008-2013, for those who competed higher education. For high school graduates and those with incomplete higher education, we count the number of years that the imputed participation rate of the student was above 0.5.

among BP students in almost 14.5 percentage points. On the other hand, the share of students out of the formal labor market between 2008-2013 is about the same (at 37%) among SC students and BP students but the total years of experience are, on average, higher among the former. Last, those who do not enroll in higher education have the highest share, at 45%, of students who do not report working in the formal labor market between 2008-2013. NE students, who have been out of school for longer than SC and BP students, also have the highest average years of experience.

Table 4.3 shows average local characteristics of the municipality where students attended High School, by type of enrollment. On average, NE students lived in municipalities with lower levels of urbanization, higher poverty, and a higher fraction of households with unemployed members, compared to SC and BP students (see Table 4.3). In turn, SC students live in denser municipalities, more urbanized, and better connected municipalities with lower poverty rates than BP students. Hence, the differences in local characteristics between SC and BP students goes in the opposite direction to what we observed in terms of individual and household characteristics. That is, while SC students were more disadvantaged in terms of household characteristics and test scores, they stem from municipalities with better local conditions, on average, than BP students.

4.3.1 Exploratory Analysis

To determine the role of Short-Cycle degrees on student's labor market trajectories we focus on three outcomes: the probability of not working in the formal labor market between 2008-2013, the probability of working in 2013, and years of experience between 2008-2013. The first corresponds to students who have not reported to hold a formal job between 2008-2013, and we define it as a binary variable that takes the value of one for students out of the formal labor market. The latter refers to the total of years of formal work between 2008-2013. This five year period corresponds to the years of student-level labor market data we have available from OLE.

To explore how SC enrollment affects labor market trajectories, we first estimate the following equation by OLS:

$$Y_i = \beta_0 + \beta_{SC} D_{SCi} + \beta_X X_i + u_i, \tag{4.1}$$

where Y_i denote labor market outcomes for student i, and D_{SCi} is a binary variable that takes the value of one if student i enrolls in a Short-Cycle program. X_i includes age, gender, number of siblings, household income level, mother's education, test scores in Saber 11, and characteristics of the municipalities where the student attended high school (as detailed in Table 4.3).

One important feature of equation (4.1) is that the omitted category is a mixture of students who did not enroll in higher education, and students who enroll in BP. Hence, it ignores that students with different counterfactuals can derive different gains from choosing a SC program. Also, u_i is an unobserved component which contains students' preferences and unobserved ability, among other factors. Since these factors can also explain enrollment choices, we see the results we present here as descriptive rather than causal evidence of the labor market effects of Short-Cycle degrees. In this sense, Table 4.4 shows a positive association between enrolling in a SC program and not working in the formal labor market, than enrolling in other programs or not entering higher education. Similarly, enrolling in a SC degree is associated with a decrease of 0.6 years of experience *versus* not enrolling in a SC program.

4.3.2 Standard Instrumental Variables

Students who choose to enroll in a Short-Cycle program differ in observable, and unobservable, characteristics from students who choose to not enroll in higher education or to enroll in a Bachelor's program. Preferences, motivation, and abilities, which we can only capture with noise using test scores, can explain enrollment choices as well as labor market outcomes. To deal with self-selection into short-cycle programs, we use an instrumental variables (IVs) strategy. Define equation (4.1) as the second-stage, or the outcome equation. In a standard 2SLS set up, the first-stage would correspond to the following:

$$D_{SCi} = \gamma_0^{SC} + \gamma_1^{SC} Z_i + \gamma_X^{SC} X_i + \epsilon_{SCi}, \tag{4.2}$$

where Z is an IV, X_i is a matrix of student-level and municipal-level characteristics, and ϵ is an unobserved component. We define the instrument in Z as follows:

Local supply of higher education by type of program: we use data from Saber 11 to identify the municipality where the student attended high school. We use this information to determine if, in 2004, there was any SC or BP institutions in close proximity to the students' high school municipality. Furthermore, we distinguish between HEIs that only offer SC programs, only offer BP programs, and those that offer both types. We use the variation in local availability by type of HEI to define a binary instrument, Z, that takes the value of one if there was a HEI specialized in SC programs in a radius of 10km from the student's HS municipality. We also define variables for availability of other type of institutions (those only offering bachelor's programs, and those offering either type).

The IV of local availability of SC programs follows the logic of cost-shifters, as in Card (1995) and Mountjoy (2019). That is, we assume that students with a HEI near their high school municipality might have lower costs of enrolling in a HE program. In contrast, students with no HEI locally would face higher costs. These costs can be in terms of access (e.g., having a HEI locally reduces transportation costs) or information (e.g., if there is a HEI near the student has less barriers to learn about the supply of programs). Table 4.5 shows the average of HEI availability in 2004, for the total sample. About 78.3% of students went to high school in a municipality with a HEI that offered both SCP and BP. This fraction reduces to 61.1% for HEI that only offered SC programs.[6] Moreover, we observe that almost 74.2% of SC students went to high school in a municipality with SC HEI, and this share reduces to 65.4% for BP and 57% for NE students.

[6]Although these local differences in the supply of HEI can proxy for the costs that students face, there can be heterogeneity in the tuition that different types of HEI can charge. There is no publicly available data on tuition by HE programs in 2005, however more recent data (González-Velosa, Rucci, Sarzosa, and Urzúa, 2015) shows that tuition in private HEI can be twice as high that of public HEI. Using imputed data for tuition, we observe that SC programs charge on average about half of the average tuition among BP.

We also use measures of enrollment in SENA in 2004, to capture the availability of public, free of charge, technical and technological programs. Around 75% of students attended high school in municipalities with positive SENA enrollment, and this share is higher among BP and SC students than among NE students. We control for this variable in our regressions, since our group of NE students consist of high school graduates who did not continue their education as well as high school graduates who enroll in SENA.

With a standard IV strategy, we can obtain Local Average Treatment Effects (LATEs) of HEI enrollment on labor market outcomes. That is, effects among students who would change their enrollment choices as the instruments change (*i.e.*, compliers). Table 4.6 shows LATEs of higher education enrollmnent on formal labor market experience. Taken as given, the results suggest that among compliers SC enrollment decreases the probability of not working in the formal labor market, and increases formal experience in 1.2 years. Both results have the opposite direction than the OLS results in Table 4.4. Only the effect on experience is statistically significant. However, as we described in the previous sections, the effects of enrolling on SC can largely depend on the counterfactual option, or what students would have chosen in the absence of SC programs. While on average SC programs can have no effect on labor market experience, there can be groups of students who, depending on their counterfactual, would benefit from choosing a SC program.

In our setting, we can define two groups of SC students according to their counterfactuals: (i) students who would have not enrolled in higher education, and (ii) students who would have enrolled in BP. Previous literature (Mountjoy, 2019;

Rouse, 1995) defines the former as the *democratization* or *expansion margin* (students enter the HE system with a SC program), while the latter is referred to as the *diversion margin* (students divert from BP and into SC programs). Changes in the instruments could shift students into SC from either of those margins (*e.g.* from NE towards SC, or from BP towards SC), which means that there can be more than one type of compliers. Moreover, each complier group can experience different effects of choosing SC. Let $LATE_{BP-SC}$ represent the local effects of choosing SC for students switching along the diversion margin; $LATE_{NE-SC}$ represents the local effects of choosing SC for students switching along the expansion margin. Previous literature (Heckman and Urzúa, 2010; Mountjoy, 2019) has shown that, in contexts where compliers are switching from different initial states, standard (univariate) IV estimates the effect of one option, (*e.g.* SC) *versus* the next-best (*i.e.*, a mixture of NE and BP).

$$\beta^{SC} = \omega LATE_{BP-SC} + (1-\omega)LATE_{NE-SC}, \tag{4.3}$$

where ω is the share of compliers along the diversion margin.[7] These are the results we report in Table 4.6.

[7] Mountjoy (2019) presents a formal derivation of these effects (*i.e.*, of one option *versus* the next-best) in the context of two-year college in the U.S. Moreover, he also shows that multivariate IV (*i.e.*, where there are two or more endogenous choices) does not necessarily identify a well-defined effect but rather a combination of students shifting along many margins).

4.4 Self-selection into Short-Cycle programs: expansion *versus* diversion

Uncovering group specific LATEs of SC programs for students along the expansion *versus* diversion margins is both an empirical and policy-relevant challenge. First, without additional assumptions, standard IV does not identify group specific sub-LATEs. Furthermore, there should be as many instruments as margins of choice for the identification of sub-LATEs. More important, SC programs might be a better match for students along the expansion or diversion margins. Overall, LATEs of one option *versus* the next-best do not provide enough information about the type of students who might, or not, benefit from a SC program.

In this section we employ the methodological framework in Heckman and Pinto (2018) and Mountjoy (2019) to, first, estimate the share of expansion-compliers and diversion-compliers induced by the variation in local availability of SC HEI. Then, with additional assumptions, we estimate subLATEs of SC programs on labor market outcomes. In particular, we find no statistical evidence that changes in the availability of SC HEI shift students along the expansion margin ($i.e.$, $1-\hat{\omega}=0$). We show how, in turn, for students along the expansion margin other local conditions, such as employment prospects, matter for the decision of whether or not to enroll in college.

4.4.1 Complier Shares

In this section we present our identification strategy to estimate ω in equation (4.3). Let D_{di} denote the student's enrollment choice. That is, D_{di} are binary variables such that $D_{di} = \mathbb{1}[\text{student } i \text{ chooses enrollment option d}]$ for $d \in \{SC, BP, NE\}$. Choices depend on students' characteristics, in X_i, such as household income, maternal education, gender, age, and household composition, as well as test scores in the High School Exit exam, and on local SC HEI availability in instrument Z. In what follows, we implicitly condition on X_i to simplify notation.

Following the logic of cost-shifters, we assume that changes in Z would induce some students to change their enrollment choices towards SC programs. Let $D_{di}(Z)$ denote potential choices, or the enrollment status student i would choose at different realizations of Z. For instance, if $D_{SCi}(1) = 1$ the student would choose a SC program when there was a SC HEI in her high school municipality. We formalize the potential changes in enrollment choices as the instrument changes as follows:

$D_{NEi}(0) \geq D_{NEi}(1)$, which means that having a SC HEI available should make NE weakly less attractive for students.

$D_{BPi}(0) \geq D_{BPi}(1)$, which means that having a SC HEI available should make BP weakly less attractive for students.

$D_{SCi}(0) \leq D_{SCi}(1)$, which means that having a SC HEI available should make SC weakly more attractive for students.

The combinations of these enrollment changes as a result of changes in the instrument satisfy the property of *monotonicity* (Imbens and Angrist, 1994) such that all students are (weakly) induced towards the same option (SC, in this case). Using these assumptions on potential enrollment choices, we can identify the total share of compliers, as well as the shares of margin-specific compliers. Formally,

$$\omega = \frac{E[D_{BP}|Z=0] - E[D_{BP}|Z=1]}{E[D_{SC}|Z=1] - E[D_{SC}|Z=0]}$$

$$1 - \omega = \frac{E[D_{NE}|Z=0] - E[D_{NE}|Z=1]}{E[D_{SC}|Z=1] - E[D_{SC}|Z=0]},$$

where the total share of compliers is given by $E[D_{SC}|Z=1] - E[D_{SC}|Z=0]$ which denotes the share of students changing their choices towards SC programs as the local availability of SC HEI changes.

To estimate the complier shares, we estimate propensity scores for each enrollment choice in $d \in \{SC, BP, NE\}$ with a multinomial logit as follows:

$$D_i = \gamma_0 + \gamma_1 Z_i + \gamma_X X_i + \epsilon_i$$

where D_i is a categorical variable that takes on three values: not enrollment, SC programs enrollment, and BP enrollment; ϵ_i is an unobserved component. Table 4.7 and Table 4.8 show the results from the multinomial logit, for each enrollment choice. Women have a lower probability of enrolling in SC programs, in 2.4 percentage points, *versus* male students. Moreover, increases in age, the number of siblings, and maternal education decrease the probability of choosing a SC program. In terms of academic performance in Saber 11, increases in one standard deviation in Math or Physics decreases the probability of choosing SC programs. The opposite effect appears for Reading, History, and Philosophy subjects. In contrast, higher scores in Saber 11 across all subjects increase the probability of enrolling in BP and decrease the probability of NE.

Table 4.8 shows the effects of municipal characteristics and of the local availability of HEI. In particular, we observe that having a HEI specialized in SC programs in a 10km radius, increases the probability of choosing SC programs in 2.9 percentage points, and decreases the probability of choosing BPs in 2.8 percentage points. Furthermore, we observe that changes in the local availability of SC HEI have no significant effect on the probability of not enrolling in higher education. The magnitude of this effect is also small, at 0.09 percentage points. It is worth noting that other type of HEI (for instance, universities) seem to shift students along the NE-BP margin, but not into or away from SC programs. However, the share of compliers who would shift into BP and out of NE as the local supply of (not SC) HEI changes is only 1.14%, which is about half of the change we observe as a result of changes in Z.

We use the multinomial logit results to estimate the share of $BP - SC$ compliers (or, students switching along the *diversion margin*). Table 4.9 shows that diversion compliers represent the majority of compliers shifting towards SC, *i.e.*, $\hat{\omega} = 0.965$. We estimate these shares for women and men, and observe similar patters. Importantly, the prevalence of BP-SC compliers is higher among men (at 3.2%) than among women (at 1.9%). These results are consistent with the descriptive analysis in the previous section, where we observe that women are less likely to enroll in SC programs. Overall we see no evidence of Z, local availability of SC HEI, inducing students to change their choices along two, but rather one, margins. We use these findings to argue that LATEs estimated with the variation of Z are likely to be mostly driven by diversion compliers that by a combination of subLATES for expansion and diversion compliers.

4.4.2 LATEs

In this section, we estimate LATEs for compliers induced into Short-Cycle programs as the local availability of SC HEI changes. Importantly, our first-stage in equation (4.2) includes Z as well as variables for the availability of other HEIs (*i.e.*, those that offer other programs and are not fully specialized in SC degrees). If some students would also change their enrollment choice as the availability of other HEIs changes, standard 2SLS does not identify the effect of one option *versus* the next-best, instead the effect it uncovers is a weighted average of subLATEs for all compliers.[8] Table 4.8 shows that as the availability of other HEI changes, some students are induced to switch between BP and NE. Note that these students would not switch towards, or away, of SC programs which is the change in behavior that interests us in this chapter. Nonetheless, standard 2SLS would attribute part of the subLATE for the BP-NE margin to the effect of SC programs. To deal with this bias component in standard 2SLS, we employ the assumption of *Partial Monotonicity* in Mountjoy (2019) and Mogstad et al. (2020a,b):

[8] Mountjoy (2019) shows that in the case of multivariate IV and multiple instruments, 2SLS identifies weighted averages of subLATEs across all margins of choice.

Let Z^- denote other variables that explain choices in equation (4.2) and are excluded from the outcome equation in (4.1). Let $D_{di}(Z = z|Z^-)$ denote the potential enrollment choice of student i when $Z = z$, keeping Z^- fixed. In our setting of SC programs, partial monotonicity states that, conditional on Z^-, all students should be induced towards, or away from, SC programs as Z changes: $D_{NEi}(0|Z^-) \geq D_{NEi}(1|Z^-)$, $D_{BPi}(0|Z^-) \geq D_{BPi}(1|Z^-)$, and $D_{SCi}(0|Z^-) \leq D_{SCi}(1|Z^-)$.

With this assumption, we isolate the effect of SC programs *versus* the next-best (which, in our case, denotes a mixture of NE and BP), as a result of the variation in Z:

$$LATE_{SC-/SC} = \frac{E(Y|Z = 1, Z^-) - E(Y|Z = 0, Z^-)}{E(D_{SC}|Z = 1, Z^-) - E(D_{SC}|Z = 0, Z^-)}$$

We estimate this effect with regressions of Y and D_{SC} on Z, Z^-, and X. As we discussed in Section 4.3.2, $LATE_{SC-/SC}$ is a combination of subLATEs for NE-SC compliers and BP-SC compliers (see equation (4.3)). However, given that we find no evidence of compliers along the NE-SC margin we attribute most of the effect on $LATE_{SC-/SC}$ to the BP-SC margin.

Table 4.10 shows the estimates of $LATE_{SC-/SC}$ on the probability of not working in the formal labor market between 2008-2013, the probability of working in 2013, and the number of years that the student works from 2008-2013. SC programs decrease the probability of not working in the formal sector in about 22 percentage points, which is almost four points higher than the effect we find with standard 2SLS (in Table 4.6). We also observe positive local effects of SC programs on participation in the formal labor market in 2013 and years of experience, but the latter is not statistically significant. In terms of heterogeneity, we observe results in the same direction for women and men but the magnitude is larger among the former. Overall, we observe positive gains of choosing SC programs among compliers, which are likely to be driven by compliers along the diversion margin, particularly for women compliers. Among the latter we estimate large and statistically significant gains of switching towards SC programs.

To explain the larger gains for women *versus* men, as well as the overall positive effects of SC programs, we estimate the average initial (before enrolling in HE) characteristics of BP-SC compliers. Table 4.11 shows that, on average, women compliers are older, from low income households, and had lower average test scores in Saber 11 across all academic subjects than male compliers and compliers in the full sample. Women compliers also lived in municipalities with lower average urbanization, as well as denser, poorer and less connected areas, than men compliers and overall compliers.

4.4.3 Responses along the *expansion* margin

In the previous section, we find that changes in the local availability of SC HEI would induce some students to divert from Bachelor's programs and into Short-Cycle programs, but would not induce students along the NE-SC margin. These results contrast with findings for the U.S.: for the state of Texas, Mountjoy (2019) shows that about 60% of compliers induced towards two-year college because of variation in distance to this option correspond to expansion margin compliers. Our estimates for Colombia, in turn, show that expansions of the supply of SC HEI are likely to divert some students from BP rather than induce students to enroll in HE with a SC program. Nonetheless, these results provide important evidence and indicate that, in our context, there are other constraints that might restrict NE students from choosing SC programs. For instance, tuition and other enrollment and attendance costs could deter NE students from enrolling in a HE program.[9] Given data limitations on HE costs, we explore the role of local employment prospects which could also affect the student's decision of not to enroll in higher education programs.

We use Census data for 2005 from IPUMS on municipal employment rates by educational level, for individuals 28-33 years old. IPUMS contains data for 445 municipalities, of which only 59 are also in our sample. As a result, we lose about 45% of our original estimating sample. Nonetheless, we perform the analysis in this subset to explore if students respond to local changes in employment when making enrollment decisions. Our main hypothesis is that changes in employment rates for SC degrees could make enrolling in a SC program more attractive. In particular, we define a binary instrument that takes the value of one if the local employment rate (at the student's High School municipality) is higher than the department employment rate, for individuals of ages 28-33 with SC degrees. Formally, let Z_E

[9] Although we have no direct measures of tuition fro our cohort, in 2005, we explore results with two sets of proxies: imputed tuition to 2005 prices using data from 2009, and the public character of some HEIs. The public character of HEI is a potential proxy for tuition, given that public HEI are typically cheaper than private HEI (by almost half, González-Velosa et al. 2015). Our results using these measures do not present different enrollment patterns than those we described in this chapter.

be the binary instrument of local employment rate above the department level.

Our smaller sample largely differs from the total sample in terms of the availability of HEI; the vast majority of students in the subsample has available (in a 10km radius) a HEI that offers other types, not SC, of programs. Hence, we further restrict the subsample to students who had a HEI offering other programs in a 10 km radius, which removes 1% of the observations. Among these students, we estimate propensity scores for each choice (SC, BP, and NE) on student's characteristics, fixed effects for region and metropolitan areas, as well as the IV of local employment in Z_E and a similar variable of employment rates for HS graduates.

Table 4.12 shows that students in municipalities with higher SC employment rates have a higher probability of enrolling in SC programs, on 8.8 percentage points, *versus* students in municipalities with lower than at the department level SC employment rates. There is no evidence of effects of differences in the local *versus* department level employment rates on the probability of choosing BP, rather the increase in SC enrollment seems to be driven by a decrease in the probability of not enrolling in higher education. The latter effects is of almost 10 percentage points. Hence, favorable employment conditions for SC degrees could drive students along the expansion margin.

While the effect of Z_E on the expansion margin is an important result, we find no empirical evidence to be confident in estimating LATEs using the variation of Z_E. First, the direction of the effect of Z_E on the probability of choosing BP is the opposite to what we would expect: rather than decreasing enrollment in BP, such that students substitute BP for SC, it increases it. Second, the results of diagnostics for weak instruments on the first-stage do not pass strict criteria (although some statistics are above the frequently used criteria of F>10). Hence, while we present results in Table 4.13 of estimating the effect of SC programs on labor market outcomes by OLS and with Z_E as instrument, we see these results as informative rather than causal effects. Overall, SC might not improve the labor market participation and experience of students who would have not enrolled in higher education but switch to a SC program because of changes in local employment prospects.

4.5 Discussion

This chapter discusses the empirical challenges associated with the estimation of the economic returns to Short-Cycle degrees, namely self-selection of students into multiple enrollment options. We focus on the context of Colombia; in our setting, students can choose not to enroll in higher education, to enroll in a Bachelor's program, or to enroll in a Short-Cycle program. In line with the literature on returns to higher education in the U.S., (Denning, 2017; Mountjoy, 2019; Rouse, 1995), we argue that the effects of Short-Cycle programs largely differ depending on the fallback alternative of students. That is, if Short-Cycle programs were not available some students would have chosen a Bachelor's program while other students would have deter from entering higher education. These students, that differ in terms of their fallback option, can derive different effects from choosing a Short-Cycle degree. We use an Instrumental Variables (IVs) approach, and exploit variation in the supply of Short-Cycle degrees at the municipal level. In particular, we define a binary IV that takes the value of one if the student when to high school in a municipality where Short-Cycle degrees were offered in a radius of 10km.

We uncover three main findings: only 2.9% of students react to the variation in the local supply of Short-Cycle programs. But we find that among these students almost 97% would divert from Bachelor's programs and into Short-Cycle degrees. The remaining 3% corresponds to students who would be induced to enter the higher education system with a Short-Cycle degree as a result of changes in the local supply. We also observe similar patterns among men and women, with shares of 85% and 75%. These findings are in stark contrast with previous evidence for the U.S.: for the state of Texas, Mountjoy (2019) finds that about two thirds of students would react to changes in proximity to two-year college by switching along the expansion margin (*i.e.*, from not enrolling in higher education towards enrolling in a two-year college); the remaining 30% would divert from four-and into two-year college (along the diversion margin). Along these lines, Denning (2017) finds that reduced tuition for two-year college increase higher education enrollment, rather than decreasing enrollment in universities.

Moreover, we estimate that Short-Cycle degrees can improve the labor market participation rates and experience of students along the diversion margin. These results are larger for women, and we find no significant effects on labor market participation among men. Similar to our findings on diversion effects of changes in supply, our results for labor market outcomes are somewhat different from previous literature. For instance, Mountjoy (2019) finds that students along the expansion margin are most likely to benefit from two-year colleges in terms of average earnings and years of education, than students along the diversion margin.

Differences in the role of Short-Cycle degrees *versus* two-year college degrees might explain our findings in contrast with the literature: while two-year college can still serve as a pathway into four-year college enrollment, Short-Cycle degrees are mostly focused in preparing students for the transition into work, rather than into Bachelor's degrees. Moreover, students from countries such as Colombia might face stricter constraints which prevents them to enter the higher education system, compared to students in the U.S. As part of our exploratory analysis, we also offer evidence to show that local employment prospects can be an important driver to induce students to enroll in higher education via Short-Cycle degrees.

4.6 Tables and Figures

Table 4.1: Summary Statistics

Variable	Mean	SD
Female	0.538	0.499
Age at Saber 11	17.019	1.583
Siblings	2.682	1.639
Household Income Level		
<1 Minimum Wages (MW)	0.272	0.445
1-2 MW	0.432	0.495
2-3 MW	0.159	0.366
>3 MW	0.137	0.344
Mother's Education Level		
Primary	0.423	0.494
Secondary	0.354	0.478
Short-Cycle program	0.109	0.312
At least Bachelor's program	0.113	0.317
Higher Education Enrollment		
Not enrolled (NE)	0.612	0.487
Short-Cycle program (SC)	0.095	0.293
Bachelor's program (BP)	0.293	0.455
Educational Attainment		
High School Graduate	0.612	0.487
Short-Cycle Incomplete	0.047	0.211
Short-Cycle Complete	0.036	0.186
Bachelor's Incomplete	0.156	0.363
Bachelor's Complete	0.149	0.356
Formal Labor Market Outcomes		
No participation (2008-2013)	0.418	0.493
Works in 2013	0.4902	0.499
Years of experience (2008-2013)	2.611	2.670
N	322,537	

Note: The sample corresponds to the universe of students who took the High School Exit Exam in 2005. The information on educational attainment comes from SPADIES, and the labor market outcomes for higher education graduates is from OLE. For high school graduates and those with higher education incomplete, we impute labor market participation and experience using household survey data from SEDLAC.

Table 4.2: Average characteristics, by type of enrollment

Variable	Short-Cycle Program	Bachelor's Program	Not Enrolled
Female	0.480	0.548	0.542
Age at Saber 11	16.777	16.543	17.284
Siblings	2.329	2.102	3.014
Household Income Level			
<1 MW	0.186	0.148	0.344
1-2 MW	0.506	0.349	0.461
2-3 MW	0.206	0.201	0.131
>3 MW	0.103	0.302	0.063
Mother's Education Level			
Primary	0.366	0.217	0.531
Secondary	0.432	0.323	0.357
Short-Cycle program	0.136	0.179	0.072
At least Bachelor's program	0.066	0.281	0.040
Standardized Test Scores from the High School Exit Exam			
Math	0.004	0.366	-0.148
Reading	0.132	0.561	-0.240
Biology	0.044	0.511	-0.218
Physics	0.005	0.345	-0.144
History	0.070	0.469	-0.202
Chemistry	0.030	0.510	-0.215
Geography	0.057	0.430	-0.182
Philosophy	0.064	0.412	-0.173
Education Attaintment			
SC Incomplete	0.494		
SC Complete	0.342	0.012	
BP Incomplete	0.102	0.501	
BP Complete	0.062	0.487	
Formal Labor Market Outcomes			
No participation (2008-2013)	0.364	0.371	0.449
Works in 2013	0.512	0.536	0.481
Experience (2008-2013)	2.395	2.070	2.903
N	30,514	94,583	197,440

Note: The sample corresponds to the universe of students who took the High School Exit Exam in 2005. The information on educational attainment comes from SPADIES, and the labor market outcomes for higher education graduates is from OLE. For high school graduates and those with higher education incomplete, we impute labor market participation and experience using household survey data from SEDLAC.

Table 4.3: Average characteristics of HS municipality, by type of enrollment

Variable	Short-Cycle Program	Bachelors' Program	Not Enrolled	All
Population density	2.853	2.464	1.955	2.189
Fraction of urban pop.	0.873	0.867	0.795	0.823
Linear distance to the main food market, km	38.022	53.553	64.414	58.732
Linear distance to the main city, km	22.627	22.882	34.545	29.997
Poverty Index	0.358	0.364	0.399	0.385
Homicide rate	0.350	0.350	0.386	0.372
GDP per capita	9.289	8.603	8.095	8.357
Unemployment rate	83.552	84.572	86.515	85.665
N	30,514	94,583	197,440	322,537

Note: The sample corresponds to the universe of students who took the High School Exit Exam in 2005. The information on local characteristics comes from the Municipal Panel from CEDE (*Centro de Estudios de Desarrollo Económico*, in Spanish).

Table 4.4: OLS Results: Labor Market Outcomes

Variable	Prob(Not in Formal Sector) 2008-2013	Prob(Working) 2013	Experience 2008-2013
$\hat{\beta}_{SC}$	-0.0169***	-0.0442***	-0.5216***
	(0.0044)	(0.0047)	(0.0247)
N	322,537	322,537	322,537
R^2	0.4337	0.4833	0.5472

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the high school level. All regressions include baseline characteristics at the individual and municipal level, as well as department fixed effects. Each outcome is estimated separately.

Table 4.5: Availability of Higher Education Institutions (HEIs), by type of programs offered

Variable	Short-Cycle Program	Bachelors' Program	Not Enrolled	All
Z_1: HEI only offers SCP	0.742	0.654	0.570	0.611
Other type of HEI	0.849	0.851	0.740	0.783
Enrollment in SENA>0 in 2004	0.798	0.819	0.710	0.750
N	30,514	94,583	197,440	322,537

Note: The sample corresponds to the universe of students who took the High School Exit Exam in 2005. The information on higher education programs and institutions is from SNIES (*Sistema Nacional de Información de la Educación Superior*, in Spanish).

Table 4.6: 2SLS Results: Labor Market Outcomes

Variable	Prob(Not in Formal Sector) 2008-2013	Prob(Working) 2013	Experience 2008-2013
$\hat{\beta}_{SC}$	-0.1848	0.1682	1.2483*
	(0.1307)	(0.1360)	(0.6553)
First-stage:			
$\hat{\gamma}_1^{SC}$	0.0312***		
	(0.003)		
N	322,537		
Kleibergen-Paap F-stat	40.441		
Cragg-Donald F-stat	104.643		
Effective F-stat	41.937		

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the high school level. All regressions include baseline characteristics at the individual and municipal level, as well as department fixed effects. Each outcome is estimated separately.

Table 4.7: First-stage: individual characteristics

Variable	Short-Cycle Program	Bachelor's Program	Not Enrolled
Female	-0.0243***	0.0302***	-0.0059***
	(0.0013)	(0.0020)	(0.0021)
Age at Saber 11	-0.0064***	-0.0243***	0.0307***
	(0.0005)	(0.0007)	(0.0008)
Siblings	-0.0034***	-0.0119***	0.0153***
	(0.0008)	(0.0015)	(0.0016)
Household Income			
1-2 MW	0.0126***	0.0150***	-0.0276***
	(0.0024)	(0.0038)	(0.0041)
>2 MW	0.0064**	0.0527***	-0.0591***
	(0.0030)	(0.0059)	(0.0065)
Mother's level of education			
Secondary	0.0133***	0.0581***	-0.0714***
	(0.0020)	(0.0033)	(0.0036)
Higher Education	-0.0043	0.2039***	-0.1996***
	(0.0032)	(0.0069)	(0.0069)
Standardized Test Scores from the High School Exit Exam			
Math	-0.0022***	0.0189***	-0.0167***
	(0.0006)	(0.0009)	(0.0009)
Reading	0.0062***	0.0434***	-0.0495***
	(0.0007)	(0.0010)	(0.0011)
Biology	-0.0003	0.0251***	-0.0249***
	(0.0007)	(0.0010)	(0.0010)
Physics	-0.0017***	0.0132***	-0.0115***
	(0.0006)	(0.0009)	(0.0009)
History	0.0014**	0.0219***	-0.0233***
	(0.0007)	(0.0009)	(0.0010)
Chemistry	-0.0005	0.0225***	-0.0221***
	(0.0007)	(0.0010)	(0.0010)
Geography	-0.0007	0.0202***	-0.0195***
	(0.0006)	(0.0009)	(0.0009)
Philosophy	0.0021***	0.0151***	-0.0172***
	(0.0006)	(0.0009)	(0.0009)
N	322,537	322,537	322,537

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the high school level. All regressions include department fixed effects. Each choice is estimated separately.

Table 4.8: First-stage (cont'd): HS municipality characteristics and proximity to HEIs

Variable	Short-Cycle Program	Bachelor's Program	Not Enrolled
Population Density	0.0040***	-0.0009	-0.0032**
	(0.0007)	(0.0010)	(0.0013)
Fraction of urban pop.	0.0546***	-0.0265***	-0.0281**
	(0.0075)	(0.0101)	(0.0114)
Distance to the main market, in km	-0.0002***	0.0001**	0.0001
	(0.0000)	(0.0001)	(0.0001)
Distance to main city, in km	0.0001	0.0001	-0.0001***
	(0.0000)	(0.0001)	(0.0001)
Poverty incidence	-0.0012	0.0612**	-0.0600*
	(0.0195)	(0.0292)	(0.0315)
Homicide rate	-0.0026	-0.0170**	0.0196**
	(0.0056)	(0.0076)	(0.0079)
GDP per capita	-0.0007	0.0010**	-0.0003
	(0.0006)	(0.0005)	(0.0006)
Unemployment rate	0.0033***	-0.0022***	-0.0011**
	(0.0003)	(0.0004)	(0.0004)
Instrument Z_1 :			
SC HEI in 10 km radius	0.0289***	-0.0280***	-0.0009
	(0.0033)	(0.0051)	(0.0054)
Availability of other HEI:			
Not specialized in SC, or only BP	0.0035	0.0114**	-0.0150***
	(0.0035)	(0.0046)	(0.0051)
SENA enrollment>0, 2004	-0.0122***	0.0147***	-0.0025
	(0.0030)	(0.0042)	(0.0045)
N	322,537	322,537	322,537

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the high school level. All regressions include department fixed effects. Each choice is estimated separately.

Table 4.9: Share of compliers along the expansion and diversion margins, for the full sample and by gender

	Expansion margin Compliers: NE-SC	Diversion margin Compliers: BP-SC	$\hat{\omega}$
All	0.0008	0.028***	0.965***
	(0.0054)	(0.0051)	(0.186)
Female	0.006	0.019***	0.748***
	(0.006)	(0.0063)	(0.254)
Male	0.005	0.032***	0.849***
	(0.0044)	(0.0061)	(0.156)

Note: The Table shows the estimated shares of students who react to the variation in the local supply of Short-Cycle HEI by: switching from not enrolling towards Short-Cycle programs, and by diverting from Bachelor's and into Short-Cycle programs. $\hat{\omega}$ is the ratio of the effect in the third column ("Diversion margin"), over the sum of column two and three. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Bootstrap standard errors in parentheses, clustered at the high school level.

Table 4.10: Local Average Treatment Effects for compliers along the diversion margin, full sample and by gender

$LATE_{SC-/SC}$	Prob(Not in Formal Sector) 2008-2013	Prob(Working) 2013	Experience 2008-2013
All	-0.2270**	0.0506	0.9207**
	(0.0893)	(0.0808)	(0.421)
Female	-0.4487***	0.2358**	1.7650***
	(0.129)	(0.112)	(0.405)
Male	-0.1745	0.0307	0.8577
	(0.1106)	(0.114)	(0.693)
Observations	322,537		

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Bootstrap standard errors in parentheses, clustered at the high school level.

Table 4.11: Average characteristics of compliers along the diversion margin, full sample and by gender

Variable	All	Female	Male
Age	18.655	20.057	17.837
Siblings	3.572	3.878	3.400
Income level: > 2 MW	0.185	0.054	0.254
Mother's education: Higher	0.266	0.196	0.307
Standardized Test Scores, Saber 11			
Math	-0.013	-0.398	0.214
Reading	0.314	0.274	0.344
Biology	0.209	-0.213	0.477
Physics	0.395	0.181	0.524
History	0.155	-0.043	0.285
Chemistry	0.158	-0.427	0.511
Geography	0.189	-0.108	0.383
Philosophy	0.242	0.107	0.328
Municipal Characteristics			
Fraction of urban pop.	0.581	0.375	0.697
Population Density	13.580	20.609	9.444
Distance to main market, km	147.159	220.108	103.669
Distance to main city, km	85.095	105.984	73.020
Poverty	0.689	0.847	0.596
Homicide rate	0.774	0.899	0.703
GDP per capita	13.320	17.127	11.104

Table 4.12: Subsample first-stage: HEI availability and Local employment conditions

Variable	Short-Cycle Program	Bachelor's Program	Not Enrolled
Local employment rates>department rate for ages 28-33, by education level			
High School	-0.0507	-0.0826***	0.1332***
	(0.0482)	(0.0207)	(0.0343)
Z_E: Short-Cycle programs	0.0884***	0.0111	-0.0995***
	(0.0235)	(0.0293)	(0.0270)
N	150,525	150,525	150,525
Compliers NE-SC/total variation	0.887***		
	(0.275)		

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the high school level. All regressions include metropolitan area (Carranza and Ferreyra, 2019) fixed effects, and controls at the municipal level. Each choice is estimated separately.

Table 4.13: Subsample 2SLS Results: Labor Market Outcomes

	Prob(Not in Formal Sector) 2008-2013		Prob(Working) 2013		Experience 2008-2013	
Variable	OLS	2SLS	OLS	2SLS	OLS	2SLS
$\hat{\beta}_{SC}$	-0.0563***	0.1031	0.0085	-0.1267	-0.2021***	-1.8011
	(0.0058)	(0.3238)	(0.0061)	(0.3182)	(0.0330)	(1.6670)
First-stage:						
$\hat{\gamma}_E^{SC}$	0.0774***					
	(0.0141)					
N	150,525					
Kleibergen-Paap F-stat	20.967					
Cragg-Donald F-stat	6.371					
Effective F-stat	16.359					

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the high school level. All regressions include metropolitan area (Carranza and Ferreyra, 2019) fixed effects, and controls at the municipal level. Each choice is estimated separately.

Appendix A:

A.1 Proofs

Convexity of costs and partial monotonicity Let $d \in \{1, 2, ..., k\}$ and d^{-} represent choices other than d. Let $U_{id}(z) = \mu_{id} - V_{id}(z)$ denote the utility of choice d evaluated at realization z of the instruments in $\mathbf{Z}$. μ_{id} represent unobserved preferences for choice d, $V_{id}(z)$ denote the costs of choosing option d. Let (z_l, z_{l^-}) and (z'_l, z_{l^-}) be two realizations of $Z_l, Z_{l^-} \in \mathbf{Z}$, and suppose $z_l < z'_l$. From element-wise convexity of V_{id} in R1 it follows that $\frac{\partial^2 V_{id}(Z_l, z_{l-})}{\partial Z_l^2} \geq 0$.

> For convex and decreasing costs, then $V_{id}(z_l, z_{l^-}) \geq V_{id}(z'_l, z_{l^-})\ \forall i$. Suppose that $Z_d \cap Z_{d^-} = \emptyset$, such that the restriction in R1.2 holds and there is no intersection in the instruments that affect each choice (i.e., $\frac{\partial V_{id-}(Z_l, z_{l-})}{\partial Z_l} = 0$). Then, $U_{id}(z_l, z_{l^-}) \leq U_{id}(z'_l, z_{l^-})$, and $U_{id^-}(z_l, z_{l^-}) = U_{id^-}(z'_l, z_{l^-})$. Thus, $(z_l, z_{l^-}) \rightarrow (z'_l, z_{l^-})$ weakly induces agents towards d such that $D_{id}(z_l, z_{l^-}) \leq D_{id}(z'_l, z_{l^-})\ \forall i$. Similarly for convex and increasing costs, then $V_{id}(z_l, z_{l^-}) \leq V_{id}(z'_l, z_{l^-})\ \forall i$. Suppose $\frac{\partial V_{id-}(Z_l, z_{l-})}{\partial Z_l} = 0$. Then, $U_{id}(z_l, z_{l^-}) \geq U_{id}(z'_l, z_{l^-})$, and $U_{id^-}(z_l, z_{l^-}) = U_{id^-}(z'_l, z_{l^-})$. Thus, $(z_l, z_{l^-}) \rightarrow (z'_l, z_{l^-})$ weakly induces agents away from d such that $D_{id}(z_l, z_{l^-}) \geq D_{id}(z'_l, z_{l^-})\ \forall i$.
>
> Consider the case of $Z_d \cap Z_{d^-} \neq \emptyset$ and let $Z_{l^-} \in Z_d \cap Z_{d^-}$, such that there is at least one instrument that affects d and d^-. From R1, it follows that $V_{id}(Z_l, Z_{l^-})$ and $V_{id^-}(Z_l, Z_{l^-})$ are element-wise convex. Suppose $\frac{\partial V_{id}(Z_l, Z_{l-})}{\partial Z_{l^-}} \leq 0$ and $\frac{\partial V_{id-}(Z_l, Z_{l-})}{\partial Z_{l^-}} \leq 0$ and assume $\frac{\partial^2 V_{id}(Z_l, z_{l-})}{\partial Z_l^2} > \frac{\partial^2 V_{id}(Z_l, z_{l-})}{\partial Z_l^2}$ such that costs decrease faster for d. Then $V_{id}(z_l, z_{l^-}) \leq V_{id}(z'_l, z_{l^-})$ implies $U_{id}(z_l, z_{l^-}) \geq U_{id}(z'_l, z_{l^-})$ and $(z_l, z_{l^-}) \rightarrow (z'_l, z_{l^-})$ weakly induces agents away from d such that $D_{id}(z_l, z_{l^-}) \leq D_{id}(z'_l, z_{l^-})\ \forall i$. Similarly, suppose $\frac{\partial V_{id}(Z_l, Z_{l-})}{\partial Z_{l^-}} \geq 0$ and $\frac{\partial V_{id-}(Z_l, Z_{l-})}{\partial Z_{l^-}} \geq 0$ and $\frac{\partial^2 V_{id}(Z_l, z_{l-})}{\partial Z_l^2} > \frac{\partial^2 V_{id}(Z_l, z_{l-})}{\partial Z_l^2}$. Then $V_{id}(z_l, z_{l^-}) \leq V_{id}(z'_l, z_{l^-})$ implies $U_{id}(z_l, z_{l^-}) \geq U_{id}(z'_l, z_{l^-})$ and $(z_l, z_{l^-}) \rightarrow (z'_l, z_{l^-})$ weakly induces agents

away from d such that $D_{id}(z_l, z_{l^-}) \leq D_{id}(z_l', z_{l^-})\ \forall i$.

Conditional/Partial monotonicity and identification of counterfactuals

Let R denote a response matrix and let $B_d = \mathbf{1}[R = d]$ for $d \in \text{supp}(D)$. That is, B_d is a binary matrix that takes the value of one every time d appears in response matrix R. Heckman and Pinto (2018) define unordered monotonicity such that:

A3 *(Unordered Monotonicity)* For any $z, z' \in \mathcal{Z}$[1] and each $d \in \{1, 2, 3, .., k\}$ either $\mathbb{1}[D_i(z) = d] \geq \mathbb{1}[D_i(z') = d]\ \forall i$, or, $\mathbb{1}[D_i(z) = d] \leq \mathbb{1}[D_i(z') = d]\ \forall i$.

Assumption A2 is a weaker version of A3, thus if A3 holds then A2 holds. The opposite does not hold; that is, partial monotonicity does not imply unordered monotonicity. For unconditional response matrix R, Heckman and Pinto (2018) show that R is unordered monotonic if and only if each $B_d = \mathbf{1}[R = d]$ is lonesum such that there are no two-way patterns in any sub-matrix of dimension 2×2. That is, there is no 2×2 sub-matrix in R that takes the form of:

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \qquad \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$$

For the IV Model in equations (2.2)-(2.3) with A1, unconditional response matrix R is not unordered monotonic. Nonetheless, from A2 it follows that $R_l(Z_{l^-})$ satisfies that each $B_l^d = \mathbf{1}[R_l(Z_{l^-}) = d]$ is lonesum, with $d \in \text{supp}(D)$. Suppose not. If B_l^d is not lonesum, then there is at least one 2×2 sub-matrix that takes either of the following forms:

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \qquad \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$$

Given that $B_l^d = \mathbf{1}[R_l(Z_{l^-}) = d]$, two-way patterns are defined by conditional response matrix $R_l(Z_{l^-})$. By definition of $R_l(Z_{l^-})$, there is a response vector $S_{la}(Z_{l^-})$ and $S_{lb}(Z_{l^-})$ such that $\mathbf{1}[S_{la}(Z_{l^-}) = d] = [1, 0]'$ and $\mathbf{1}[S_{lb}(Z_{l^-})] = [0, 1]'$. Then, $D_{ad}(z, Z_{l^-}) \geq D_{ad}(z', Z_{l^-})$ and $D_{bd}(z, Z_{l^-}) \leq D_{bd}(z', Z_{l^-})$. This contradicts A2. Hence, conditional response matrix $R_l(Z_{l^-})$ and conditional binary matrices for each d, $B_l^d(Z_{l^-})$, are lonesum. Then, applying Theorem T-6 in Heckman and Pinto (2018) to the case of conditional response matrix, the counterfactuals for responses

[1] Recall $z \in \mathcal{Z}$ is a combination of $(Z_{i,1}, ..., Z_{i,L})$

g are identified by:

$$E\left(Y_d \mid g \in \sum\nolimits_{dl}(q|Z_{l^-})\right) = \frac{b_{dl}(q|Z_{l^-})B_l^{d+}(Z_{l^-})L_{Yl}^d(Z_{l^-})}{b_{dl}(q|Z_{l^-})B_l^{d+}(Z_{l^-})\Pr_{Z_l}(D=d|Z_{l^-})}$$

where $B_l^{d+}(Z_{l^-})$ is the Moore-Penrose pseudoinverse.

A.2 Probability of groups and their average baseline characteristics

This appendix follows the identification results in Heckman and Pinto (2018), and the results for response vectors $S_l(z_{l^-})$ and binary matrix $B_l^d(Z_{l^-})$ for choice $d \in \{1, 2, ..., k\}$ in Section 2.2.3. Let $P_{gl}(Z_{l^-})$ represent a vector containing the shares of each type of response in matrix $R_l(Z_{l^-})$. These shares can be identified from:

$$P_g(Z_{l^-}) = B_l^+(Z_{l^-})\Pr_{Z_l}(D|Z_{l^-}) \tag{A.1}$$

where $B_l^+(Z_{l^-})$ is the Moore-Penrose pseudoinverse.[2] Vector $\Pr_{Z_l}(D|Z_{l^-})$ contains the propensity score for choices $d \in \{1, 2, ..., k\}$ evaluated at instrument Z_l, conditional on Z_{l^-}. I denote the shares as functions of Z_{l^-}, to indicate that the prevalence of different types of responses can vary depending on the evaluation point in Z_{l^-}. Let $\Pr_{Z_l}(D=d|Z_{l^-}) = [\Pr_{Z_l}(D=d|z_l, Z_{l^-}), \Pr_{Z_l}(D=d|z_l', Z_{l^-})]$ be the propensity score of choice d evaluated at realizations $z_l, z_l' \in \text{supp}(Z_l)$, conditional on Z_{l^-}. Thus,

$$\Pr_{Z_l}(D|Z_{l^-}) = [\Pr_{Z_l}(D=1|Z_{l^-}), \Pr_{Z_l}(D=2|Z_{l^-}), ..., \Pr_{Z_l}(D=k|Z_{l^-})]' \tag{A.2}$$

and its dimension is $(l \cdot k) \times 1$. Intuitively, the identification of shares of responses is a weighed combination of observed choices and potential behavior at different values of the instruments, given Z_{l^-}.

Let $E[X_{gl}(Z_{l^-})]$ denote the average of baseline variables and $E[Y_{gl}(Z_{l^-})]$ the average counterfactuals for response g in matrix matrix $R_l(Z_{l^-})$. To identify the average of baseline variables and counterfactuals of the responses in matrix $R_l(Z_{l^-})$ define $\omega(Z_{l^-}) = B_l(Z_{l^-}) \circ P_{gl}(Z_{l^-})'$. That is, $\omega(Z_{l^-})$ is the element-wise multiplication of matrix $B_l(Z_{l^-})$ and the shares of responses in $P_{gl}(Z_{l^-})$. Let XD_d denote a vector of baseline variables and choices interactions, for $X \in \mathbf{X}$. For each choice $d \in \{1, ..., k\}$ define $L_{Xl}^d(Z_{l^-}) = [E[XD_d|z_l, Z_{l^-}], E[XD_d|z_l', Z_{l^-}]]$ which is the observed average of baseline variable X when option d is chosen, evaluated at real-

[2] For instance, if $B_l(Z_{l^-})$ has full-rank then $B_l^+(Z_{l^-}) = (B_l(Z_{l^-})'B_l(Z_{l^-}))^{-1} B_l(Z_{l^-})'$.

izations z_l, z_l' conditional on Z_{l^-}. Vector $L_{Xl}(Z_{l^-})$ stacks choice specific $L^d_{Xl}(Z_{l^-})$ such that $L_{Xl}(Z_{l^-}) = [L^{d=1}_{Xl}(Z_{l^-}), L^{d=2}_{Xl}(Z_{l^-}), ..., L^{d=k}_{Xl}(Z_{l^-})]'$. The average of baseline variables for each response, g, in matrix $R_l(Z_{l^-})$ corresponds to:

$$E[X_{gl}(Z_{l^-})] = \omega^{+}(Z_{l^-})L_{Xl}(Z_{l^-}) \tag{A.3}$$

The shares of responses, g, in equation A.1 and their average baseline variables in equation A.3 can be identified if $\text{rank}\,(B_l(Z_{l^-})) = N_{S,z_{l^-}}$.[3]

A.3 Multiple choices of childcare and identification of conditional LATEs

This section illustrates the identification problem that arises from using the unconditional variation in the lottery outcome. Let Y denote children development (e.g., cognitive, socio-emotional, or nutritional). Y_d denotes potential development, with $d \in \{h, s, l\}$. Observed development results from a switching regression model:

$$Y = Y_l D_l + Y_s D_s + Y_h D_h$$

where D_d takes the value of one if child care option d is chosen. The results in section 2.3.5 imply the following unconditional responses,

Table A.1: Unconditional responses to the lottery outcome

Z_1	g_1	g_2	g_3	g_4	g_5	g_6
0	h	s	l	h	s	s
1	h	s	l	l	l	h

From the switching regression model above, it follows that

$$\begin{aligned} E[Y|Z_1 = 1] &= E[Y_l D_l + Y_s D_s + Y_h D_h | Z_1 = 1] \\ &= E[Y_l D_l | Z_1 = 1] + E[Y_s D_s | Z_1 = 1] + E[Y_h D_h | Z_1 = 1] \\ &= E[Y_l | D_l = 1, Z_1 = 1]\Pr(D_l = 1 | Z_1 = 1) \\ &+ E[Y_s | D_s = 1, Z_1 = 1]\Pr(D_s = 1 | Z_1 = 1) \\ &+ E[Y_h | D_h = 1, Z_1 = 1]\Pr(D_h = 1 | Z_1 = 1) \end{aligned}$$

[3]This follows from the identification result in Heckman and Pinto (2018) for unconditional matrix B.

By A1 and the unconditional responses in the previous table,

$$\begin{aligned}E[Y|Z_1=1] = E[Y_l|D_l(0)=1,D_l(1)=1]Pr(D_l(0)=1,D_l(1)=1)\\+E[Y_l|D_h(0)=1,D_l(1)=1]Pr(D_h(0)=1,D_l(1)=1)\\+E[Y_l|D_s(0)=1,D_l(1)=1]Pr(D_s(0)=1,D_l(1)=1)\\+E[Y_s|D_s(0)=1,D_s(1)=1]Pr(D_s(0)=1,D_s(1)=1)\\+E[Y_h|D_h(0)=1,D_h(1)=1]Pr(D_h(0)=1,D_h(1)=1)\\+E[Y_h|D_s(0)=1,D_h(1)=1]Pr(D_s(0)=1,D_h(1)=1)\end{aligned}$$

Similarly,

$$\begin{aligned}E[Y|Z_1=0] = E[Y_l|D_l(0)=1,D_l(1)=1]Pr(D_l(0)=1,D_l(1)=1)\\+E[Y_s|D_s(0)=1,D_s(1)=1]Pr(D_s(0)=1,D_s(1)=1)\\+E[Y_s|D_s(0)=1,D_l(1)=1]Pr(D_s(0)=1,D_l(1)=1)\\+E[Y_s|D_s(0)=1,D_h(1)=1]Pr(D_s(0)=1,D_h(1)=1)\\+E[Y_h|D_h(0)=1,D_h(1)=1]Pr(D_h(0)=1,D_h(1)=1)\\+E[Y_h|D_h(0)=1,D_l(1)=1]Pr(D_h(0)=1,D_l(1)=1)\end{aligned}$$

Subtracting the conditional expectations,

$$\begin{aligned}E[Y|Z_1=1]-E[Y|Z_1=0] =\\ \underbrace{E[Y_l-Y_h|D_h(0)=1,D_l(1)=1]Pr(D_h(0)=1,D_l(1)=1)}_{l-h\text{ margin}}\\+\underbrace{E[Y_l-Y_s|D_s(0)=1,D_l(1)=1]Pr(D_s(0)=1,D_l(1)=1)}_{l-s\text{ margin}}\\+\underbrace{E[Y_h-Y_s|D_s(0)=1,D_h(1)=1]Pr(D_s(0)=1,D_h(1)=1)}_{h-s\text{ margin}}\end{aligned}$$

From Table A.1, $\Pr(g_4)=\Pr(D_h(0)=1,D_l(1)=1)$, $\Pr(g_5)=\Pr(D_s(0)=1,D_l(1)=1)$, and $\Pr(g_6)=\Pr(D_s(0)=1,D_h(1)=1)$. Replace in the reduced form:

$$\begin{aligned}E[Y|Z_1=1]-E[Y|Z_1=0] = E[Y_l-Y_h|D_h(0)=1,D_l(1)=1]\Pr(g_4)\\+E[Y_l-Y_s|D_s(0)=1,D_l(1)=1]\Pr(g_5)\\+E[Y_h-Y_s|D_s(0)=1,D_h(1)=1]\Pr(g_6)\end{aligned}$$

By an analogous argument, with A1 and from the responses in Table A.1:

$$\begin{aligned}
E[D_l|Z_1=1] &= \Pr(D_l=1, Z_1=1) \\
&= \Pr(D_l(1)=1) \\
&= \Pr(D_l(0)=1, D_l(1)=1) \\
&+ \Pr(D_h(0)=1, D_l(1)=1) + \Pr(D_s(0)=1, D_l(1)=1) \\
&= \Pr(g_3) + \Pr(g_4) + \Pr(g_5)
\end{aligned}$$

Similarly,

$$\begin{aligned}
E[D_l|Z_1=0] &= \Pr(D_l=1, Z_1=0) \\
&= \Pr(D_l(0)=1) \\
&= \Pr(D_l(0)=1, D_l(1)=1) \\
&= \Pr(g_3)
\end{aligned}$$

Thus,

$$E[D_l|Z_1=1] - E[D_l|Z_1=0] = \Pr(g_4) + \Pr(g_5)$$

If $\Pr(g_4)=0$ and $\Pr(g_6)=0$, then the Wald estimator $\beta^{\text{Wald}} = \frac{E[Y|Z_1=1]-E[Y|Z_1=0]}{E[D_l|Z_1=1]-E[D_l|Z_1=0]}$ can be written as:

$$\begin{aligned}
\frac{E[Y|Z_1=1]-E[Y|Z_1=0]}{E[D_l|Z_1=1]-E[D_l|Z_1=0]} &= \frac{E[Y_l - Y_s|D_s(0)=1, D_l(1)=1]\Pr(g_5)}{\Pr(g_5)} \\
&= E[Y_l - Y_s|D_s(0)=1, D_l(1)=1]
\end{aligned}$$

However, if $\Pr(g_4) \neq 0$, with $\Pr(g_6)=0$, then

$$\begin{aligned}
\frac{E[Y|Z_1=1]-E[Y|Z_1=0]}{E[D_l|Z_1=1]-E[D_l|Z_1=0]} &= \\
&\frac{E[Y_l - Y_s|D_s(0)=1, D_l(1)=1]\Pr(g_5)}{\Pr(g_4)+\Pr(g_5)} \\
&+ \frac{E[Y_l - Y_h|D_h(0)=1, D_l(1)=1]\Pr(g_4)}{\Pr(g_4)+\Pr(g_5)}
\end{aligned}$$

which denotes the effect of childcare choice l *versus* the next-best. That is, the choice parents would have made if l was no longer available. Meanwhile, if $\Pr(g_6) \neq 0$ then

$$\frac{E[Y|Z_1=1]-E[Y|Z_1=0]}{E[D_l|Z_1=1]-E[D_l|Z_1=0]} = \frac{E[Y_l-Y_s|D_s(0)=1,D_l(1)=1]\Pr(g_5)}{\Pr(g_4)+\Pr(g_5)} + \frac{E[Y_l-Y_h|D_h(0)=1,D_l(1)=1]\Pr(g_4)}{\Pr(g_4)+\Pr(g_5)} - \frac{E[Y_s-Y_h|D_s(0)=1,D_l(h)=1]\Pr(g_6)}{\Pr(g_4)+\Pr(g_5)}$$

In the latter, univariate IV identifies the effect of l *versus* the next-best and a bias component stemming from responses along the $s-h$ margin. Compared to the binary case, the last term resembles defiers and threatens the identification of treatment effects.

The role of conditional rules of parental choice of childcare

Following the empirical patterns of choice and the behavior described in Table 2.2, I apply the following restrictions: $\Pr(g_4|Z_2 > z_2^*) = \Pr(g_4^+) = 0$ and $\Pr(g_6|Z_2 < z_2^*) = \Pr(g_6^-) = 0$. To simplify notation, let z_2^- denote a realization of Z_2 such that $z_2^- < z_2^*$, and z_2^+ denote a realization of Z_2 such that $z_2^+ > z_2^*$. Thus,

$$\frac{E[Y|Z_1=1,z_2^-]-E[Y|Z_1=0,z_2^-]}{E[D_l|Z_1=1,z_2^-]-E[D_l|Z_1=0,z_2^-]} = \frac{E[Y_l-Y_s|D_s(0,z_2^-)=1,D_l(1,z_2^-)=1]\Pr(g_5^-)}{\Pr(g_4^-)+\Pr(g_5^-)} + \frac{E[Y_l-Y_h|D_h(0,z_2^-)=1,D_l(1,z_2^-)=1]\Pr(g_4^-)}{\Pr(g_4^-)+\Pr(g_5^-)}$$

which is the conditional effect of l *versus* the next-best. By a similar argument,

$$\frac{E[Y|Z_1=1,z_2^+]-E[Y|Z_1=0,z_2^+]}{E[D_s|Z_1=1,z_2^+]-E[D_s|Z_1=0,z_2^+]} = \frac{E[Y_s-Y_l|D_s(0,z_2^+)=1,D_l(1,z_2^+)=1]\Pr(g_5^+)}{\Pr(g_5^+)+\Pr(g_6^+)} + \frac{E[Y_s-Y_h|D_s(0,z_2^+)=1,D_h(1,z_2^+)=1]\Pr(g_6^+)}{\Pr(g_5^+)+\Pr(g_6^+)}$$

which denotes the conditional effect of s *versus* the next-best.

Separability of conditional LATEs: homogeneity assumption

Assume:

$$E[Y_s|D_s(0,z_2^+)=1,D_l(1,z_2^+)=1] = E[Y_s|D_s(0,z_2^+)=1,D_h(1,z_2^+)=1],$$

which states that the mean counterfactual outcome from choosing s would have been the same for compliers along the $s-l$ and $s-h$ margin. To show how to secure separate identification of conditional LATEs, first define the set of counterfactuals that can be identified from outcomes-choice interactions such as:

$$E[YD_l|Z_1=1,z_2^+]-E[YD_l|Z_1=0,z_2^+]= \\ E[Y_l|D_s(0,z_2^+)=1,D_l(1,z_2^+)=1]\Pr(g_5^+)$$

Similarly,

$$E[YD_h|Z_1=1,z_2^+]-E[YD_h|Z_1=0,z_2^+]= \\ E[Y_h|D_s(0,z_2^+)=1,D_h(1,z_2^+)=1]\Pr(g_6^+)$$

where $\Pr(g_5^+)$ and $\Pr(g_6^+)$ are identified from the first-stage:

$$E[D_l|Z_1=1,z_2^+]-E[D_l|Z_1=0,z_2^+]=\Pr(g_5^+)$$

$$E[D_h|Z_1=1,z_2^+]-E[D_h|Z_1=0,z_2^+]=\Pr(g_6^+)$$

hence, the following counterfactuals can be identified:

$$\frac{E[YD_l|Z_1=1,z_2^+]-E[YD_l|Z_1=0,z_2^+]}{E[D_l|Z_1=1,z_2^+]-E[D_l|Z_1=0,z_2^+]}=E[Y_l|D_s(0,z_2^+)=1,D_l(1,z_2^+)=1]$$

and,

$$\frac{E[YD_h|Z_1=1,z_2^+]-E[YD_h|Z_1=0,z_2^+]}{E[D_h|Z_1=1,z_2^+]-E[D_h|Z_1=0,z_2^+]}=E[Y_h|D_s(0,z_2^+)=1,D_h(1,z_2^+)=1]$$

In turn, the counterfactual of choosing s, Y_s, cannot be separately identified for each complier group without the homogeneity assumption,

$$\frac{E[YD_s|Z_1=1,z_2^+]-E[YD_s|Z_1=0,z_2^+]}{E[D_s|Z_1=1,z_2^+]-E[D_s|Z_1=0,z_2^+]}= \\ E[Y_s|D_s(0,z_2^+)=1,D_l(1,z_2^+)=1]\frac{\Pr(g_5^+)}{\Pr(g_5^+)+\Pr(g_6^+)} \\ + \\ E[Y_s|D_s(0,z_2^+)=1,D_h(1,z_2^+)=1]\frac{\Pr(g_6^+)}{\Pr(g_5^+)+\Pr(g_6^+)}$$

Imposing the homogeneity assumption,

$$\frac{E[YD_s|Z_1=1,z_2^+]-E[YD_s|Z_1=0,z_2^+]}{E[D_s|Z_1=1,z_2^+]-E[D_s|Z_1=0,z_2^+]} = E[Y_s|D_s(0,z_2^+)=1,D_l(1,z_2^+)=1]$$
$$= E[Y_s|D_s(0,z_2^+)=1,D_h(1,z_2^+)=1]$$

It follows that the following conditional LATEs can be identified:

$$\begin{aligned}\text{LATE}(z_2^+)_{s-l} &= \frac{E[YD_l|Z_1=1,z_2^+]-E[YD_l|Z_1=0,z_2^+]}{E[D_l|Z_1=1,z_2^+]-E[D_l|Z_1=0,z_2^+]} \\ &\quad - \frac{E[YD_s|Z_1=1,z_2^+]-E[YD_s|Z_1=0,z_2^+]}{E[D_s|Z_1=1,z_2^+]-E[D_s|Z_1=0,z_2^+]} \\ &= E[Y_l-Y_s|D_s(0,z_2^+)=1,D_l(1,z_2^+)=1]\end{aligned}$$

$$\begin{aligned}\text{LATE}(z_2^+)_{h-l} &= \frac{E[YD_h|Z_1=1,z_2^+]-E[YD_h|Z_1=0,z_2^+]}{E[D_h|Z_1=1,z_2^+]-E[D_h|Z_1=0,z_2^+]} \\ &\quad - \frac{E[YD_s|Z_1=1,z_2^+]-E[YD_s|Z_1=0,z_2^+]}{E[D_s|Z_1=1,z_2^+]-E[D_s|Z_1=0,z_2^+]} \\ &= E[Y_h-Y_s|D_s(0,z_2^+)=1,D_h(1,z_2^+)=1]\end{aligned}$$

Testing homogeneity

If $\text{LATE}(z_2^-)_{s-l}$ and $\text{LATE}(z_2^+)_{s-l}$ are separately identified, then

$$\lim_{Z_2\to z_2^{*+}} E[Y_l-Y_s|D_s(0,z_2^-)=1,D_l(1,z_2^-)=1] =$$
$$\lim_{Z_2\leftarrow z_2^{*-}} E[Y_l-Y_s|D_s(0,z_2^+)=1,D_l(1,z_2^+)=1]$$

That is, if the counterfactuals for complier group $s-l$ are indentified then in a neighborhood of z_2^* they should be comparable. In addition, the groups should be similar in their average baseline characteristics.

A.4 Alternative cost structure: disruption of existing childcare supply

Assume that $V_s(0) < V_s(1)$, such that the lottery increased the cost of choosing s. Assume that $V_s(0) < V_h$ and $V_s(0) < V_l(0)$, which states that in the absence of the lottery the costs of s would have been lower than those of h or l. The underlying assumption is that in the absence of the lottery, parents would have not changed their childcare decisions. Since all children in the sample were at s initially, I assume the cost of s are below all other child care alternatives before the expansion of supply took place.

Figure A.1: Relative costs that induce parents away from choosing s

(a) $s - l$ and $s - h$ margins (b) $s - l$ margin

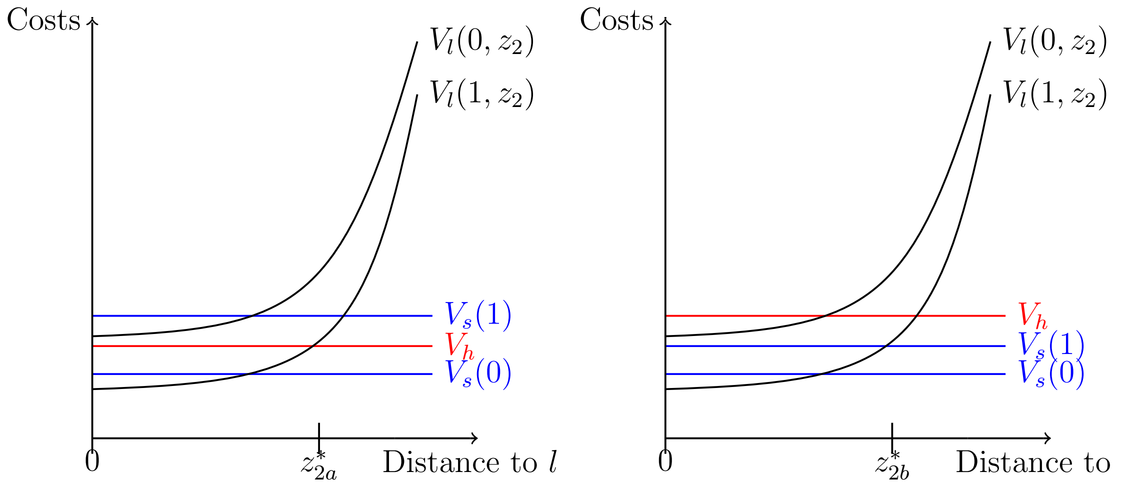


Figure A.1 shows examples of costs of childcare that would induce parents to switch away from s and into h and l. Panel (a) displays a case where below z^*_{2a} the costs of l are below the costs of s and h, when winning the lottery. Above z^*_{2a}, h is the cheapest childcare option. In this scenario, parents would be induced to switch between $s - l$ until the point where distance to the center is too high, at z^*_{2a}. After such distance to the large center, parents would switch between s and into h. These are parents who, because of changes in the supply of care, face higher costs of their existing childcare s and this increase is large enough to induce them to change their choice towards h or l. In panel (b), in contrast, there is also an increase in the cost of s but not large enough to induce parents towards h. These scenario would represent parents who face higher costs of caring for their children at home, with less family support available to do so. As a result, below z^*_{2b} parents switch away from their existing childcare option s and into l. Meanwhile, above z^*_{2b} s is the least expensive option and parents behave as s-always-takers.

B.1 Within class composition of Hybrid Schools

Given that Hybrid schools can offer either type of track, its effect on educational achievement are likely to differ depending on wheter or not the student choose a vocational track within H schools. To illustrate, let HA refer to academic tracks within Hybrid schools. Potential outcomes at school H are a combination of potential outcomes for those who choose academic and vocational tracks:

$$Y_H = D_{HA}Y_{HA} + D_{HV}Y_{HV},$$

where D_{HA} and D_{HV} are binary variables that take the value of one for students who choose academic and vocational tracks within H schools, respectively. I decompose the effect of A *versus* H schools among Z_A compliers as follows:

$$\begin{aligned}
&E[Y_A - Y_H|D_H(z) = 1, D_A(z') = 1] \\
&\qquad = E[Y_A - (D_{HA}Y_{HA} + D_{HV}Y_{HV})|D_H(z) = 1, D_A(z') = 1] \\
&\qquad = E[Y_A - D_{HA}Y_{HA} - D_{HV}Y_{HV} - D_{HA}Y_A \\
&\qquad + D_{HA}Y_A|D_H(z) = 1, D_A(z') = 1] \\
&\qquad = E[D_{HA}(Y_A - Y_{HA}) + D_{HV}(Y_A - Y_{HV})|D_H(z) = 1, D_A(z') = 1] \\
&\qquad = \theta_A E[Y_A - Y_{HA}|D_{HA}(z) = 1, D_A(z') = 1] \\
&\qquad + \theta_V E[Y_A - Y_{HV}|D_{HV}(z) = 1, D_A(z') = 1]
\end{aligned}$$

where θ_A denotes the share of students in H schools who choose an Academic track, and θ_V denotes the share of students in H schools who choose a Vocational track, given $\{D_H(z) = 1, D_A(z') = 1\}$. Hence, the effect of A *versus* H is a combination of the effect of H-academic *versus* A, and H-vocational *versus* A. Note that the first component compares students who choose A tracks but in different school-types (one specialized, A, and one offering also vocational tracks). Since the choice of track is fixed, this first component would capture school effects such as specialization, quality, and class composition. In turn, the second component combines the effects of differences in track (academic and vocational), school specialization, quality, and class composition.[4]

[4] A similar argument follows for $E[Y_H - Y_V|D_V(z) = 1, D_H(z') = 1]$, which also combines effects of track choice among H students.

C.1 Data restrictions

The master sample consists of the universe of test-takers of Saber 11 in 2005, who were between 14-24 years of age and account for 93.5% of the full set of students taking the test in that year. We restrict the sample to students who have data on all socioeconomic variables and test scores, which removes only 0.3% of students. We matched data on enrollment on higher education from SPADIES to the master sample, and restrict the start date of the first program recorded to be after the semester of 2005 when the student took the test, which removes 1.1% of the observations. We set an enrollment window of five years after graduation, and remove students who started their program on or after 2012 (about 3% of the sample). We also restrict the graduation year to be no less than two years- for those in short-cycle programs- and four years- for those in four-year college (removes 1.5% observations). In addition, we set the graduation age to be between 19-30 which reduces the sample by 0.1%. We remove 0.004% of the sample which corresponds to those who switched, or enrolled, in more than four programs between 2005 and 2015.

We further restrict the sample of those who graduated and have data on wages from OLE to start working at, or after, 20 years of age (removes 0.25% of the sample). For almost 4.8% of students who appear in OLE, enrollment and graduation do not match with SPADIES: two percent show as graduated in OLE but only enrolled in SPADIES, while the remaining 2.8% graduated according to OLE but do not report ever enrolling in a higher education institution in SPADIES. We removed the latter 2.8% from the sample, given that they never report enrolling in SPADIES, but since OLE contains information on wages we assume that the 2% who enrolled but never graduated in SPADIES actually did so. Lastly, 4% of those who graduated in SPADIES do not appear in OLE which means that for this fraction of students there is no information on wages. The final dataset consists of $369,541$ students, of a total of $403,209$ initially in the master sample (about 91% of students who took Saber 11 in 2005 are in our estimating sample).

Finally, we restrict our estimating sample to high schools with more than 20 students. The reason for this restriction is twofold: our instruments are constructed based on the municipality of the student's high school, and we estimate with clustered standard errors at this level. If the clusters are too small we won't be able to do inference. This restriction leaves an estimating sample of $322,537$ observations.

C.2 Imputation of wages and labor market participation

To impute participation in the labor market and average monthly wages, we use household survey data between 2008 and 2013. We use the set of homogenized household surveys in SEDLAC, which correspond to the Integrated Household Survey (GEIH, in Spanish). We restrict each sample to those individuals who were 14-24 old in 2005. Let Y denote labor market variables such as participation and wages.

$$Y = \beta_0 + \beta_X X + u$$

For labor market participation, we estimate the equation above with a LPM. X contains sex, age, number of household members, household income level, a dummy that takes the value of one if the individual lives in an urban area, and regional fixed effects (results in Table 2). For average wages, we estimate the equation above with a quantile regression and predict coefficients for the 25th, 50th, and 75th percentile. In this case X contains sex, age, age squared, number of household members, household income level, a dummy that takes the value of one if the individual lives in an urban area, and regional fixed effects (results are in Table 3 and 4).

Table 2: Regression Results: LPM of probability of working in 2013

Variables	HS Graduates	HE Incomplete
Male	0.350***	0.135***
	(0.014)	(0.017)
Age	0.013***	0.027***
	(0.003)	(0.003)
Number of members in main household	-0.019***	-0.028***
	(0.005)	(0.005)
$1-2$ MW	0.080***	0.140***
	(0.027)	(0.043)
$2-3$ MW	0.154***	0.254***
	(0.031)	(0.043)
> 3 MW	0.190***	0.340***
	(0.031)	(0.040)
Urban area	0.020	-0.030
	(0.022)	(0.035)
Region: Oriental	0.068**	0.040*
	(0.028)	(0.023)
Region: Central	-0.023	0.038*
	(0.025)	(0.021)
Region: Pacifica	0.000	-0.015
	(0.028)	(0.025)
Region: Bogotá	0.067*	0.067**
	(0.034)	(0.027)
Observations	4,722	8,141

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table 3: Regression Results: hourly wage in main occupation

	High School Graduates		
Variables	q25	q50	q75
Male	461.627***	434.299***	172.838
	(100.972)	(105.942)	(143.657)
Age	437.637	131.550	-6.856
	(284.343)	(273.086)	(328.158)
Age^2	-8.053	-1.763	0.952
	(5.356)	(5.101)	(6.104)
Number of members in main household	-117.627***	-121.458***	-173.951***
	(16.687)	(17.445)	(24.616)
$1-2$ MW	1,203.005***	1,318.647***	1,507.471***
	(112.909)	(128.563)	(158.184)
$2-3$ MW	1,419.765***	1,598.469***	2,040.413***
	(150.009)	(138.681)	(180.284)
> 3 MW	2,023.496***	2,102.490***	2,847.308***
	(142.385)	(153.329)	(277.423)
Urban area	-236.529**	-315.042**	-756.664***
	(102.791)	(151.966)	(237.149)
Region: Oriental	315.999**	367.526***	197.873
	(135.960)	(123.872)	(142.295)
Region: Central	155.621	236.217	354.958**
	(128.751)	(152.566)	(164.369)
Region: Pacifica	-110.810	-113.379	-47.057
	(135.958)	(116.014)	(231.480)
Region: Bogotá	655.060***	510.973***	326.354*
	(189.608)	(150.499)	(196.395)
Constant	-4,739.014	-359.963	2,686.551
	(3,726.144)	(3,609.618)	(4,344.162)
Observations	3,090	3,090	3,090

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the high school level.

Table 4: Regression Results: hourly wage in main occupation

	Higher Education Incomplete		
Variables	q25	q50	q75
Male	240.217***	342.924***	423.750**
	(89.161)	(86.817)	(184.014)
Age	680.753**	210.822	-61.922
	(265.982)	(295.035)	(509.594)
Age^2	-11.772**	-2.368	3.341
	(4.981)	(5.542)	(9.509)
Number of members in main household	-208.221***	-277.203***	-438.176***
	(26.766)	(21.133)	(41.265)
$1-2$ MW	1,503.986***	1,853.146***	1,342.269***
	(182.953)	(159.451)	(265.630)
$2-3$ MW	2,170.080***	2,771.923***	2,532.960***
	(192.120)	(173.614)	(278.575)
> 3 MW	3,077.149***	3,808.032***	4,681.549***
	(151.087)	(160.508)	(285.479)
Urban area	-408.050***	-308.839***	-229.820
	(117.800)	(92.751)	(269.067)
Region: Oriental	341.604**	368.984***	687.261**
	(141.283)	(134.636)	(293.303)
Region: Central	31.376	-175.898*	-285.977
	(119.496)	(90.773)	(193.988)
Region: Pacifica	152.830	183.826	103.817
	(139.976)	(154.098)	(225.862)
Region: Bogotá	378.608***	243.773**	155.136
	(127.699)	(124.243)	(227.950)
Constant	-8,028.344**	-1,587.394	3,238.408
	(3,506.050)	(3,893.089)	(6,742.760)
Observations	5,378	5,378	5,378

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the high school level.

Bibliography

Acton, R. (2020, April). Effects of Reduced Community College Tuition on College Choices and Degree Completion. *Education Finance and Policy*, 1–71.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association 91*(434), 444–455. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Attanasio, O., F. Cunha, and P. Jervis (2019, December). Subjective Parental Beliefs. Their Measurement and Role. Technical Report w26516, National Bureau of Economic Research.

Attanasio, O. P., V. D. Maro, and M. Vera-Hernández (2013, September). Community Nurseries and the Nutritional Status of Poor Children. Evidence from Colombia. *The Economic Journal 123*(571), 1025–1058. Publisher: Oxford Academic.

Barrera-Osorio, F. and H. Bayona-Rodríguez (2019, June). Signaling or better human capital: Evidence from Colombia. *Economics of Education Review 70*, 20–34.

Bassi, M. and S. Urzua (2010). Educación en Chile: El desafío está en la calidad | Publications. *Notas Técnicas* (IDB-TN-271).

Bernal, R., O. Attanasio, X. Peña, and M. Vera-Hernández (2019, April). The effects of the transition from home-based childcare to childcare centers on children's health and development in Colombia. *Early Childhood Research Quarterly 47*, 418–431.

Bernal, R. and C. Fernández (2013, November). Subsidized childcare and child development in Colombia: Effects of Hogares Comunitarios de Bienestar as a function of timing and length of exposure. *Social Science & Medicine 97*, 241–249.

Bernal, R. and S. M. Ramírez (2019, June). Improving the quality of early childhood care at scale: The effects of "From Zero to Forever". *World Development 118*, 91–105.

Bertrand, M., M. Mogstad, and J. Mountjoy (2021). Improving Educational Pathways to Social Mobility: Evidence from Norway's "Reform 94". *Journal of Labor Economics*, 713009.

Bettinger, E. and A. Soliz (2016, October). *Returns to Vocational Credentials: Evidence from Ohio's Community and Technical Colleges. A CAPSEE Working Paper.* Center for Analysis of Postsecondary Education and Employment. Publication Title: Center for Analysis of Postsecondary Education and Employment.

Betts, J. R. (2011, January). Chapter 7 - The Economics of Tracking in Education. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 3, pp. 341–381. Elsevier.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013, October). Salience and Consumer Choice. *Journal of Political Economy 121*(5), 803–843. Publisher: The University of Chicago Press.

Brunello, G. and D. Checchi (2007, October). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy 22*(52), 782–861.

Brunner, E., S. Dougherty, and S. Ross (2019). The Effects of Career and Technical Education: Evidence from the Connecticut Technical High School System. *EdWorkingPaper No. 19-112*.

Caetano, C. and J. C. Escanciano (2020). Identifying Multiple Marginal Effects With a Single Instrument. *Econometric Theory*, 1–31. Publisher: Cambridge University Press.

Camacho, A., J. Messina, and J. Uribe Barrera (2017, February). The Expansion of Higher Education in Colombia: Bad Students or Bad Programs? SSRN Scholarly Paper ID 2921965, Social Science Research Network, Rochester, NY.

Canaan, S. (2020, July). The long-run effects of reducing early school tracking. *Journal of Public Economics 187*, 104206.

Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling, Aspects of labour market behaviour: essays in honour of John Vanderkamp. ed. *LN Christofides, EK Grant, and R. Swidinsky*.

Carranza, J. E. and M. M. Ferreyra (2019). Increasing Higher Education Access: Supply, Sorting, and Outcomes in Colombia. *Journal of Human Capital 13*(1), 95–136. Publisher: University of Chicago.

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014). Local Identification of Nonparametric and Semiparametric Models. *Econometrica 82*(2), 785–809. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA9988.

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007, July). Instrumental variable estimation of nonseparable models. *Journal of Econometrics 139*(1), 4–14.

Chetverikov, D., A. Santos, and A. M. Shaikh (2018, August). The Econometrics of Shape Restrictions. *Annual Review of Economics 10*(1), 31–63. Publisher: Annual Reviews.

Chmielewski, A. K. (2014, May). An International Comparison of Achievement Inequality in Within- and Between-School Tracking Systems. *American Journal of Education 120*(3), 293–324. Publisher: The University of Chicago Press.

Cunha, F. (2015). Subjective Rationality, Parenting Styles, and Investments in Children. In P. R. Amato, A. Booth, S. M. McHale, and J. Van Hook (Eds.), *Families in an Era of Increasing Inequality: Diverging Destinies*, National Symposium on Family Issues, pp. 83–94. Cham: Springer International Publishing.

Cunha, F., I. Elo, and J. Culhane (2013, June). Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation. Technical Report w19144, National Bureau of Economic Research.

Das, M. (2005, February). Instrumental variables estimators of nonparametric models with discrete endogenous regressors. *Journal of Econometrics 124*(2), 335–361.

Dean, J. T. and S. Jayachandran (2019). Attending kindergarten improves cognitive but not socioemotional development in India.

Denning, J. T. (2017, May). College on the Cheap: Consequences of Community College Tuition Reductions. *American Economic Journal: Economic Policy 9*(2), 155–188.

D'Haultfœuille, X. and P. Février (2015). Identification of Nonseparable Triangular Models With Discrete Instruments. *Econometrica 83*(3), 1199–1210. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA10038.

Dougherty, S. M. (2018, March). The Effect of Career and Technical Education on Human Capital Accumulation: Causal Evidence from Massachusetts. *Education Finance and Policy 13*(2), 119–148.

Elacqua, G., M. F. Prada, P. Navarro-Palau, and S. Soares (2019, August). Does Technical Education Improve Academic Outcomes?: Evidence from Brazil. Technical report, Inter-American Development Bank.

Feng, J. (2020, July). Matching Points: Supplementing Instruments with Covariates in Triangular Models. *arXiv:1904.01159 [econ]*. arXiv: 1904.01159.

Ferreyra, M. M., C. Avitabile, J. Botero Álvarez, F. Haimovich Paz, and S. Urzúa (2017). *At a Crossroads : Higher Education in Latin America and the Caribbean.* Directions in Development—Human Development. World Bank, Washington, DC.

Ferreyra, M. M., A. Franco Hernandez, T. Melguizo, and A. M. Sanchez Diaz (2020, October). Estimating the Contribution of Short-Cycle Programs to Student Outcomes in Colombia.

Freyberger, J. and J. L. Horowitz (2015, November). Identification and shape restrictions in nonparametric instrumental variables estimation. *Journal of Econometrics 189*(1), 41–53.

Goff, L. (2020, September). A Vector Monotonicity Assumption for Multiple Instruments. *arXiv:2009.00553 [econ]*. arXiv: 2009.00553.

González-Velosa, C., G. Rucci, M. Sarzosa, and S. Urzúa (2015). Returns to Higher Education in Chile and Colombia | Publications. *IDB WORKING PAPER SERIES IDB-WP-587*.

Hanushek, E. A., G. Schwerdt, L. Woessmann, and L. Zhang (2017). General Education, Vocational Education, and Labor-Market Outcomes over the Lifecycle. *Journal of Human Resources 52*(1), 48–87. Publisher: University of Wisconsin Press.

Heckman, J. J. and R. Pinto (2018). Unordered Monotonicity. *Econometrica 86*(1), 1–35. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA13777.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006, August). Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432. Publisher: MIT Press.

Heckman, J. J. and S. Urzúa (2010, May). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics 156*(1), 27–37.

Heckman, J. J., S. Urzúa, and E. Vytlacil (2008). Instrumental Variables in Models with Multiple Outcomes: the General Unordered Case. *Annales d'Économie et de Statistique* (91/92), 151–174. Publisher: [GENES, ADRES].

Hojman, A. and F. López Bóo (2019, August). Cost-Effective Public Daycare in a Low-Income Economy Benefits Children and Mothers. Technical report, Inter-American Development Bank.

Hull, P. (2018, February). IsoLATEing: Identifying Counterfactual-Specific Treatment Effects with Cross-Stratum Comparisons. SSRN Scholarly Paper ID 2705108, Social Science Research Network, Rochester, NY.

Imbens, G. W. and J. D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–475. Publisher: [Wiley, Econometric Society].

Imbens, G. W. and W. K. Newey (2009). Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity. *Econometrica 77*(5), 1481–1512. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA7108.

Jepsen, C., K. Troske, and P. Coomes (2014, January). The Labor-Market Returns to Community College Degrees, Diplomas, and Certificates. *Journal of Labor Economics 32*(1), 95–121. Publisher: The University of Chicago Press.

Kane, T. J. and C. E. Rouse (1999, March). The Community College: Educating Students at the Margin between College and Work. *Journal of Economic Perspectives 13*(1), 63–84.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016, August). Field of Study, Earnings, and Self-Selection*. *The Quarterly Journal of Economics 131*(3), 1057–1111.

Kline, P. and C. R. Walters (2016, November). Evaluating Public Programs with Close Substitutes: The Case of Head Start. *The Quarterly Journal of Economics 131*(4), 1795–1848. Publisher: Oxford Academic.

Kreisman, D. and K. Stange (2020, January). Vocational and Career Tech Education in American High Schools: The Value of Depth Over Breadth. *Education Finance and Policy 15*(1), 11–44.

Kutscher, M., S. Nath, and S. Urzúa (2020). Centralized Admission Systems and School Segregation: Evidence from a National Reform. pp. 52.

Larrañaga, O., G. Cabezas, and F. Dussaillant (2013). Informe completo del estudio de la Educación Técnico Profesional | El PNUD en Chile. *UNDP*.

Lee, S. and B. Salanié (2018). Identifying Effects of Multivalued Treatments. *Econometrica 86*(6), 1939–1963. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA14269.

Lee, S. and B. Salanié (2020, July). Filtered and Unfiltered Treatment Effects with Targeting Instruments.

Leigh, D. E. and A. M. Gill (2003, February). Do community colleges really divert students from earning bachelor's degrees? *Economics of Education Review 22*(1), 23–30.

Malamud, O. and C. Pop-Eleches (2010, February). General Education versus Vocational Training: Evidence from an Economy in Transition. *The Review of Economics and Statistics 92*(1), 43–60.

Marcotte, D. E. (2019, September). The Returns to Education at Community Colleges: New Evidence from the Education Longitudinal Survey. *Education Finance and Policy 14*(4), 523–547.

Matzkin, R. L. (1991). Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models. *Econometrica 59*(5), 1315–1327. Publisher: [Wiley, Econometric Society].

Minaya, V. and J. Scott-Clayton (2020, August). Labor Market Trajectories for Community College Graduates: How Returns to Certificates and Associate's degrees Evolve Over Time. *Education Finance and Policy*, 1–62.

Minnesota Population Center (2019). Integrated Public Use Microdata Series, International. Version 7.2 [dataset]. Minneapolis, MN: IPUMS, 2019. https://doi.org/10.18128/D020.V7.2.

Mogstad, M., A. Torgovitsky, and C. Walters (2020a). The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables. Technical Report w25691, National Bureau of Economic Research, Cambridge, MA.

Mogstad, M., A. Torgovitsky, and C. R. Walters (2020b, July). Policy Evaluation with Multiple Instrumental Variables. Working Paper 27546, National Bureau of Economic Research. Series: Working Paper Series.

Mountjoy, J. (2019, February). Community Colleges and Upward Mobility. SSRN Scholarly Paper ID 3373801, Social Science Research Network, Rochester, NY.

Mulcahy, H. and E. Savage (2016). Uncertainty: A little bit not sure. Parental concern about child growth or development. *Journal of Child Health Care: For Professionals Working with Children in the Hospital and Community 20*(3), 333–343.

Ollikainen, J.-P. (2021, January). Comprehensive school reform and labor market outcomes over the lifecycle: Evidence from Finland. *Labour Economics 68*, 101952.

Pekkarinen, T., R. Uusitalo, and S. Kerr (2009, August). School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics 93*(7), 965–973.

Pinto, R. (2019). Noncompliance as a Rational Choice: A Framework that Exploits Compromises in Social Experiments to Identify Causal Effects. *Unpublished manuscript*, 63.

Rau, T., S. Urzúa, and L. Reyes (2015, September). Early Exposure to Hazardous Waste and Academic Achievement: Evidence from a Case of Environmental Negligence. *Journal of the Association of Environmental and Resource Economists 2*(4), 527–563. Publisher: The University of Chicago Press.

Rodriguez, J. and F. Saltiel (2020). Preschool Attendance, Parental Investment, and Child Development: Experimental Evidence from Bangladesh.

Rouse, C. E. (1995). Democratization or Diversion? The Effect of Community Colleges on Educational Attainment. *Journal of Business & Economic Statistics 13*(2), 217–224. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Sacerdote, B. (2011, January). Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 3, pp. 249–277. Elsevier.

Sanchez, C. (2018). Understanding School Competition Under Voucher Regimes. pp. 84.

Sevilla, M. P. and L. Sepúlveda (2015). Selección en liceos polivalentes. Análisis de los mecanismos de adscripción de estudiantes a la modalidad científico humanista o técnico profesional. *Fondo de Investigación y Desarrollo en Educación-FONIDE, MINEDUC.*.

Silliman, M. and H. Virtanen (2019). Labor Market Returns to Vocational Secondary Education. Working Paper 65, ETLA Working Papers.

Soto, R. and A. Torche (2004). Spatial Inequality, Migration, and Economic Growth in Chile. *Cuadernos de Economía 41*(124), 401–424. Publisher: Instituto de Economia, Pontificia Universidad Catolica de Chile.

Vytlacil, E. (2002). Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica 70*(1), 331–341. Publisher: [Wiley, Econometric Society].

Wang, F., E. Puentes, J. Behrman, and F. Cunha (2020, April). You Are What Your Parents Think: Height and Local Reference Points. SSRN Scholarly Paper ID 3167023, Social Science Research Network, Rochester, NY.

Xie, H. (2020, January). Generalized Local IV with Unordered Multiple Treatment Levels: Identification, Efficient Estimation, and Testable Implication. *arXiv:2001.06746 [econ]*. arXiv: 2001.06746.

Zilic, I. (2018, February). General versus vocational education: Lessons from a quasi-experiment in Croatia. *Economics of Education Review 62*, 1–11.