

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Essays on Education, Housing and Public Policy

Permalink

<https://escholarship.org/uc/item/46p6s4s5>

ISBN

9798288861895

Author

Fuenzalida Bello, Joaquin

Publication Date

2025-05-15

Peer reviewed|Thesis/dissertation

Essays on Education, Housing and Public Policy

by

Joaquin Fuenzalida Bello

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

David Card, Chair

Emmanuel Saez

Rucker Johnson

Spring 2025

Essays on Education, Housing and Public Policy

Copyright 2025
by
Joaquin Fuenzalida Bello

Abstract

Essays on Education, Housing and Public Policy

by

Joaquin Fuenzalida Bello

Doctor of Philosophy in Economics

University of California, Berkeley

David Card, Chair

This dissertation examines the effects of public policies in housing and education markets, highlighting how targeted interventions influence educational outcomes, housing affordability, and market dynamics. Chapter 1 explores the intergenerational impact of homeownership subsidies, providing novel evidence from a large-scale housing policy in Santiago, Chile. Leveraging quasi-experimental variation, my coauthors and I show that subsidies significantly enhance children's educational achievements, attendance, and early labor market outcomes. Improvements are particularly pronounced for boys and children from larger families, highlighting reduced household overcrowding and increased parental engagement as crucial mediators of these benefits. Chapter 2 examines the equilibrium effects of online education expansion on higher education markets in Brazil. Utilizing detailed administrative data and a structural equilibrium model, my coauthors and I document how the rapid proliferation of online degrees reshapes market structure, lowers tuition fees, and redistributes enrollment patterns, benefiting older students but diverting younger students away from high-quality in-person education. These findings underline the critical role that regulatory frameworks can play in balancing accessibility and educational quality. Finally, Chapter 3 investigates housing supply elasticity using Chile's homebuyer subsidy program as a quasi-experimental shock to housing demand. Through detailed property transaction data, my coauthor and I estimate a low elasticity of housing supply, uncovering that subsidies primarily lead to price increases rather than significant expansions in supply, especially in suburban areas. This low responsiveness highlights inefficiencies in the subsidy design, suggesting the need for complementary supply-side interventions. Together, these chapters emphasize the nuanced ways in which housing and education policies influence economic and social outcomes, demonstrating the importance of context-specific evaluations and highlighting opportunities to optimize public investments for greater equity and efficiency.

Acknowledgments

Coauthor information: Chapter 1 is coauthored with Felipe Vial and Harrison Wheeler. Chapter 2 is coauthored with Nano Barahona, Caue Dobbin, and Sebastián Otero. Chapter 3 is coauthored with Felipe Sepúlveda. I’m incredibly thankful for all your guidance, the commitment you brought to our work, and for showing me how rewarding research can be.

I would like to express my sincere gratitude to David Card, Javier Feinmann, Marco Gonzalez-Navarro, Rucker Johnson, Patrick Kline, Timothy McQuade, Ricardo Perez-Truglia, Demian Pouzo, Emmanuel Saez, Felipe Sepúlveda, and Cristopher Walters for invaluable feedback and guidance. I am also thankful for the many discussions with seminar participants at UC Berkeley that have helped shape my work. I gratefully acknowledge financial support from the IRLE, the Burch Center, Stone Center, and the Fisher Center at UC Berkeley, the California Policy Lab, CEGA, and CONICYT Chile. I also extend my thanks to the Ministries of Housing (MINVU), Education (MINEDUC), and Labor (MINTRAB) for providing data access.

I am truly grateful to the professors who supported me throughout this journey — David Card, Emmanuel Saez, Rucker Johnson, and Ricardo Perez-Truglia during my PhD, and Francisco Gallego and Jeanne Lafortune during my undergraduate and master’s years. Your guidance has shaped me in so many ways, and I’ll always be thankful for your mentorship.

I am forever grateful to the many Bay Area friends who have supported me and helped to make this time of my life so enjoyable. To all my housemates over the years—Nico, Pat, Alex, Justine, Nacha, and Vale—thank you for making this journey so much better. A special thanks to my five-year housemate and officemate, Javi— from shared walls to shared deadlines, thanks for being the best kind of officemate and housemate, one who made the PhD grind genuinely fun. And to Dirtybirds, thank you for the memories, laughs, and for introducing me to so many new experiences.

To my volleyball crew—Nico, Michi, Clau, Jessa, Justine, Pipe Arteaga, Alex, and Pat—thank you for the rallies on and off the court. For tennis, I’m especially grateful to Pipe Vial, Pipe Arteaga, and Harry for helping me rediscover the game. I also thank my Richmond tennis friends at RTA, who helped me survive the Job Market through tennis—especially Tomás, Manny, Sal, Nando, and Rob.

To my Chilean friends in LP, thank you for your unwavering support—no matter the distance. I always look forward to reuniting in Chile at the end of the year and picking up right where we left off.

Finally, to my family—Mom, Dad, Cata, Dani, Jose, Javi, Andrés, Cata, Mane, and Javier—thank you for your constant love and support. To Mom and Dad: thank you for always believing in me and encouraging me to follow my dreams. And to Cata and Dani: thank you for being there through it all, especially during the Job Market season—with Tomi bringing joy at just the right moments.

Contents

Contents	ii
List of Figures	iv
List of Tables	vi
1 The Effect of a Homebuyer Subsidy on Children	3
1.1 Introduction	3
1.2 Housing, the <i>DS01</i> Subsidy and Education in Chile	8
1.3 Data and Summary Statistics	12
1.4 Design	16
1.5 Results	19
1.6 Mechanisms	22
1.7 Early and Long-run Outcomes	27
1.8 Conclusion	31
2 The Effects of Widespread Online Education on Market Structure and Enrollment	53
2.1 Introduction	53
2.2 Setting and data	59
2.3 The Effects of Online Entry on Market Outcomes	67
2.4 Model	76
2.5 Counterfactuals	87
2.6 Conclusion and discussion	91
3 The Role of Housing Supply in Shaping Affordability: A Quasi-experimental Approach	93
3.1 Introduction	93
3.2 The Chilean Homebuyer Voucher	98
3.3 Data	102
3.4 Housing supply elasticity	103

3.5	Validation of the methodology and effects on housing prices: an Event Study approach	113
3.6	Conclusions	117
	Bibliography	119
A	Appendix of The Effect of a Homebuyer Subsidy on Children	130
A.1	Additional Figures	130
A.2	Additional Tables	146
B	Appendix of The Effects of Widespread Online Education on Market Structure and Enrollment	155
B.1	Additional Figures and Tables	155
B.2	Data Construction	164
B.3	Empirical Model Details	168
C	Appendix of The Role of Housing Supply in Shaping Affordability: A Quasi-experimental Approach	172
C.1	Additional Figures and Tables	173

List of Figures

1.1	Voucher Subsidy Amount and House Prices	33
1.2	First Stage	34
1.3	McCrary test after application	35
1.4	Regression Discontinuity and Event Study: Grades	36
1.5	RD plot for main achievement outcomes	37
1.6	Heterogeity by Age at Application: Grades	38
1.7	Heterogeity by Household Size: Grades	39
1.8	Heterogeity by Neighborhood of Origin: Grades	40
1.9	RD of parents' behavior	41
1.10	RD on hours of sleep	42
2.1	Expansion of online education	64
2.2	Expansion of online education enrollment by categories of field of study	65
2.3	Expansion of online education enrollment by age group	66
2.4	Expansion of online education enrollment by categories of field of study	70
2.5	Estimated parameters from demand and pricing decision	83
2.6	Changes in total value added by age group under each counterfactual	91
3.1	Growth in housing prices and stock across Santiago's CZ (2005-2020)	96
3.2	Voucher allocation over time	99
3.3	Prices of purchased properties by voucher type and price-subsidy function	101
3.4	Event Study on Local Prices	115
3.5	Distribution of Price Effects (2011-2019)	116
A1	Public spending, financial support and tax relief for homebuyers	130
A2	Subsidy Assistance by Income level	131
A3	Housing Tenure	132
A4	Quantitative Housing Deficit	132
A5	Qualitative Housing Deficit	133
A6	Evolution of Overcrowding by Income Quintile	134
A7	Distribution of Vouchers and Housing Prices for Recipients	134
A8	Cutoff Evolution by Call and Tier	135

A9	Score Distribution	136
A10	Distribution of average neighborhood quality	137
A11	Distance between houses and to the city center	137
A12	Maps of the distribution of winners and users	138
A13	Evolution of winners	139
A14	Reapplicants and Winners around the cutoff	139
A15	Heterogeity by Demographics: Grades	140
A16	Heterogeneity on Neighborhood: alternative measures	141
A17	Heterogeneities in high school attainment and years of college	142
A18	RD of children’s behavior regarding food consumption	143
A19	RD of children’s expectations	144
A20	RD of parents’ expectations about children’s education	145
B1	Value-added distribution for in-person and online degree programs	155
B2	Cost of building infrastructure as a function of distance	156
B3	Changes in various outcomes by age group under each counterfactual	157
B4	Estimated parameters from value added model	167
C1	Transactions, voucher recipients and neighborhood quality by CZ	176
C2	Homeownership and moving probability by income level	177
C3	Neighborhood choice by voucher recipients	178
C4	Evolution of prices	178
C5	Distribution of prices	179
C6	Histogram of Census Zone’s exposure to each voucher type	179
C7	Reduced forms and first stages scatter-plots	180
C8	Event Study on Local Prices with and without Pre-trends Adjustment	181
C9	Heterogeneous effects in an Event Study framework	182

List of Tables

1.1	Differences in Demographics at application	43
1.2	Differences in Outcomes at application	44
1.3	Achievement outcomes	45
1.4	Achievement outcomes: Robustness	46
1.5	Achievement outcomes by Gender	47
1.6	School Quality Outcomes	48
1.7	Labor Market outcomes for parents	49
1.8	Day-care and Preschool Attendance	50
1.9	High School and Post-Secondary outcomes	51
1.10	Labor Market Outcomes for Children Not in College	52
2.1	Comparison between online and in-person programs	67
2.2	Effects of introducing an additional online degree	74
2.3	Effects of introducing an additional online degree on enrollment by age groups	76
2.4	Elasticity and diversion ratios	81
2.5	Counterfactual exercises	88
2.6	The effects of the online education expansion	89
3.1	Vouchers eligibility criteria and structure from 2015 (in US dollars)	100
3.3	Share of new constructions and CBD distance – Low price properties	108
3.2	IV estimates for housing supply	109
3.4	Share of redevelopment and CBD distance	111
3.5	Implied effect on real prices (2011-2019)	112
A1	Score Components	146
A2	Neighborhood characteristics	147
A3	Optimal Bandwidths by Outcome	148
A4	Differences in demographics for voucher Users	149
A5	Achievement outcomes: Bandwidth selection	150
A6	Achievement outcomes by Age at application	151
A7	Achievement outcomes by Household size	152
A8	Achievement outcomes by Neighborhood quality	153

A9	High School and Post-secondary outcomes by age	154
B1	Areas of Study and Degrees	158
B2	Relationship between distance and probability of opening an online degree for selected years	159
B3	Demand parameters	160
B4	Diversion ratios by age group	161
B5	Marginal cost parameters	162
B6	Fixed cost parameters	163
B7	Value added parameters	166
B8	Effects of internet expansion on degree availability and enrollment	168
C1	OLS results for housing supply	173
C2	OLS results for other measures of prices	174
C3	Reduced forms and first stages	175

Dissertation Introduction

This dissertation examines the effects of public policies in housing and education, highlighting how targeted interventions shape educational outcomes, housing affordability, and market dynamics.

Chapters 1 and 3 analyze the impact of a homeownership subsidy in Chile, focusing on its effects on children's education, their labor trajectories, and the housing market. Chapter 2 investigates the equilibrium effects of online education expansion on higher education markets in Brazil, emphasizing its consequences for students. These policies provide a framework to explore key questions for policymakers: Can homeownership vouchers foster social mobility by improving children's education and labor prospects? What are the unintended consequences of such policies on the housing market? How does the expansion of online education affect student outcomes?

In Chapter 1, coauthored with Felipe Vial and Harrison Wheeler, we examine whether homeownership can enhance intergenerational mobility by fostering human capital accumulation in children. We study a large-scale homebuyer subsidy program for low-income families in Santiago, Chile, leveraging an arbitrary discontinuity in subsidy assignment to estimate its causal impacts on primary, secondary, and tertiary education outcomes. We find that children in subsidy-recipient families exhibit significant improvements in attendance, grades, class rankings, and achievement test scores compared to those in families that narrowly missed the cutoff. These gains are not driven by sorting into different schools or changes in neighborhood quality but are most pronounced in larger families, suggesting that alleviating overcrowded living conditions plays a key role. Children from winning households also report sleeping more hours on average, which is in line with improved housing conditions. Additionally, subsidy-winning parents reduce their labor force participation, engage more in their children's education, and express higher educational aspirations for them. The program also increases preschool attendance early in life and raises high school graduation, college attendance, and employment rates later on. Overall, the findings indicate that homebuyer subsidies can significantly improve the economic prospects of children from low-income households.

In Chapter 2, coauthored with Nano Barahona, Caue Dobbin, and Sebastián Otero, we analyze the rapid growth of Brazil's private online higher education sector and its effects on market structure and college enrollment. Exploiting variation in online education penetration

across regions and fields of study, we find that online programs increase enrollment among older students but divert younger students from higher-quality in-person programs. Increased competition reduces tuition costs for in-person programs but also leads to a decline in their availability. Using an equilibrium model of college education, we estimate that, without online education, the average student would experience a 3.4% higher value-added. While younger students benefit from fewer online options, older students are disadvantaged. Our results suggest that targeted policies restricting online education to older cohorts could improve overall student outcomes.

In Chapter 3, coauthored with Felipe Sepúlveda, we use the same homebuyer subsidy program to examine its effects on Santiago’s housing market. As housing affordability becomes a growing concern, many countries are implementing housing assistance programs to mitigate rising costs and expand access. A key factor in evaluating these programs is housing supply elasticity, which measures how responsive the market is to changes in demand. Using quasi-experimental variation from Chile’s 2011 homeownership subsidy program and detailed administrative data on property transactions from 1998 to 2020, we estimate Santiago’s housing supply elasticity and the program’s impact on property prices. Our findings indicate an elasticity of 0.543, with new construction responding at twice the rate of rental unit conversions to owner-occupied housing. This limited rental conversion constrains opportunities for households seeking homeownership. We estimate that the policy increased housing prices by an average of 8.9

Taken together, the chapters of this dissertation illustrate how housing and education policies shape economic and social outcomes through both direct and indirect channels. The findings emphasize the need to consider not only the immediate effects of these policies but also their broader market and neighborhood implications. By exploring the role of homeownership in intergenerational mobility, the market effects of online education expansion, and the housing price response to subsidies, this dissertation provides evidence on how policy design and market conditions interact to produce varied outcomes. These insights contribute to a deeper understanding of the trade-offs and unintended consequences of urban and education policies, offering valuable guidance for researchers and policymakers striving to enhance economic mobility and expand access to education and housing.

Chapter 1

The Effect of a Homebuyer Subsidy on Children

1.1 Introduction

Housing is scarce and expensive in most of the world. Rapid urbanization in low- and middle-income countries has outpaced the availability of quality housing in city centers ([UN-Habitat, 2022](#)), while in high-income countries, affordable housing remains a persistent challenge. Given the potential for housing stability to positively influence various family outcomes, there is significant interest in housing policy as a tool to expand housing access and improve economic prospects for vulnerable populations. In the United States, approximately \$67 billion, or 1% of total federal spending, is allocated to housing assistance each year, with most of it directed toward rental support.¹

Recent empirical studies have questioned whether such programs can actually deliver on their promise. For example, wages and employment of adults appear unaffected - or even decline - after receiving housing assistance across a variety of contexts ([Barnhardt et al., 2017](#); [Belchior et al., 2023](#); [Galiani et al., 2017](#); [Jacob and Ludwig, 2012](#); [Ludwig et al., 2013a](#); [Van Dijk, 2019](#)). However, much less is known about the impact of housing policies on children. Early investment in children's health and education have been some of the most impactful public programs to date ([Hendren and Sprung-Keyser, 2020](#)), and the environment that children grow up in is critical to their schooling and later-life earnings ([Chetty et al., 2016a](#); [Chyn, 2018](#); [Kling et al., 2007b](#); [Pollakowski et al., 2022](#)). Thus, it is conceivable that housing assistance delivered to families with children could have a large impact on children's long-run educational attainment and employment.

This paper provides new evidence on the causal effects of a unique housing policy—a homeownership subsidy for low-income families—on children's schooling and early-career employment outcomes. Our study is situated in Chile, a middle-income country and member

¹Source: [Peter G. Peterson Foundation](#) (2024).

of the OECD, where housing affordability issues persist despite substantial homeownership rates. We focus on families in Santiago, the capital of Chile—a rapidly growing metropolitan area comparable in size and economic significance to Madrid,² but with particularly strained housing affordability, making it an ideal setting for studying housing interventions.

In 2011, Chile introduced the *DS01* program, a major housing policy initiative that provides significant subsidies for low-income families to purchase homes in the private market. These subsidies cover up to about 85% of the purchase price of a home up to a maximum value of \$40,000, representing a wealth transfer that is equivalent to about five times the average recipient’s annual family income. Applicants need to pay the remaining portion of the home’s value using their savings. From 2011 to 2020, there were over 155,900 applications for the program, with 31,873 families awarded subsidies. Housing purchases of the subsidy recipients accounted for more than half of all home purchases by the poorest 20% of households in Chile from 2012 to 2017.³

A key contribution of this study is our access to novel administrative data from the Chilean Housing Ministry on applicants to the subsidy and the status of their application (whether they were awarded a subsidy and if they used it). In addition, we have information on the children of the applicant families that can be linked to schooling data from the Ministry of Education and (formal) labor market outcomes from the Ministry of Labor. The education data includes enrollment data for public daycare facilities, preschools, primary, secondary, and tertiary educational systems,⁴ as well as children’s grades, class rankings, absenteeism, test scores, high school completion, preschool attendance, and post-secondary outcomes. Furthermore, it includes responses to questionnaires on topics like food consumption, drug use, educational expectations, and parental involvement, completed by parents and children during the standardized SIMCE tests. We further link employer-employee data to capture early labor market outcomes, creating a comprehensive life-cycle view of children’s educational and economic trajectories following the subsidy.

We estimate the impact of the *DS01* policy on children’s outcomes by leveraging a discontinuity in the assignment mechanism that determines which households are awarded the subsidy. Applicants are scored based on family characteristics, and subsidies are awarded to the most vulnerable applicants up to a cutoff that varies annually with the budget and is unknown ex-ante. Awardees need to use the voucher in the private market to purchase a house, and they can choose any home that meets minimum habitability standards and complies with price restrictions. The voucher must be used in the region of application within

²Santiago has a population of 6.5MM and Madrid 6.8MM. Their GDP per capita is estimated to be USD 29,507 and USD 36,911, respectively. However, Santiago’s population density is almost twice as high (25.4 vs 15 residents per square mile). Source: World Bank and World Population Review.

³See [Figure A2](#).

⁴This cross-level tracking is a unique advantage of the Chilean administrative data structure. In the U.S., educational data is often available only at the district level, making it difficult to follow children whose families relocate.

three years,⁵ and applicants cannot sell or rent the house before five years.

Among households just above the cutoff point, we find that 62% of awarded applicants use the voucher to purchase a housing unit. This take-up rate is slightly higher than some housing programs in the US (Chetty et al., 2016a; Kling et al., 2005) or India (Barnhardt et al., 2017), but lower than the take-up rate for free housing measured in some programs in low-income countries (Agness and Getahun, 2024; Camacho et al., 2022). The high take-up rate can be explained by the monetary benefits of a homeownership subsidy in relation to temporary housing and the flexibility to choose the location of the unit. Imperfect compliance reflects the need to complement the voucher with savings or a mortgage in order to buy the house, as well as the limiting nature of the price cap on potential homes, which prevents applicants from selecting better neighborhoods in the city, and limits their housing choices. Empirically, we find that voucher users tend to relocate to neighborhoods that are comparable to their locations at the time of application but with slightly cheaper housing, higher population density, and lower average education levels. Moreover, over half of users relocate within one mile of their previous residence.

Our research design uses a standard fuzzy regression discontinuity framework that compares outcomes for families with a score just above the threshold to qualify for the subsidy and those with a score just below. Throughout, we report two sets of estimates: 1) an intention-to-treat (ITT) estimate based on the “reduced form” effect of scoring just above the threshold in a given round, and 2) an average treatment effect (ATE) estimate, which scales the reduced form effect by the fraction of marginal winners who buy a house, and captures the causal effect of actually using the subsidy to purchase a home on future children’s outcomes. Consistent with the validity of our approach, we find no evidence that applicants can manipulate their scores around the cutoff, or that families look or behave differently on either side of the cutoff prior to application.

We find that receiving a homebuyer subsidy leads to significant improvements in primary and secondary educational outcomes, with gains in both cognitive and non-cognitive skills. Children in subsidy-winning households experience increases of 0.27 standard deviation units in grades, a 7.9 percentage point rise in school rank, and a 0.26 standard deviation improvement in math and verbal test scores. They are more likely to be enrolled at school, be in a grade appropriate for their age, and are less likely to be chronically absent. They are also more confident in reaching their future goals as adults. These results are consistent across alternative specifications and are robust to various bandwidth choices, suggesting that composition effects, pre-existing differences in household characteristics, or unobserved confounding factors do not drive the positive impacts of the subsidy. Interestingly, the gains arising from winning the subsidy and buying a home are stronger for boys, and we do not find important differences depending on the age at the time of application.

We examine several mechanisms that may underlie these positive impacts, including parental responses and family resources, school quality, neighborhood environment, and

⁵This restriction is extendable.

housing quality. We find no evidence that changes in school quality or neighborhood environment are important in our context. We believe this is due to several factors particular to our setting: most households do not relocate far from their original home, those who do move tend to settle in neighborhoods with similar characteristics, and public schools in Chile allow children to attend schools outside their municipality of residence. These factors likely limit the influence of neighborhood and school quality on the observed outcomes. However, we find that gains in a child's academic performance are much stronger in larger families, suggesting that the subsidy may alleviate overcrowding and provide more favorable conditions for academic success. In addition, children from marginally winning households report sleeping more hours after being awarded the subsidy, supporting the hypothesis that improved housing conditions can explain the estimated effects.

We also find that family resources do not improve for marginal subsidy winners. In fact, a secondary contribution of our work is finding that parents' incomes and employment rates decline after winning the subsidy, consistent with results from earlier studies of other forms of housing assistance (Barnhardt et al., 2017; Kling et al., 2005; Van Dijk, 2019).⁶ Relatedly, improved household finances due to the subsidy do not translate into greater consumption for children, as measured by food. However, parents are 1.8 percentage points (3.5%) more likely to monitor their children's grades, 4.5 percentage points (5.5%) more likely to congratulate them for academic achievements, and 4.9 percentage points (10%) more likely to assist with study efforts after receiving the subsidy. Parents also expect their children to attend college and graduate school at higher rates.⁷ This increased involvement, alongside improved educational expectations, likely contributes to the observed gains in children's academic outcomes (Guryan et al., 2008; Boneva and Rauh, 2018).

Finally, we consider how early and later investments in human capital formation are impacted by the homebuyer subsidy. We find that children of subsidy users are 9.4 pp more likely to attend preschool. We lack enough time to observe the full extent of college behavior for our entire sample of children. However, for older children (ages 7-18 at the time of their families' application for the early application cohorts), we find that end-of-high school performance outcomes - including high school graduation, their average grades, and entrance exam scores for college - are significantly higher for children of marginally winning families relative to those from marginally losing families. Moreover, they attend college and receive professional degrees at higher rates. Conversely, they drop out of college at slightly elevated rates. For students who do not attend college, we look at early-career labor market outcomes and find suggestive evidence that the children of subsidy users are more likely to be employed, but we cannot yet detect increases in annual earnings.

This paper builds on a recent literature considering the effects of housing policies on individual-level outcomes. The effect of rent control (Diamond et al., 2019), eviction

⁶More generally, the characteristics of parents at the time of application seem to matter relatively little for the relative impacts of the subsidy.

⁷These findings are consistent with Kumar (2021), who show that parents are more optimistic about their children's future after receiving housing assistance.

([Collinson et al., 2023](#)), rental assistance ([Jacob and Ludwig, 2012](#)), and public housing ([Van Dijk, 2019](#)) on adult outcomes have all been studied. Less is known about how children are impacted by such policies. Studies of housing voucher programs have found mixed results on children’s educational attainment and earnings: positive effects in [Chetty et al. \(2016a\)](#) and [Pollakowski et al. \(2022\)](#); null effects in [Jacob and Ludwig \(2012\)](#)). Similarly, mixed findings have been found in work on forced moves from disadvantaged neighborhoods due to public housing demolitions ([Chyn, 2018](#); [Jacob, 2004](#)) and slum-clearance programs ([Rojas-Ampuero and Carrera, 2022](#)). On the other hand, programs that provide free housing in the low-income country context have been found to significantly improve children’s years of schooling ([Agness and Getahun, 2024](#); [Kumar, 2021](#); [Camacho et al., 2022](#)).

Our paper is the first to study a large-scale homebuyer subsidy, an initiative that promotes homeownership itself in contrast with rental assistance or public housing ([Kling et al., 2007b](#); [Chetty et al., 2016a](#); [Chyn, 2018](#); [Van Dijk, 2019](#); [Schwartz et al., 2020](#); [Pinto, 2022](#)). Significantly, the subsidy provides recipients with a voucher to purchase a home in the private market. In contrast, “free” or other public housing programs often require tenants to relocate to neighborhoods with potentially very different characteristics and schools ([Barnhardt et al., 2017](#); [Kumar, 2021](#); [Camacho et al., 2022](#); [Belchior et al., 2023](#); [Agness and Getahun, 2024](#)).

Homebuyer subsidies offer a scalable alternative to free housing programs, particularly in developed countries, and have become part of current policy discussions in the 2024 U.S. presidential election. Vice President Kamala Harris recently proposed a \$25,000 down-payment assistance program for first-time homebuyers.⁸ The proposal has garnered public support with 57% of Americans in favor according to a recent poll.⁹ To date, the only form of homeownership assistance in the U.S. has been a limited extension of the Housing Choice Voucher (HCV) program. While the broader HCV program provides \$18 billion in rental assistance to 2.2 million families annually ([Ellen, 2018](#)), the HCV homeownership program has just 9,673 active participants since 2015.¹⁰ The renewed focus on homebuyer subsidies reflects a growing recognition of housing stability’s role in family and child outcomes, an area our study explores in depth.

Our data enable us to follow children from daycare through high school and, for earlier cohorts, into college and the labor market. To our knowledge, this is the first study to examine the impacts of housing assistance across all levels of education: daycare, preschool, primary, secondary, and post-secondary schooling. Furthermore, we capture both cognitive

⁸This subsidy would be available to individuals who have consistently paid their rent on time for the past two years, with larger amounts for those whose parents did not own a home ([Campaign, 2024](#)). At a campaign event, Harris stated: “I know what homeownership means. It’s more than a financial transaction. It’s so much more than that. It’s more than a house ... It’s financial security. It represents what you will be able to do for your children.”

⁹Source: [YouGov poll](#).

¹⁰Source: [HCV Homeownership Dashboard](#). Figure current as of November 11, 2024.

and non-cognitive outcomes.¹¹ Non-cognitive outcomes (like attendance, grades, and aspirations) have often been overlooked in housing assistance research which typically centers on measures such as test scores and graduation rates.¹² For example, our work is the first to show that while parents reduce employment in response to housing assistance, they subsequently increase their involvement in supporting their child’s academic success. This research underscores the importance of examining both cognitive and non-cognitive skills, across every level of schooling, to fully understand the impact of housing assistance on educational and long-term life outcomes.

We conclude that there are significant benefits to homebuyer subsidies and, more generally, homeownership for low-income families. Studying the impacts on parents alone would overlook the main source of social gains: improvements to human capital formation in children. A back-of-the-envelope calculation extrapolating the gains in educational attainment finds that the homeownership subsidy increases a child’s present value of lifetime earnings by \$22,403 USD, more than covering the costs of the program. Our findings underscore the potential for homeownership subsidies to yield long-term benefits in children’s human capital development, positioning this policy as a promising tool for reducing intergenerational inequality.

The rest of the paper is organized as follows. Section 1.2 discusses the homebuyer subsidy and our Chilean context in more detail. Section 1.3 describes the sources of administrative data we combine for this project. Section 1.4 discusses our research design. Section 1.5 presents our main findings on K-12 educational outcomes. Section 1.6 discusses mechanisms. Section 1.7 presents evidence on early and later-life outcomes. Section 1.8 concludes.

1.2 Housing, the *DS01* Subsidy and Education in Chile

Housing in Chile

Chile has a long tradition of subsidizing housing for disadvantaged families.¹³ Until 2011, these policies primarily targeted the poorest 40% of the population by providing large subsidies to help them purchase homes. The introduction of the *DS01* homeownership program in 2011 marked a significant policy shift, expanding subsidies to include middle-income families and extending coverage to the poorest 60% of the population.

¹¹The Ministry of Education records grades and attendance as early as in first grade, providing a unique opportunity to track achievement across a child’s entire school history

¹²Some of the few exceptions include Kling et al. (2007b) (mental health), Jacob et al. (2015) (attendance), and Agness and Getahun (2024) (aspirations). Non-cognitive skills have proven to be critical predictors of high school completion, post-secondary enrollment, and earnings (Becker et al., 2010; Jackson, 2018; Carlana et al., 2022; Card et al., 2024)

¹³Policies dating back to 1906 were introduced to improve housing conditions for the poorest households.

In 2017, 60.4% of households owned the house they were living in, while only 21.9% of households rented their unit. Panel A of [Figure A3](#) illustrates homeownership rates by income quintile in Chile, showing a relatively stable ownership rate across the income distribution. The relatively equal distribution of homeownership in Chile may reflect the country's long-standing focus on housing subsidies for low-income families.¹⁴ This pattern has been stable during the last decades. By contrast, homeownership rates in the U.S. are much more skewed by income. In 2019, only 40% of the bottom income quartile owned their homes, compared to nearly 90% in the top quartile.

Despite high homeownership rates, Chile faces a significant housing deficit in both the quantity and quality of available units. In 2017, more than 500,000 housing units were needed to solve the quantitative deficit.¹⁵ This issue is particularly salient for the poorest 20%, who are 2.5 times more likely to face such a deficit than the richest 20% (Panel B of [Figure A4](#)). Concerning the qualitative deficit in housing units, Panel A of [Figure A5](#) shows that 1,250,000 units were needed to address overcrowding, poor sanitation, and substandard construction. Again, this issue is particularly pressing for the lower end of the income distribution. More than 35% of households in the bottom income quintile live in these units, while less than 10% do so in the top quintile (Panel B of [Figure A5](#)).

During the period 2012-17, 23% of all sold housing units were bought only using the *DS01* subsidy, and 10% were purchased using a combination of subsidy and credit. As shown in [Figure A2](#), there is a large amount of variation across the income distribution. More than 50% of the houses purchased by the bottom quintile were bought using either the subsidy or a combination of subsidy and credit. In contrast, only 18% of the households used this benefit in the top quintile. These figures highlight the critical role that housing subsidies play in enabling homeownership among Chile's lower-income families.

The *DS01* Policy

The *DS01* policy is a housing subsidy aimed at enabling homeownership for lower- to middle-income families in Chile. Through this subsidy, households apply for an up-front capital voucher to cover part of the cost of purchasing an existing housing unit in the private

¹⁴Source: Survey of Consumer Finances 2019, Federal Reserve

¹⁵Quantitative deficit: 70,000 houses were severely damaged, 300,000 houses had more than one household living in the same house, and 180,000 had more than one family in the same house, with more than 2.5 individuals per bedroom (Panel A of [Figure A4](#)). Overcrowding is defined as having more than 2.5 people per bedroom in the housing unit and is an important challenge that many families in Chile face. In particular, more than 6% of households live in overcrowded houses. As seen in [Figure A6](#), this is particularly important for the bottom quintile of the income distribution, as more than 12% of the households face this problem, compared to less than 3% in the top quintile. This situation differs markedly from the U.S., where overcrowding is defined less stringently as more than one person per room. Even by this more lenient measure, only 3.3% of U.S. households experienced overcrowding in 2019, compared to much higher rates of overcrowding among Chile's lower-income households (source: U.S. Census Bureau, American Community Survey, 2019).

market. Applicants can select any house that meets the minimum habitability standards and complies with price restrictions. Applicants must specify one of Chile’s 16 regions and select a preferred *comuna* (municipality) within it. However, this choice of *comuna* is not binding upon receipt of the subsidy.¹⁶

The subsidy is divided into three tiers, each targeting different income groups based on their vulnerability. The “high” subsidy tier is reserved for households within the poorest 60% of the population and requires minimum savings of \$1,000. This subsidy provides winners with a \$20,000 voucher for housing up to \$40,000. The “medium” and “low” subsidies are targeted at the 80% and 90% most vulnerable and require savings of \$1,400 and \$2,800, respectively. They provide approximately \$12,500 and \$8,000 vouchers – though varying in the home purchase price – for housing up to \$55,000 and \$80,000, respectively. Panel A of [Figure 1.1](#) shows the subsidy amount as a function of the house sale price separated by tier. The amount is fixed for the high subsidy as long as the price of the housing unit is greater than the subsidy amount, and it decreases in the value of the house for the medium and low subsidies, up to a minimum amount. In addition, the pricing cap for the high subsidy is more restrictive than the medium and low subsidies.

In Panel B of [Figure 1.1](#), we show the distribution of the subsidy amount as a percentage of the house’s value for the different tiers of the subsidy. On average, the voucher covers 85% of the house price in the high subsidy tier, compared to 55% in the medium tier and 18% in the low tier.¹⁷ Our analysis focuses on the high subsidy in Santiago, Chile’s largest region, which houses nearly half of the country’s population. We focus on the high subsidy tier because it aligns more closely with other studies on free or heavily subsidized housing, as it requires minimal upfront spending and no mortgage financing. It is also the most targeted program. We have a separate project that examines the effects of the medium and low subsidies, as well as the impact of the program across other regions of Chile. However, given the distinct nature of these subsidies and the differences in housing contexts outside Santiago, this paper concentrates on the high subsidy tier in Santiago.

After households submit their applications, a scoring system is used to determine eligibility. Each application receives a score based on various demographic factors, such as household size, single-parent status, presence of children or elderly members, and the household’s savings. In [Table A1](#), we show the components of the score and their impact on the overall score. Applications with the highest scores within each region and tier are selected to receive the voucher. The number of winning applications in each region depends on that year’s budget, making the cutoff unknown ex-ante. The voucher must be used in the application region within three years (which is extendable), and the housing unit cannot be sold or rented before five years.

[Figure A8](#) displays the cutoff scores by application call and tier. Changes in the pool of

¹⁶Chile has 346 *comunas*, 52 of which are located in Santiago.

¹⁷[Figure A7](#) presents the distribution of the voucher amounts and the value of the houses purchased for the recipients by tier.

applicants drive variance in cutoff scores across application calls, as the budget allocation across regions and subsidy tiers remains relatively stable over time. For instance, the large drop in the cutoff score in the second call of 2019 for the low and medium subsidies is due to a reduced number in the pool of applicants.¹⁸

Education in Chile

Since our analysis focuses extensively on educational outcomes, we begin with an overview of the Chilean educational system. Preschool education in Chile consists of three levels: *Sala Cuna* (SC, up to two years old), *Nivel Medio* (NM, ages 2 to 4), and *Nivel Transición* (NT, ages 5 to 6). Starting in 2015, the upper level of Transition education (Kindergarten) became mandatory. All the other levels remain non-mandatory.¹⁹ Primary education (*Educación Básica*) in Chile serves children ages 7 to 14 and covers grades 1 through 8.²⁰ As of 2020, primary education coverage was at 99.7% for children aged 7 to 14.

Secondary education (*Educación Media*), covering grades 9 through 12, is also mandatory. In 2020, secondary school attendance reached 87.7% for children aged 15 to 18.²¹ A distinctive feature of the Chilean education system is that students are not required to attend a public school within their municipality. In our sample, over 30% of students attend a school outside their municipality. This contrasts with findings from other studies where location is a key determinant of school achievement due to factors like access to amenities, housing stability, and differential access to schools. (Chetty et al., 2016a; Barnhardt et al., 2017; Camacho et al., 2022).

Tertiary education in Chile is divided into three types: universities, professional institutes (*Institutos Profesionales*), and technical training centers (*Centros de Formación Técnica*). Universities in Chile offer programs across a broad spectrum, including the humanities, social sciences, natural sciences, engineering, and medicine. Professional institutes provide shorter, career-oriented degrees in applied fields like business administration, accounting, and technical engineering. Technical training centers, on the other hand, offer short-cycle vocational programs, typically two years in length, to equip students with practical skills

¹⁸Figure A13 shows the full evolution of applicants and winners over different application calls.

¹⁹Preschool coverage is estimated at 33% for *Sala Cuna*, 49% for *Nivel Medio*, and 95% for *Nivel Transición*, with the higher coverage in the latter driven by the mandatory status of Kindergarten (Cite MINE-DUC).

²⁰Chilean schools are classified into three types based on funding: public, voucher, and private institutions. Public institutions are fundamentally funded by the government and are usually managed by local municipalities. Voucher schools are primarily privately financed by a combination of government funds and tuition paid by the students' families. Private institutions are financed exclusively through the tuition paid by students' families. Among students, 40% attended public institutions, 53% were enrolled in voucher schools, and 7% attended private schools.

²¹Of these students, 37% attended public institutions, 51% were in voucher schools, and 8% were enrolled in private schools.

in areas like electronics, mechanics, healthcare, and computer technology.²² Admission to traditional universities is primarily based on the PSU (*Prueba de Selección Universitaria*) entry exam.²³ In contrast, professional institutes and technical training centers generally do not require the PSU for admission, focusing primarily on secondary school completion.

1.3 Data and Summary Statistics

Our analysis is based on four primary data sources: subsidy records, labor market data for applicant families, educational records for applicants' children, and neighborhood characteristics. These data were accessed through agreements with the Chilean Ministries of Housing, Labor, and Education. We linked records across ministries using personal identifiers for both applicants and household members.

Subsidy Data

The subsidy dataset, provided by the Ministry of Housing, includes details on the application region, year, and specific program call. This information, combined with the vulnerability score, allows us to completely replicate the subsidy assignment process. This dataset also includes detailed applicant information, including address of origin, age, gender, nationality, disability status, marital status, and household composition (number of children, elderly members, and other adults). Additionally, the dataset contains financial information, such as savings, loans, and self-reported earnings. For households that utilize the subsidy, additional data include the date the subsidy was used, voucher amount, house value, and address of the purchased housing unit.²⁴

The subsidy dataset provides information on household structure at the time of application, including unique identifiers for each household member, enabling us to link these records to education and labor market data. The data also contains some basic demographic information like the family member's age, gender, and relationship to the primary applicant. Since household composition is recorded at the time of application, information on new household members is only available for applicants who reapply. To ensure compara-

²²In 2020, 52% of tertiary students were enrolled in universities, 29% in professional institutes, and 19% in technical training centers.

²³Traditional universities, which include both public and private institutions under the Council of Rectors (CRUCH), evaluate applicants based on PSU scores, secondary school grades, and class rank. More competitive programs, such as engineering, medicine, and law, often require higher PSU scores. Some private universities also require the PSU, though many place less emphasis on the score and offer alternative admission pathways.

²⁴Our main sample focuses on regular subsidy calls, excluding instances where all applicants were awarded the subsidy or where subsidies targeted households affected by natural disasters. If applicants appealed their initial score, we retained the higher score, as this score ultimately determined subsidy receipt.

bility, we exclude household members born after the application date, as this information is generally unavailable for most awarded households.

Education Data

Our educational data, provided by the Ministry of Education²⁵, covers the period from 2008 to 2023. This dataset is linked to all household members from the subsidy data, offering a comprehensive view of educational outcomes alongside household characteristics. The dataset includes information on daycare and preschool enrollment, primary and secondary school enrollment, grades, attendance, standardized test scores (SIMCE), as well as tertiary education entry test scores, enrollment, and graduation status. In addition, it has information on the school or college they are enrolled in, including whether it is a private, public, or voucher school and its location.

Our key academic performance measures include grades, dropout and repetition indicators, class rank percentiles, and average test scores, including math and verbal subject tests. In addition, we construct an indicator of whether they were in the grade they were supposed to be in according to their date of birth and an indicator of whether they were enrolled at school. We also have information on absenteeism, measured as the percentage of class days missed during the academic year. Additionally, we construct an indicator for chronic absenteeism, defined as missing more than 10% of class days.

Other outcomes include a school quality indicator based on the school's national ranking (via SIMCE scores), as well as indicators for attending a public or private school, priority student classification, and changes in school or comuna of residence. Priority student status is largely based on belonging to the most vulnerable 80% of households, as determined by the *Registro Social de Hogares*. Public schools receive additional government funding for each priority student they enroll.

During the SIMCE test, both students and their parents complete questionnaires addressing topics such as food consumption, drug use, educational expectations, and parental involvement. These questionnaires specifically ask about expectations for educational attainment, with options ranging from not completing high school to completing graduate studies. Using these responses, we measure changes in both children's and parents' expectations for the child's highest educational attainment, categorizing outcomes as high school, college degree, or graduate degree. We also consider whether children have improved expectations regarding their own grades and future goals. Lastly, we explore parental engagement by studying how likely they are to know their child's grades, help them study, and celebrate their academic achievements.

To analyze high school completion and post-secondary outcomes, we collapse the data to a single observation per student, restricting the sample to those who were in school at the time of application. Then, we generate indicators for whether they graduated from high

²⁵MINEDUC: Agencia de Calidad de la Educación (2024).

school, took the college entry exam, attended, dropped out, or graduated from college. In addition, we present the average grades in high school for the graduates and the standardized scores in math, verbal, history, and science for the ones taking the entry exam.

For college dropouts and graduates, we limit the sample to students who enrolled in a post-secondary program. For tertiary attendance, we restrict the sample to students who graduated from high school. However, our results do not change if we use the full sample. Finally, for students who enroll in post-secondary programs, we calculate the number of years attended and create indicators for institutional quality.²⁶

Labor Market Data

We obtained an employer-employee dataset from the Ministry of Labor that provides detailed wage information and payment dates for each applicant with unemployment insurance coverage from January 2002 to December 2023, covering all formal employment in Chile. The dataset includes information on monthly earnings received from each employer, contract type (permanent or temporary), work location, and workers' educational attainment. Using these data, we construct yearly variables to measure employment status, total months worked, and annual earnings. We further disaggregate earnings and months worked by contract type, allowing us to differentiate between permanent and temporary employment.

Because the dataset is sourced from unemployment insurance records, it excludes self-employed individuals, independent contractors, civil servants, and informal sector workers. However, approximately 90% of applicants in our sample appear in the dataset at least once, indicating that most have engaged in the formal labor sector at some point.

In our main analysis, we include working-age individuals, defined as individuals between 18 and 65 years old, and impute zero earnings for any year in which earnings are not reported. We then restrict the sample to the working-age population, defined as individuals between 18 and 65 years old. Our results are robust to different sample selections, including: (1) excluding individuals under 25 until their first appearance in the dataset, (2) excluding each person until their initial entry in the dataset, and (3) excluding each person from both their first appearance and after their last appearance in the dataset. We further analyze the subsidy's effects conditional on employment, examining both earnings and months worked. To handle years with no employment, we use Poisson regressions for annual earnings and months worked. A limitation of the dataset is the absence of hours worked, restricting our analysis of the intensive margin to the number of months in which earnings are reported.

Neighborhood Data

To examine the significance of location in our analysis, we gather neighborhood information from three primary sources to understand applicants' origins and destinations. First, the

²⁶We consider whether the student attended a university (versus a technical institution) and the number of years the institution was accredited by the Ministry of Education.

2017 Census data provides essential indices such as education levels, overcrowding, housing material quality, housing deficits, population density, and socioeconomic characteristics. Second, data from *Espacio Público* (2017),²⁷ offers information on neighborhood amenities, including green spaces, educational facilities, supermarkets, public transit stops, and health-care centers. Third, data from the *Servicio de Impuestos Internos* (Chilean IRS) provides information on housing unit and square foot valuations and land sizes.

After extensive data cleaning, we geolocated both the origin and final addresses of applicants, successfully matching 98.5% of origin addresses and 93.2% of final addresses.²⁸ We then linked each address to its corresponding census tract. In Chile, there are 46,087 census tracts across 346 municipalities, with Santiago containing 3,426 tracts within 52 municipalities.²⁹

Summary Statistics

For our main sample, we keep applications that we are able to link to at least one household member in the education data. This matching process yields 131,251 applications for the high subsidy in Santiago, representing 84.2% of the full sample. Restricting to applicants with children under 18 at the time of application reduces the sample to 90,362 applications. The first four columns of Table A4 present summary statistics for all applicants, winners, subsidy users, and non-users, respectively. Column (5) shows the differences in demographic characteristics between users and non-users, calculated by taking the difference between columns (3) and (4). As shown in column (1), most applicants linked to the education data are female and in their mid-to-late thirties, with about two-thirds being single parents. The average household comprises four members, including 1.5 children attending school. They have applied 1.2 times before, on average, and the average number of years of schooling in their neighborhood of residence is ten.

Column (2) shows that winners' demographic characteristics are largely similar to the rest of the sample, with notable differences in household size, number of school-aged children, and normalized score. This is not surprising considering that the number of household members, and especially the number of children in the household, is a crucial component of the score. Column (5) indicates that among winners, subsidy users are, on average, two years younger and have fewer children in school. They also have lower scores and have applied fewer times on average. Finally, they report a lower income and tend to live in slightly worse neighborhoods than non-users in terms of average years of schooling. Simple comparisons

²⁷Espacio Público employs a geographic division called *Unidad Vecinal*, which slightly differs from census tracts. We linked our locations to these units to maintain consistency.

²⁸The difference in geolocation success rates is primarily due to data quality; origin addresses were consistently structured by region, *comuna*, street, and number, whereas final addresses were often combined into a single cell with varied formats across applications.

²⁹To enhance matching accuracy, we created a crosswalk to correctly align neighborhoods with addresses that referenced specific buildings, villas, or condominiums.

between winning and losing households—or between users and non-users—would be biased by these demographic differences. Our research design relies on comparisons between winning and losing applicants at the cutoff, where the two groups are statistically indistinguishable on a host of measures.

Column (1) of [Table 1.7](#) provides summary statistics on the labor participation of the children’s parents. On average, 46% of parents are employed in a given year. Including non-working parents as having zero employment, the average number of months worked is 4, with 2.7 months under permanent contracts and 1.3 months under temporary contracts. In terms of earnings, parents average 63.2 UF per year, or approximately \$3,384 USD, again with non-working parents counted as zero earners. Around 75% of these earnings are from permanent contracts, while the remaining 25% come from temporary contracts.

1.4 Design

Our research design relies on comparing the children of applicants who barely lost the subsidy to those of applicants who barely won. We focus on two key parameters: first, a reduced-form estimate of the effect of winning the subsidy (ITT); and second, an adjusted estimate that accounts for the likelihood of subsidy usage among winners compared to those who barely missed receiving it. We also detail our approach to reapplications, describe our estimation strategy, and present evidence on the validity of our regression discontinuity (RD) design.

ITT Estimates

We first estimate the effect of winning the subsidy on a child’s human capital formation without accounting for whether the family ultimately used the subsidy. To capture this intent-to-treat (ITT) effect, we model the outcome using the following specification in [Equation 1.1](#):

$$y_{itc} = \alpha_0 + \alpha_1 f(\text{Score}_{ic}) + \beta_{ITT} D_{ic} + \alpha_2 D_{ic} \cdot f(\text{Score}_{ic}) + \Pi X_{ict} + \varepsilon_{itc} \quad (1.1)$$

The y_{itc} is the outcome of interest of the applicant i , at time t , and call c . The Score_{ic} represents the applicant score in call c . The X_{ict} controls flexibly for a set of application characteristics, including gender by age of the applicant and year fixed effects. The D_{ic} is a dummy variable that takes the value of 1 if the applicant’s score is greater than 0, indicating that they won the subsidy. The $\alpha_1 f(\cdot) + \alpha_2 D_{ic} f(\cdot)$ is a flexible function capturing the relationship between the application score and the outcome. We allow for different parameters on either side of the cutoff. Our coefficient of interest, β_{ITT} , estimates the effect of winning the subsidy on the outcome for applicants at the cutoff score.

To implement our estimation, we need to take a stand on the functional form of $f(\cdot)$, the weights applied to data near and far from the cutoff, as well as the bandwidth to use. In our main results, $f(\cdot)$ is modeled using a linear polynomial with a triangular kernel. We also

explore alternative functional forms and weights in robustness checks. We select the optimal distance from the cutoff using the bandwidth selection procedure proposed by [Calonico et al. \(2014\)](#) for each outcome at the time of the application. Throughout the analysis, we cluster standard errors at the applicant level.

LATE Estimates

Given the application process, many unsuccessful applicants reapply in subsequent years, particularly those near the cutoff. In particular, approximately 70% of applicants just below the cutoff reapply, and around 60% of these eventually win the subsidy ([Figure A14](#)). Therefore, comparisons between winners and losers around the cutoff may include losing applicants who later win, potentially benefiting from the subsidy as their children experience improved housing. This dynamic likely causes the ITT estimate to understate the subsidy's effect on children's educational attainment.

To address this, we model whether an application ultimately uses the subsidy around the cutoff. We rely on variation in subsidy-using rates driven by the discontinuous jump in the probability of winning the subsidy at the cutoff. This first step is captured in [Equation 3.6](#). We then use predictions from that model to capture how educational outcomes of interest vary with how likely the family is to use the subsidy around the cutoff in [Equation 1.3](#).

$$Used_{itc} = \gamma_0 + \gamma_1 f(Score_{ic}) + \beta D_{ic} + \gamma_2 D_{ic} \cdot f(Score_{ic}) + \Gamma X_{ict} + u_{itc} \quad (1.2)$$

$$y_{itc} = \delta_0 + \delta_1 f(Score_{ic}) + \beta_{LATE} \hat{Used}_{itc} + \delta_2 D_{ic} \cdot f(Score_{ic}) + \Delta X_{ict} + v_{itc} \quad (1.3)$$

Here, $Used_{itc}$ is a binary variable indicating whether applicant i in call c used the subsidy to purchase a house by time t . The remaining terms are defined as in the previous section. In [Equation 1.3](#), β_{LATE} captures the Local Average Treatment Effect (LATE), representing the causal effect of using the subsidy on the outcome for applicants at the cutoff. In practice, we estimate this set of equations through a two-stage least squares (2SLS) regression. The distance from the cutoff we impose to limit our estimation sample is computed using the optimal bandwidth proposed by [Calonico et al. \(2014\)](#). Standard errors are clustered at the applicant level.

Identification of β_{LATE} relies on the discontinuous jump in the winning probability at the cutoff. This increased probability drives higher subsidy usage rates among applicants who barely win compared to those who barely lose. A causal interpretation of β_{LATE} requires that all other factors influencing children's educational outcomes remain continuous around the cutoff ([Lee and Lemieux, 2010](#); [Cattaneo et al., 2020b](#)). As the score is used exclusively for the *DS01* subsidy, no other policy changes occur at the discontinuity. Moreover, since the cutoff depends on the application call demand and is unknown before the application, an applicant can't manipulate their scores to cross the threshold. In the following section, we provide evidence that scores were in fact not manipulated, and that other characteristics and pre-application outcomes remain smooth around the cutoff.

Households receive a bonus score each time they have applied unsuccessfully in the past. In Section 1.4, we show that while this bonus increases applicants' scores when reapplying, there is no discontinuity in the number of past applications at the threshold. In other words, winners and losers near the cutoff do not statistically differ in the number of prior applications. Further, in Section 1.5, we demonstrate that our main estimates remain robust when controlling for the number of previous applications.

Panel A of Figure 1.2 reports the first stage for the high subsidy in Santiago, using Equation 3.6 and pooling across application calls. We find that around 62% of winning applicants eventually use the voucher to purchase a home, compared to 38% of losing applicants, resulting in a 23 percentage point increase at the threshold. This jump indicates a sufficiently strong first stage for estimating the policy's effects. Moreover, this result suggests that the ITT estimates need to be rescaled by a factor of approximately 4 to 5 to calculate the LATE of using the voucher to purchase a home. In Panel B of Figure 1.2, we show how take-up rates for winning and losing applicants evolve over time. Take-up gradually increases for both groups, with a relatively stable 25 percentage point difference on average.

Take-up for the *DS01* subsidy is higher than in some housing programs (Chetty et al., 2016a; Barnhardt et al., 2017; Kling et al., 2005) but lower than programs with nearly complete take-up (Agness and Getahun, 2024; Camacho et al., 2022). In our setting, monetary benefits and increased housing stability and quality are likely drivers of the take-up rates. However, requirements for savings and housing price caps, which limit neighborhood options, may deter some winning families from using the subsidy.

Design Validity

To assess the validity of our design, we test for potential manipulation of application scores around the cutoff. Specifically, we examine whether there is a discontinuity in the distribution of applicants' scores, which would indicate strategic behavior. Additionally, we assess the continuity of observable characteristics and prior outcomes for both applicants and their children near the cutoff, ensuring comparability between winners and losers.

Figure 1.3 presents the results of a score manipulation test for the high subsidy, using a local polynomial fit to the empirical distribution of the score on either side of the cutoff. This test evaluates whether there is a structural break in the score distribution at the cutoff (Cattaneo et al., 2020a). While maintaining the flavor of the McCrary test (Lee and Lemieux, 2010; McCrary, 2008), this approach does not require pre-binning the scores into a histogram. With a p -value of 0.966, we cannot reject that scores are not manipulated. This is unsurprising, given the varying nature of the cutoffs across application calls.

Another key assumption in the RD design is that observable characteristics (predetermined variables) should remain continuous across the cutoff, ensuring that observed outcome differences are attributable to subsidy receipt rather than preexisting differences between winners and losers. In Table 1.1, we apply our ITT specification from Equation 1.1 to demographic characteristics at the time of application and find no significant discontinuities

in any demographic variables. Notably, there is no significant difference in the number of prior applications at the cutoff; both winners and losers averaged 1.2 previous applications. These findings are robust to different bandwidth selections.

Finally, we test for “pretrends,” or differences in the outcomes of interest prior to application, by estimating models of the form in [Equation 1.1](#) on outcomes in the year before the program call. Panel A of [Table 1.2](#) shows that, while children in subsidy-winning households performed slightly better at the time of application, these differences are statistically insignificant and economically small relative to the control mean. Overall, our findings confirm that the design is valid and that subsequent results are not driven by prior differences in children’s academic performance.

In addition to checking for differences at the time of application, [Section 1.5](#) considers whether winning and losing applicants followed different trends prior to applying. To do this, we estimate versions of our main specification on outcomes at each time period prior to application and find no evidence for pretrends on our main education outcomes.

Children at the Cutoff

Since our RD design focuses on comparisons near the cutoff, our estimates of the homebuyer subsidy’s effects are most relevant to children in this range. We briefly characterize the sample of children for this subgroup in [Table 1.2](#). The average student grade is 5.7 out of 7, with an average class rank in the 49th percentile. Approximately 17% of students are not in the expected grade based on their birth date, indicating prior grade retention, with 4.4% retained in a given year. The annual dropout rate for the control group is 0.5%, and the cumulative dropout rate averages around 4.4%. Absenteeism is about 10% of school days, with chronic absenteeism—defined as missing more than 10% of days—reaching 36%.

Panel B shows that children of winning and losing applicants generally attend schools with comparable characteristics. About 34% attend public schools, while only 0.4% are enrolled in private institutions. The remaining 65% attend voucher schools. The average class size is 36 students, and the average school quality, measured by percentile rank in test scores, is at the 44th percentile. Additionally, 62% of students are classified as priority students, enabling public and voucher schools to receive additional funding. Finally, around 12% of the sample switched schools in a given year, and 2.6% reported changing their municipality of residence.

1.5 Results

Main Results: Primary and Secondary Education Achievement

We begin by examining the effect of the homebuyer subsidy on K-12 academic performance. Panel A of [Figure 1.4](#) presents regression discontinuity results for grades after the application,

using a linear polynomial with triangular kernel weights and pooling across years and calls. Grades are negatively correlated with the vulnerability score on both sides of the threshold. However, we find a 0.035-point increase at the threshold, equivalent to an improvement of approximately 0.06 standard deviations.

In Panel B, we use the panel structure of our data to trace the evolution of grade differences before and after the application. Before the application, there are no significant grade differences between winners and losers. After the application, however, we observe a significant and sustained improvement in grades for children in winning households, starting in the year following the award. This grade increase shows no substantial dynamics over time, remaining relatively constant after being awarded the voucher.

Table 1.3 displays the regression discontinuity results for additional achievement outcomes, with both ITT and ATE coefficients estimated using equations 1.1 and 1.3, respectively. Each cell reports an individual RD estimate calculated using the optimal bandwidth proposed by Calonico et al. (2014) and fitting a linear polynomial on each side of the cutoff, with standard errors clustered at the applicant level. The results reveal significant positive effects across most outcomes. The ATE estimates suggest a 0.27 standard deviation increase in grades, a 0.08 percentage point rise in class rank percentile, and a 0.32 standard deviation improvement in test scores. Moreover, we observe a 2% increase in school attendance, a 9% increase in the probability of students being in the correct grade for their age, and reductions in both grade repetition and dropout rates. Although there is no significant effect on the overall percentage of school days missed, we find a 4.1 percentage point decrease in chronic absenteeism. Figure 1.5 provides graphical evidence of the discontinuities across some of these achievement outcomes.

These results are notably larger than those found in Jacob (2004), which investigates the educational impact of housing assistance following the Chicago public housing demolitions—one of the few studies to examine intermediate secondary school outcomes, particularly non-cognitive skills. The authors found no significant effects on GPA, standardized test scores, absenteeism, dropout rates, or age-grade mismatch. Similarly, our results are more pronounced than the effects on standardized test scores reported in Schwartz et al. (2020) for housing vouchers in New York City.

Most of the related literature assessing secondary education outcomes uses broader metrics, such as school enrollment, high school completion, years of schooling, and exit test scores. However, Jackson (2018) highlights that non-cognitive indicators like attendance, course grades, and grade repetition can differ from test score outcomes and have meaningful long-term impacts on student success. In Section 1.7, we benchmark our findings against this literature by presenting results for high school graduation, university entry scores, and post-secondary outcomes.

In Figure A19, we show that children also report higher academic expectations following the subsidy award. There is a positive but insignificant increase in their belief that they will achieve good grades and attend post-secondary education. Moreover, children in awarded households are 3.1 percentage points (significant at the 1%-level) more likely to be confident

in reaching their future goals as adults (Panel C). These results align with previous research showing that housing assistance improves children’s expectations regarding their academic success (Agness and Getahun, 2024).

Robustness

Table 1.4 demonstrates that our results are robust across various model specifications. In Columns (3) and (4), we show that the estimates remain consistent regardless of whether we exclude controls or include controls for the outcome at the time of application. Column (5) confirms that the results are not sensitive to bandwidth selection, as using the average bandwidth yields similar estimates.

Further robustness checks in Columns (6) and (7) indicate that the results hold when excluding weights or using a second-degree polynomial instead of a linear polynomial. Finally, Column (8) shows that our estimates are, if anything, stronger when excluding the COVID-19 period (2020 onward) despite the sample size being reduced by half. In Table A5, we also demonstrate that our results remain consistent across a range of bandwidths from 30 to 100, in intervals of ten.

Heterogeneities

Table 1.5 shows that the impact of housing assistance is larger for boys than for girls across all main academic achievement outcomes. Specifically, boys from awarded households improve their grades by 0.34 standard deviations and increase their class rank by 10 percentage points, whereas girls experience smaller gains of 0.20 standard deviations in grades and 6 percentage points in class rank. Similarly, boys’ test scores increase by 0.5 standard deviations, and they are 10.2 percentage points (12.5%) more likely to be in the correct grade for their age, while the effects on these outcomes for girls are not statistically significant.

These findings contrast with results from the Moving to Opportunity (MTO) experiment (Chetty et al., 2016a; Kling et al., 2007b) and other housing subsidy programs (Pollakowski et al., 2022), where girls experienced larger educational gains than boys. Prior studies suggest that girls adapt more readily to new neighborhood norms, engage in less risky behavior, and face fewer challenges with relocation compared to boys (Clampet-Lundquist et al., 2011). In our setting, however, most awarded households do not move far from their original location, and relocations rarely involve significant changes in neighborhood quality. Additionally, children are no more likely to switch schools after receiving the subsidy. Therefore, unlike in the MTO studies, our setting does not expose children to substantial changes in neighborhood or school characteristics. We provide further evidence of this in Section 1.6. Our findings regarding heterogeneity in gender align more closely with studies on housing relocation in Chicago and Australia (Deutscher, 2020; Jacob, 2004; Jacob et al., 2015).

The age of the child at the time of receiving housing assistance has also proven to be important in other settings. Longer exposure to improved neighborhood and school environ-

ments during critical developmental stages can be beneficial, while older children may face greater challenges adjusting socially and academically after relocation. However, as shown in [Figure 1.6](#), age at application does not appear to be a significant source of heterogeneity in our setting. The ITT effect on grades remains stable across age groups, ranging between 0.03 and 0.05, corresponding to a 0.2–0.4 standard deviation increase. If anything, older children at application perform slightly better academically, though this difference is not significant. Results for other achievement measures are presented in [Table A6](#).

Our findings differ from studies that report larger benefits for younger children receiving housing assistance ([Agness and Getahun, 2024](#); [Chetty et al., 2016a](#); [Gennetian et al., 2012](#); [Jacob et al., 2015](#); [Kling et al., 2007b](#); [Schwartz et al., 2020](#)). However, most of these studies look at differences in high school completion or earnings in a particular year after the move, finding that the effects are driven by a larger dosage of childhood exposure to improved neighborhood environments. In our setting, families tend to remain close to their original location and experience little improvement in neighborhood conditions on average. Moreover, by tracking achievement outcomes yearly after the application, we capture intermediate cognitive and non-cognitive outcomes, adding insight into achievement dynamics over time.

Our results align with those of [Deutscher \(2020\)](#), who found that relocation effects are stronger during late childhood or teenage years. In [Section 1.7](#), we present findings on high school graduation, which may be more comparable with other studies. Regardless, we believe that these differences can partially be explained by modest changes in neighborhood and school characteristics after the move.

Finally, we find no evidence of significant heterogeneity based on other applicant characteristics, including marital status, gender, and self-reported income, as shown in [Figure A15](#).

1.6 Mechanisms

Neighborhood

Neighborhood quality plays a critical role in children’s educational outcomes by shaping access to resources, ensuring safety, and influencing peer networks. Studies consistently demonstrate that exposure to neighborhoods with higher socioeconomic status can improve long-term outcomes by providing better access to schools, libraries, and safe recreational spaces that promote a positive learning environment ([Chetty et al., 2016a](#)). Low-crime, socioeconomically stable neighborhoods also contribute to reduced stress and better mental well-being, which support children’s academic focus and resilience ([Cutler and Glaeser, 1997](#)). Social exposure to peers from higher socioeconomic backgrounds within these neighborhoods can further elevate academic aspirations, setting higher educational benchmarks and expectations ([Sacerdote, 2001](#)). Together, these studies find that neighborhood quality can be an influential factor in improving educational opportunities and outcomes, particularly for children from disadvantaged backgrounds.

To characterize the initial and final neighborhoods of subsidy users, [Table A2](#) displays neighborhood characteristics at the census tract level. Columns (1) and (2) show average neighborhood quality in the initial and final locations, respectively. Column (3) reports the difference in neighborhood quality, and column (4) shows the number of applicants. Panel A shows that users tend to relocate to neighborhoods with slightly lower average education levels and higher population density. However, these areas also have less overcrowding, lower housing deficits, and better housing materials, indicating trade-offs in neighborhood quality. Panel B examines amenities per capita, revealing that users' new neighborhoods generally offer fewer amenities, including educational and healthcare facilities, supermarkets, bus stops, and drugstores. Panel C shows that, on average, they move to slightly cheaper neighborhoods.

However, households that use the subsidy in our setting generally remain close to their original location and experience only modest changes in neighborhood quality. [Figure A10](#) displays the distribution of neighborhood quality based on residents' average years of education. Panel A shows that differences between initial and final neighborhoods are minimal. Although users generally move to slightly less educated areas, almost 50% experience no change in neighborhood quality, and most changes are close to zero, suggesting very small shifts in neighborhood quality. Additionally, [Figure A11](#) displays the distribution of distances from users' initial and final addresses to the city center, as well as the distance between these two locations. Over half of users relocate within one mile of their previous residence, with new homes generally situated slightly farther from the city center. These findings suggest that applicants do not move far from their original neighborhoods, and when they do, they tend to relocate to areas with similar characteristics.³⁰

We formally consider whether variation in neighborhood quality impacts the gains we find for children. Our data include initial locations for all applicants and final locations for households that received and used the voucher subsidy. Unfortunately, final location data are unavailable for non-awarded households.³¹ Consequently, our neighborhood analysis focuses on heterogeneity by neighborhood quality at the time of application. We classify neighborhoods into terciles of quality (bad, medium, good) based on average years of education within each census tract.³² In [Figure 1.8](#), we present ITT estimates for grades by neighborhood

³⁰[Figure A12](#) presents the distribution of census tracts for voucher-awarded applicants at the time of application (Panel A) and for voucher users (Panel B). It also shows the distribution of users' final locations (Panel C) and the average years of schooling by census tract (Panel D). These maps reveal two key insights: first, the geographic distribution of voucher winners closely resembles that of voucher users, indicating minimal selection differences between the groups. Second, while users' final locations tend to be more concentrated in some peripheral neighborhoods, these areas exhibit similar average years of schooling as their initial locations. This visual evidence aligns with previous findings, suggesting that users experience only minor changes in neighborhood quality upon moving.

³¹The Housing Ministry does not collect final location data for non-awarded applicants, and our other data sources do not provide location granularity sufficient to track neighborhood changes. However, we applied for data access from the Ministry of Social Development, which includes locations for all families.

³²When applying, applicants specify a preferred municipality (*comuna*), though this preference is non-

type of origin, revealing minimal variation in effects across neighborhood types. Children from higher-quality neighborhoods show slightly smaller gains, with effects no longer statistically significant at the upper end. [Table A8](#) presents results for other outcomes, with the strongest effects observed among households originating from “medium” human capital neighborhoods. Results remain consistent across alternative measures of neighborhood quality, as shown in [Figure A16](#).

Our findings diverge from the broader literature emphasizing the role of neighborhood quality in educational attainment ([Chetty et al., 2016a](#); [Deutscher, 2020](#); [Laliberté, 2021](#); [Schwartz et al., 2020](#); [Van Dijk, 2019](#)). These studies often examine policies that facilitate moves to substantially better neighborhoods and schools. In our context, most households do not relocate to significantly improved neighborhoods or switch schools.³³ Our results align with those from [Jacob et al. \(2015\)](#) and [Haltiwanger et al. \(2020\)](#), which found limited educational impacts from neighborhood quality alone, and from [Haltiwanger et al. \(2020\)](#), which reported no significant effects for families moving to lower-quality neighborhoods.

School Quality

School quality is a key predictor of a child’s human capital formation. Access to high-quality schools enhances teacher quality, curriculum rigor, available resources, and peer influences, all of which improve high school completion rates and college enrollment, especially for disadvantaged children. Consequently, school quality has proven to be an important mediator in studies of the MTO experiment and other housing mobility programs ([Chetty et al., 2016a](#); [Barnhardt et al., 2017](#); [Camacho et al., 2022](#)). However, in our context, families tend to remain close to their original residences, and Chile’s public school system does not require students to live in the same municipality as their school. Consequently, changes in school quality are less likely to be a relevant factor in our setting.

To assess this directly, we examine changes in school quality using school characteristics as outcomes in our main design. These results are presented in [Table 1.6](#). We find no evidence that children from subsidy-winning households sort into different schools following relocation. Specifically, they are equally likely to attend public or private schools, with no significant changes in schools’ average test scores or class sizes. Furthermore, we observe no increase

binding when choosing a home. We explore differential effects at the *comuna* level, similar to [Van Dijk \(2019\)](#). However, too few households apply to municipalities outside their residence.

³³In related research, we examine how neighborhood quality influences outcomes for recipients of the medium and low subsidy tiers. Our findings in that context differ from those in this paper, as neighborhood quality appears to play a significant role. Specifically, the null overall effects we observe seem to be driven by positive outcomes for children moving from low-quality neighborhoods and negative outcomes for those moving from high-quality neighborhoods. This pattern likely reflects the greater mobility in neighborhood characteristics among medium and low subsidy users, which we attribute to less restrictive price caps, allowing for more varied relocation choices and potentially explaining the contrasting findings.

in school switching rates among these students; if anything, they are slightly less likely to switch schools post-application, though this difference is not statistically significant.³⁴

We attribute these findings to the limited relocation distance in our setting and the flexibility of Chile’s public school system, which allows students to attend schools outside their municipality of residence. This feature may also explain the absence of differential effects by neighborhood type. Our results align with studies that allow for more flexibility in housing location choices, where changes in neighborhood or school quality may be less pronounced (Agness and Getahun, 2024; Kumar, 2021).

Housing Conditions

Given the limited effects of neighborhood and school quality, housing conditions emerge as a potential mechanism influencing children’s educational outcomes. Poor housing quality—characterized by overcrowding, limited space, and inadequate facilities—can increase stress, reduce focus, and lead to higher rates of illness, all of which negatively impact school performance. Overcrowded conditions, in particular, restrict access to quiet, private spaces necessary for studying and rest. Research shows that children in improved housing environments experience academic benefits due to reduced stress and better sleep quality (Goux and Maurin, 2005; Lavy et al., 2012).

To investigate housing conditions as a potential mechanism, especially overcrowding, we analyze the effect of the homebuyer subsidy on standardized grades by household size at the time of application. These results are reported in Figure 1.7. The findings indicate that the effect of the voucher increases with household size, with children from larger households benefiting more significantly. Specifically, children in households with three or fewer members show a modest, statistically insignificant improvement of 0.034 standard deviations. In contrast, children from households with more than four members experience a gain of over 0.08 standard deviations, more than doubling the effect seen in smaller households. For larger households, the ATE reaches 0.32 standard deviations. Table A7 confirms that these patterns hold across our other main achievement outcomes.

We interpret household size at application as an effective proxy for overcrowding. Approximately 9% of households in the lowest 40% of the income distribution live in overcrowded conditions, defined as more than 2.5 people per bedroom (Figure A6). These findings indicate that improved housing conditions—particularly increased living space—are likely a key mechanism driving the positive impact of housing assistance on children’s educational outcomes. Our results align with studies in developing countries where housing renovations are associated with substantial gains in children’s academic performance (Cattaneo et al., 2009; Galiani et al., 2017; Kumar, 2021).

³⁴Approximately 12% of children in our sample switch schools annually, excluding transitions between elementary, middle, and high school.

In addition, using children’s answers from the SIMCE survey, in [Figure 1.10](#) we show that children in marginally winning households did not report sleeping a significantly different number of hours before the application to the subsidy. However, after being awarded the subsidy, the fraction that reports sleeping more than 7 hours a night is 0.034 percentage points higher for marginal winners, which corresponds to a 5% increase. We take these estimates as supporting evidence that improved housing conditions might be contributing to the gains in children’s educational outcomes.

Wealth Effect and Parental Responses

When families receive housing assistance, parents may adjust their labor market behavior due to changes in household stability, income, and time allocation. Economic theory suggests that increased stability from housing assistance may reduce the need for immediate work if the subsidy alleviates financial pressures ([Moffitt, 1992](#)). This effect could lead parents to reduce their working hours or exit the labor market, reallocating time to leisure, child-rearing, or other household activities. Alternatively, housing assistance might act as a stabilizing factor, enabling parents to pursue better employment opportunities or skill development as stable housing alleviates stress and allows focus on long-term goals ([Jacob and Ludwig, 2012](#)). Empirically, recent studies find that parents’ income and employment rates often decline after receiving housing assistance ([Barnhardt et al., 2017](#); [Kling et al., 2005](#); [Van Dijk, 2019](#)). In what follows, we consider whether a parent’s labor market response to the subsidy has downstream effects on their children’s education.

We present our results on labor market outcomes using a Poisson regression in [Table 1.7](#). The Poisson regression allows us to interpret our findings in percentage terms, while better handling of zeros through periods of unemployment. We find that after receiving the subsidy, parents reduce formal labor participation by 0.9% and report earnings in fewer months by a similar margin. This decrease is primarily driven by a reduction in work in permanent positions. Similarly, we find a reduction in their yearly earnings of 1.5% on average, which is mostly coming from jobs with permanent contracts. These findings align with literature suggesting a modest reduction in labor participation following housing assistance.

These findings suggest that the finances of subsidy-winning households improve sufficiently that parents work less on the margin. We might expect children to directly benefit from these improvements as well, if the family increases consumption in response. We test this possibility in [Figure A18](#). We present ITT estimates from the RD analysis on how likely children are to have breakfast (Panel A), lunch (Panel B), and fruit (Panel C) daily. There are no differences in consumption prior to the subsidy. After the subsidy, food consumption remains largely unchanged. The marginally-winning child has fruit at slightly elevated rates (significant at the 10%-level), though at similar levels prior to the application. Taken together, we do not find evidence that the subsidy is directly increasing children’s consumption.

Reduced work hours could also allow parents more time and energy to support their children’s education (Kalil and Ziol-Guest, 2008). In Figure 1.9, we use responses from the SIMCE questionnaire to assess how subsidy receipt impacts parental involvement in education. We find that parents in awarded households significantly increase their engagement after winning the subsidy. Specifically, parents are 1.8 percentage points (3.5%) more likely to keep track of their children’s grades (Panel A), 4.5 percentage points (5.5%) more likely to congratulate them for academic achievements (Panel B), and 4.9 percentage points (10%) more likely to help them study (Panel C). Importantly, we observe no significant differences in parental engagement prior to the application.

In Figure A20, we use responses from the SIMCE questionnaire to assess whether subsidy receipt improves parents’ aspirations for their children. We focus on high school completion (Panel A), college completion (Panel B), and graduate studies (Panel C), both before and after their *DS01* application. Before applying, expectations were similar on either side of the threshold. However, post-application, parents in winning households are 2.1 percentage points (3.2%) more likely to expect their child to complete college and 3.3 percentage points (25.4%) more likely to expect graduate-level studies.

These results align with previous research showing that housing assistance can improve a parent’s outlook for their children (Kumar, 2021). In our context, the improved expectations are paired with greater involvement in their children’s academics. In summary, our findings indicate that parental responses play a crucial role in mediating the positive effects of the subsidy on children’s academic performance. Parents from subsidy-receiving households reduce their formal labor market participation and increase their involvement in their children’s education by tracking grades, celebrating achievements, and providing study support. This increased engagement, alongside raised educational expectations, likely contributes to the educational gains we observe for children, as parental involvement has been shown to significantly boost children’s achievement (Guryan et al., 2008; Boneva and Rauh, 2018).

1.7 Early and Long-run Outcomes

Pre-school and Daycare

Preschool education is widely recognized as a critical foundation for children’s academic and social development, equipping them with essential cognitive and social skills that shape future educational outcomes. Research consistently associates preschool attendance with improved test scores, higher high school graduation rates, and increased college enrollment (Heckman and Raut, 2016; Deming, 2009; Gray-Lobe et al., 2022). For disadvantaged children, preschool is particularly beneficial, as it provides structured learning environments that might otherwise be inaccessible, helping to bridge early achievement gaps and promote educational equity.

In Table 1.8, we present the impact of winning the subsidy (ITT) and using it to purchase

a home (ATE) on daycare and preschool attendance. We analyze all levels collectively and then by specific categories: *Sala Cuna* (ages 0–2), *Nivel Medio* (ages 2–4), pre-kindergarten (age 5), and kindergarten (age 6). The results indicate that winning a voucher is associated with a 2.2 percentage point increase in preschool attendance among children ages 0–5, reflecting a 4.5% rise relative to the control mean. This effect is primarily driven by increased enrollment in early preschool, while we observe no significant differences in daycare attendance or for children over age 5.

To our knowledge, this study is the first to explore the impact of housing assistance on daycare and preschool attendance. Previous studies on housing programs have typically focused on later educational outcomes, such as high school or college achievements. By examining preschool attendance, particularly across detailed age groups, our analysis offers new insights into how housing stability may affect young children during formative years.

Although we observe an increase in preschool attendance, we do not find that primary and secondary school outcomes are stronger for children who were preschool-aged at the time of application. This may suggest that housing stability alone is insufficient to yield significant advantages in subsequent schooling for younger children or that the benefits may require a longer time to materialize. Given our study’s current time frame, we lack sufficient data to explore the potential long-term impacts of preschool attendance on high school completion or college outcomes for these children.

Post-secondary Education

Improvements in secondary academic performance, coupled with elevated parental expectations, can lay the groundwork for increased post-secondary attainment. Research shows that students are more likely to complete college when their parents hold high educational aspirations (Kumar, 2021; Zhang et al., 2011). Given the positive effects we observe in primary and secondary education, these academic gains may extend into post-secondary outcomes (Heckman et al., 2006; Chetty et al., 2014; Oreopoulos and Salvanes, 2011; Jackson, 2018). We explore this directly by analyzing the impact of homebuyer subsidies on various post-secondary outcomes.

Table 1.9 presents our findings for high school completion, college entry exam scores, and college enrollment. We find that winning the subsidy increases high school graduation rates by 1.4 percentage points (1.6%), corresponding to a 5.8% increase in completion rates among subsidy users. Children from awarded households also see an average improvement in high school grades of 0.038 points. While these students are equally likely to take the college entry exam as their non-awarded counterparts, those who do take the exam score 0.043-0.062 standard deviations higher in math, verbal, history, and science exams. This corresponds to a 0.25 standard deviation improvement for families that use the subsidy to buy a home.

Additionally, students in awarded households are 1.8 percentage points (3%) more likely to attend college, which corresponds to a 14% increase relative to the control mean for

subsidy users. Those attending college extend their studies by approximately 1.5 months (a 7% increase with respect to the control mean). For families using the subsidy, this impact is closer to 6.5 months. They are also more likely to enroll in a university rather than a technical institution. However, conditional on attending college, they are slightly more likely to drop out, although this difference is not statistically significant. Lastly, we find no indication that they attend institutions with significantly longer accreditation periods by the Ministry of Education.

Our findings on high school completion and post-secondary enrollment are in line with results from other housing policies when considering the ITT estimates (Chetty et al., 2016a; Jacob et al., 2015; Camacho et al., 2022; Kumar, 2021; Laliberté, 2021). When examining subsidy users, our ATE estimates are among the higher impacts observed in the literature, though slightly below those of Agness and Getahun (2024). Our results also compare in magnitude to conditional cash transfers, tax credit programs (Dahl and Lochner, 2012; Behrman et al., 2005), and early childhood interventions (Deming, 2009; Heckman et al., 2010), indicating that stable housing support can yield lasting educational benefits on par with financial and early childhood investments.

Following the approach of Chetty et al. (2016a), Table A9 presents results by age at application. We find that effects on test scores and high school completion are primarily driven by children already in high school at the time of application. This contrasts with prior studies that generally observe higher benefits from longer policy exposure (Chetty et al., 2016a; Chyn and Katz, 2021; Agness and Getahun, 2024). However, consistent with the broader literature, we observe that effects on college attendance are strongest for children under age 13, supporting the view that early exposure to housing stability has a lasting impact on post-secondary outcomes.

In Figure A17, we present results for high school completion and college enrollment, examining heterogeneity across demographic and neighborhood characteristics. Echoing our findings in primary and secondary education, applicants' individual characteristics do not significantly alter outcomes. However, the effects remain stronger for larger households and those from low- and medium-quality neighborhoods. Interestingly, age at application shows no significant variation for students still in school at application. We also find no effect for students who were 18 or older at the time, implying minimal preexisting differences in educational outcomes before receiving the subsidy.

Labor Market

Educational attainment and performance in primary, secondary, and post-secondary schooling are powerful predictors of labor market outcomes, including employment stability, earnings potential, and career progression. Research has demonstrated that academic skills acquired in early education foster both cognitive and non-cognitive abilities, which are crucial for workplace productivity and adaptability (Card, 1999; Heckman et al., 2006; Heckman and Kautz, 2012). Skills developed through early academic achievement—such as critical

thinking, perseverance, and social adaptability—are particularly influential for long-term professional success. The positive educational outcomes observed in our early cohorts likely pave the way for improved labor market prospects, as the cumulative benefits of academic success reinforce career progression and earnings over time.

Table 1.10 presents the results of a Poisson regression on labor market outcomes for children not enrolled in post-secondary education. We find that children from subsidy-winning households are 1.2% more likely to be employed and work, on average, 3% more months per year. This increase, which equates to roughly 0.15 additional months of employment annually, is observed across both permanent and temporary positions. Although we observe a slight increase in yearly earnings for this group, the effect is not statistically significant. Given our study’s time frame, these young adults represent children who were older at the time of application or part of the policy’s earliest cohorts. Additional longitudinal data will be necessary to capture the full impact on labor market outcomes for those who were younger at the time of application.

Our findings align with the broader literature on housing assistance and labor market outcomes, albeit with modest effects compared to other policy contexts (Jacob and Ludwig, 2012).³⁵ However, while studies like the MTO experiment documented limited labor market improvements for adults (Chetty et al., 2016a), our results suggest that even small increases in stability and employment during formative years can yield incremental but positive labor market outcomes for young adults.

In terms of magnitude, our effects are smaller than the substantial gains in labor market outcomes seen in studies like Dahl and Lochner (2012), which analyzed income support policies and found significant increases in earnings and employment. This may reflect differences in how direct income transfers, such as tax credits, versus housing-based support influence labor market behavior. Nonetheless, our results suggest that housing stability provides a modest, yet meaningful, boost to employment consistency among young adults, especially those from lower-income backgrounds. These incremental gains may amplify over time, indicating that housing assistance programs could play a valuable role within a larger strategy to enhance long-term labor market outcomes for children from disadvantaged households.

Net benefits calculation

The *DS01* subsidy increased educational attainment at every point in a child’s school career. However, we currently lack enough data to observe earnings increases for those who attend college. To calculate the net benefits of the program, we perform a simple back-of-the-envelope calculation that extrapolates increases in college attendance to lifetime earnings through the returns to college education in Chile. Following Chetty et al. (2016a), we use the

³⁵Our estimates for months worked are similar to findings from the Welfare-to-Work Voucher Program, which noted temporary reductions in labor supply immediately following assistance but longer-term improvements in employment stability (Mills et al., 2006).

2022 average earnings of \$23,005 USD for the entire Chilean population as our benchmark.³⁶ We use our estimate of a 10% increase in college attendance due to the subsidy with a 62% return to college degrees in Chile ([González-Velosa et al., 2015](#)).³⁷

Taken together, we estimate the total present value of lifetime benefits to children from the homebuyer subsidy is equal to \$22,403 USD. For the average winning household with 1.7 children, this estimate is \$38,085 USD. Subsidy-winning households see their incomes decline 1.5% from a baseline of \$3,389 USD.³⁸ The remaining present value of lifetime benefits to household income from the homeowner subsidy is equal to -\$704 USD.³⁹ At an average subsidy cost of \$20,000 USD, net benefits to the average family are \$17,381 USD, and the marginal value of public funds is 1.84. This is comparable to programs that subsidize college education, like CUNY’s Pell Grant program and Tennessee’s Hope Scholarship ([Hendren and Sprung-Keyser, 2020](#)).

1.8 Conclusion

The UN estimates that 1.6 billion people globally live in inadequate housing ([UN-Habitat, 2022](#)), ranging from slums to overcrowded or otherwise poor-quality homes. Relatedly, housing tenure and homeownership are often recognized as pivotal mechanisms for breaking cycles of poverty, as stable housing plays a fundamental role in wealth accumulation and intergenerational mobility ([Chetty et al., 2014](#); [Chetty and Hendren, 2018](#)). However, few housing policies have directly aimed to expand private homeownership among low-income families, despite its potential to drive long-term economic mobility.

This paper examines the causal effects of a generous and large-scale homebuyer subsidy in Santiago, Chile. We find that, like other housing policies before it, parents slightly reduce their employment and earnings in response. However, the focus of this paper - and perhaps where the focus of the impact of most housing policies should be - is on the children of subsidy recipients. This has been cited as a primary motivation for recent proposals of homebuyer subsidies in the U.S. context.⁴⁰

³⁶Average earnings number from [Instituto Nacional de Estadísticas \(INE\)](#). At a 5% interest rate, this gives a PDV lifetime (30-year) earnings estimate of \$361,345 USD.

³⁷While not all college attendees will graduate, we think this is still a conservative estimate. Our estimates of college attendance do not contain all of the students with higher grades and performance in K-12, so that estimate is likely to increase over time as more data becomes available. Moreover, even some college attendance is likely to improve earnings, and we see improved labor market outcomes for those who do not attend college.

³⁸To contextualize this number, many parents are not employed in the formal sector in our sample.

³⁹Because the average winning applicant is 37 years old, we assume there are 23 more years they could work.

⁴⁰U.S. Vice President Kamala Harris said the following in a recent campaign speech proposing a homebuyer subsidy: “Together, we will build what I call an ‘opportunity economy’ ... an economy where everyone can compete and have a real chance to succeed; everyone, regardless of who they are or where they start, has an opportunity to build wealth for themselves and their children.” [Source: whitehouse.gov](#).

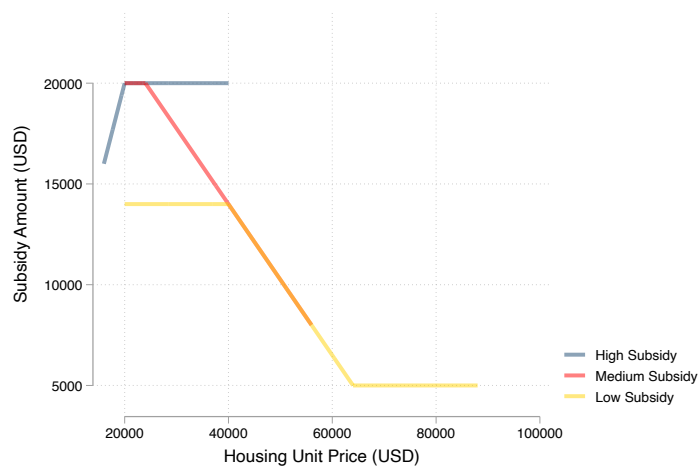
Our findings reveal substantial academic gains among children of subsidy recipients, evidenced by improved grades, class rankings, and achievement test scores, as well as reductions in chronic absenteeism. Increased preschool attendance rates highlight the policy's early positive influence on school readiness and early learning. Parents' expectations for their children's futures, and the aspirations of the children themselves, improve. Additionally, we observe that these educational benefits are more pronounced in larger families, suggesting that alleviating overcrowded living conditions is a key pathway through which the subsidy impacts child development. This is supported by marginally winning children reporting more hours of sleep on average. We do not find that increased resources translate into more consumption for children, even when less than 80% report having breakfast every day. While parents reduce their work, they also become more involved and invested in their child's education.

The evidence suggests that these effects are not driven by relocation to higher-quality schools or neighborhoods, underscoring the importance of housing stability and parental involvement in fostering human capital independently of external environmental changes. As children transition to adulthood, we find that those from subsidy-recipient families are more likely to complete high school, attend college, and secure employment if they enter the labor market directly compared to their peers from non-subsidy households.

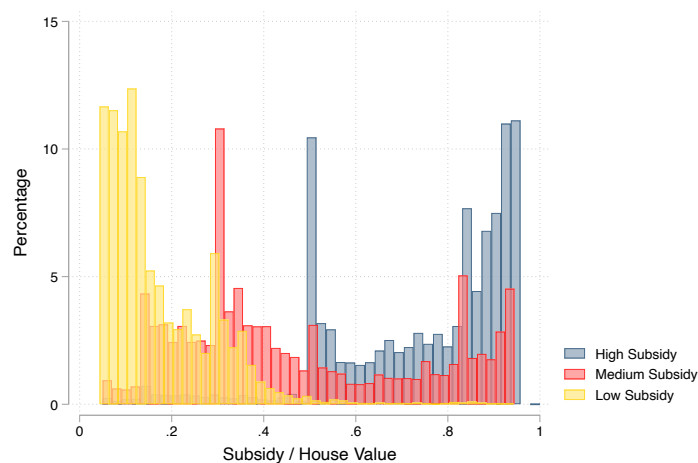
The program's benefits appear substantial relative to its costs. A preliminary calculation indicates that the policy yields an estimated \$38,000 USD increase in the present value of lifetime earnings for children against a program cost of approximately \$20,000 USD per household. These findings suggest that well-designed, targeted homebuyer subsidies can be a powerful policy tool for improving children's educational attainment and, ultimately, enhancing intergenerational mobility, making a strong case for integrating such initiatives within broader social welfare and housing policies.

Figures

Figure 1.1: Voucher Subsidy Amount and House Prices



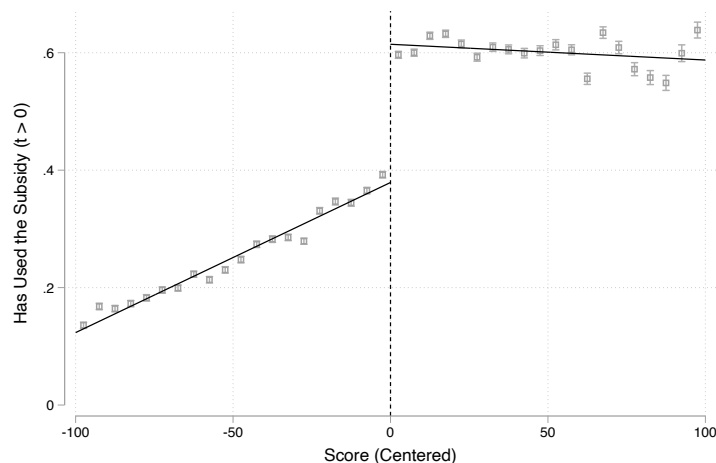
Panel A: Subsidy Amount vs Unit Price



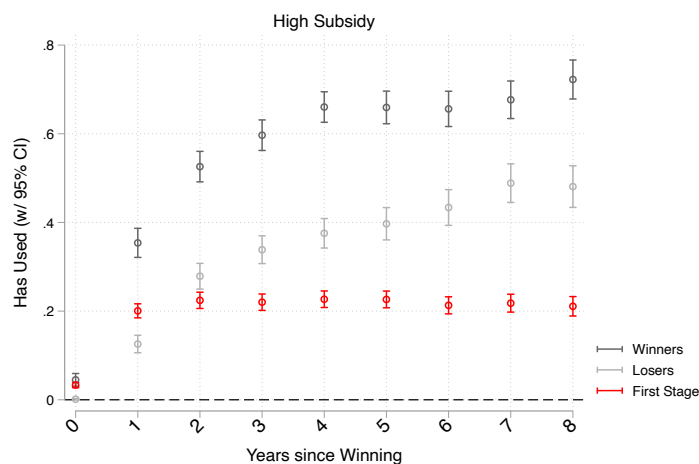
Panel B: House Price Covered by the Voucher

Notes: Panel A displays the voucher subsidy amount as a function of the housing unit price in USD by subsidy tier. The amount is fixed for the high subsidy as long as the house price exceeds the subsidy amount, while it decreases with house value for medium and low subsidies, reaching a minimum. The pricing cap for the high subsidy is more restrictive than for the medium and low subsidies. Panel B shows the distribution of the voucher amount relative to the house value by subsidy tier, covering 85% of the house price on average for the high subsidy, and 55% and 18% for the medium and low subsidies, respectively.

Figure 1.2: First Stage



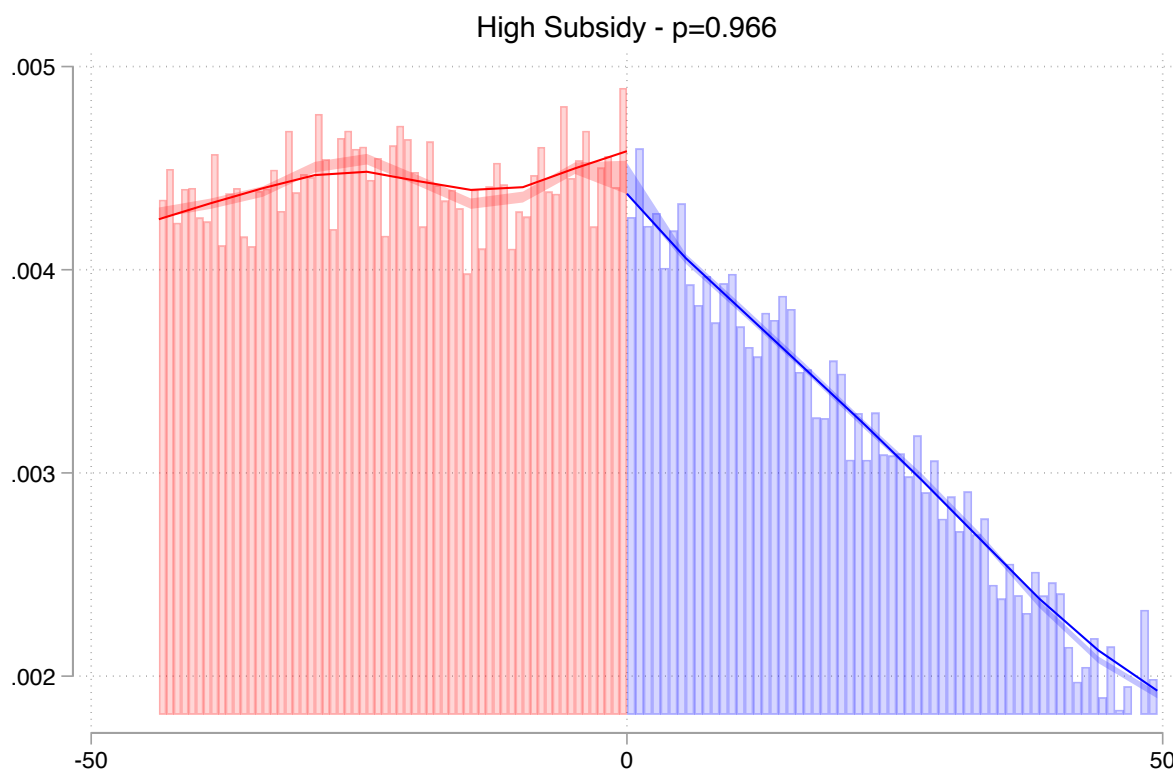
Panel A: Pooled First Stage



Panel B: Evolution of First Stage over time

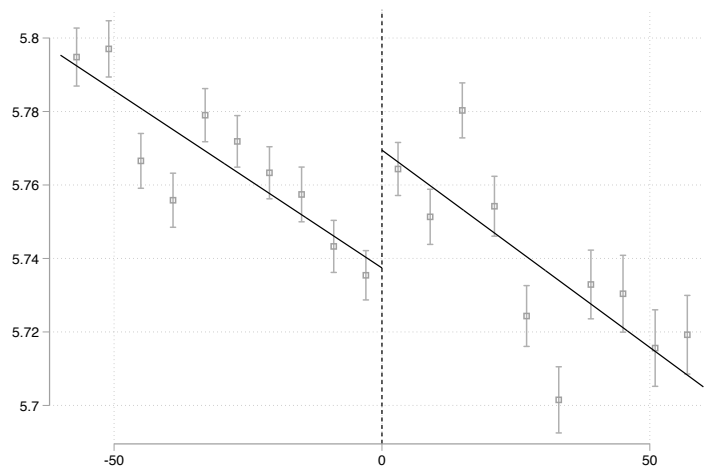
Notes: Panel A displays the first stage of the regression discontinuity, running [Equation 3.6](#) on having used the subsidy after application and pooling across application calls and years. We find a 0.25 percentage point jump around the threshold, suggesting we will need to rescale the ITT estimates by approximately four to find the LATE of buying a house using the voucher. Panel B shows the evolution of the cumulative fraction of applicants that use the subsidy in each period after the application for both awarded and non-awarded applicants, pooling application calls and controlling for the score using a polynomial degree 1. The first stage in each period is given by the difference between the fraction for winners and losers, and it is stable at 0.22 percentage points over time.

Figure 1.3: McCrary test after application

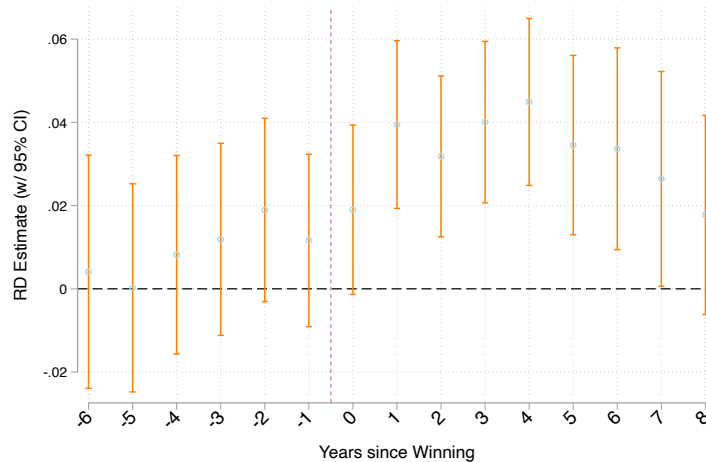


Notes: This figure shows the result of the McCrary test for manipulation of the scores pooling across application calls. We find a p-value of 0.966, suggesting we cannot reject the no-manipulation of the scores, providing evidence of the validity of the regression discontinuity design. This is not surprising, given the varying nature of the cutoff scores across application scores.

Figure 1.4: Regression Discontinuity and Event Study: Grades



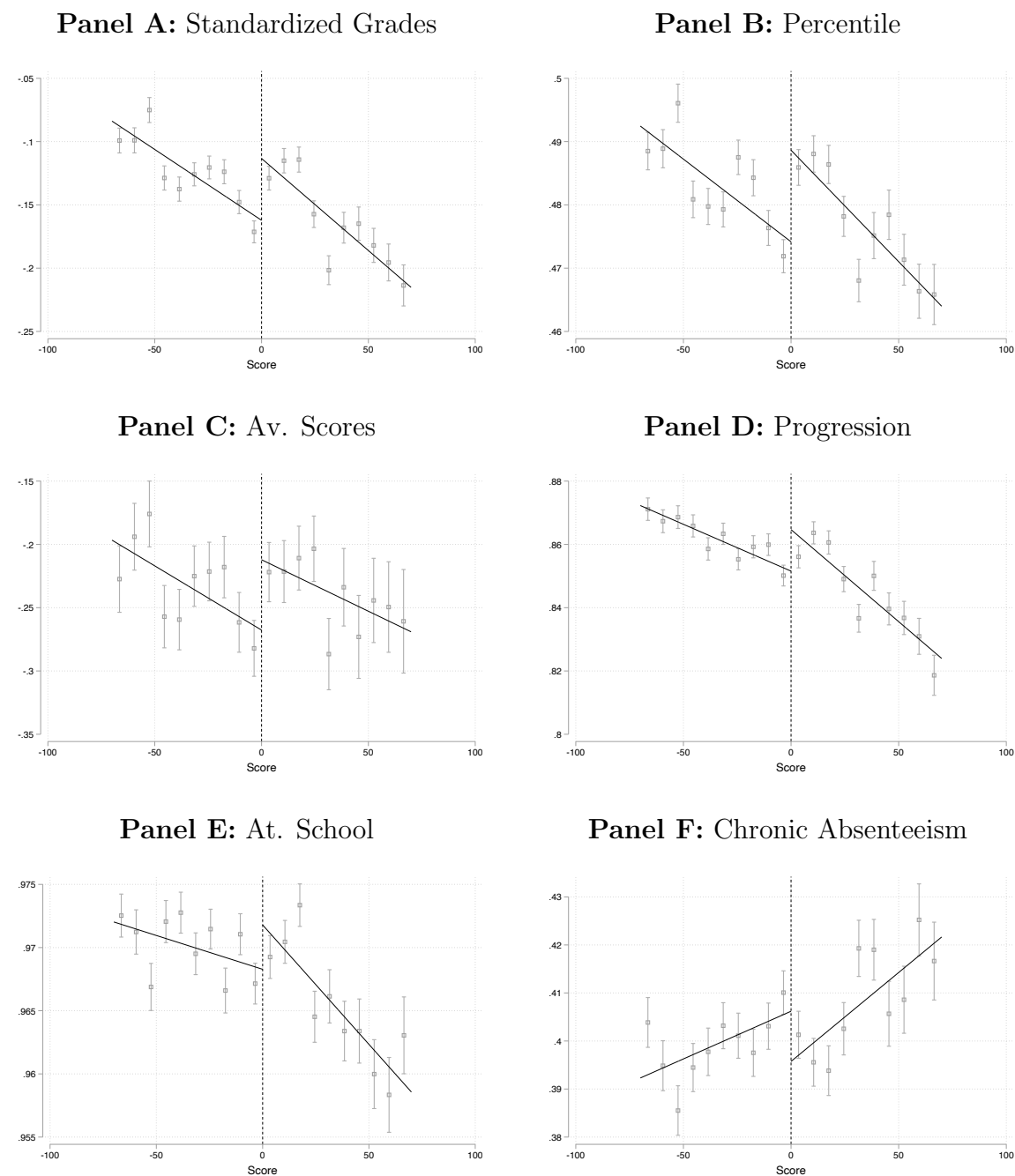
Panel A: Regression Discontinuity



Panel B: Event Study

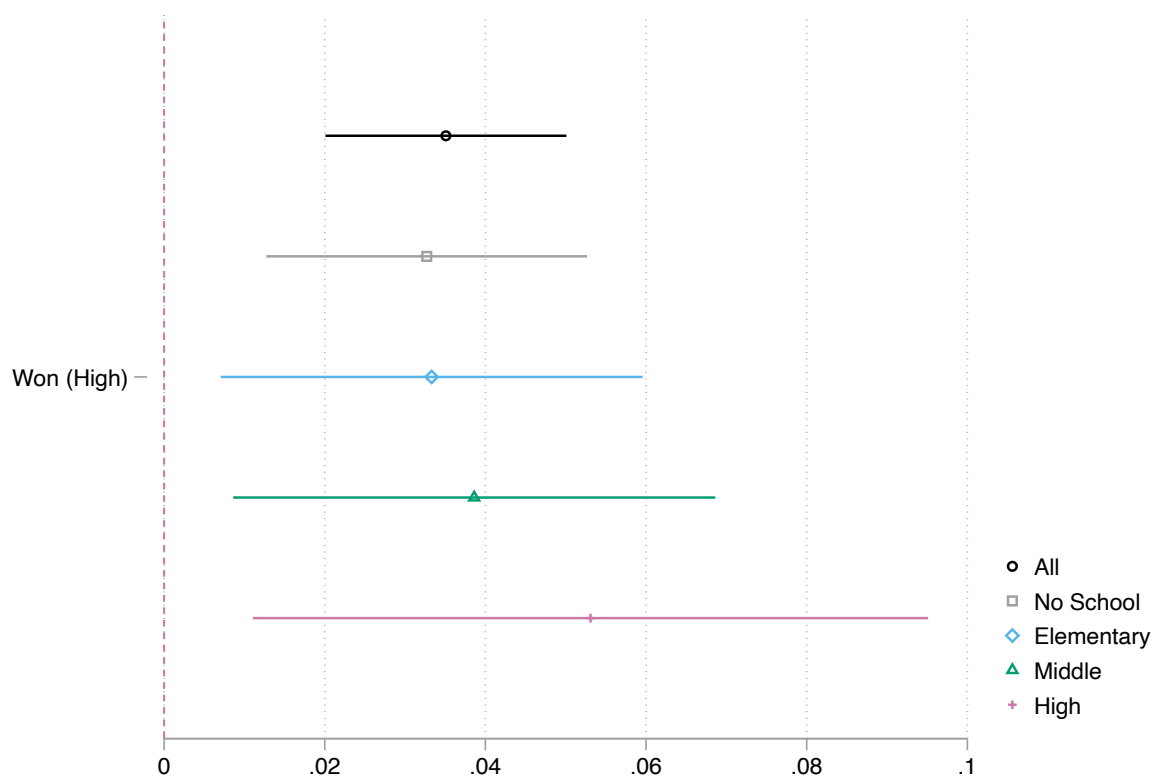
Notes: Panel A displays the coefficient of Equation 1.1 on grades for the high subsidy, shown in Table 1.3. Panel B displays the evolution of grades over time in relation to application, pooling across calls. Standard errors are clustered at the applicant level and consider scores using the IMSE-optimal bandwidth. Controls include year and gender-by-age fixed effects.

Figure 1.5: RD plot for main achievement outcomes



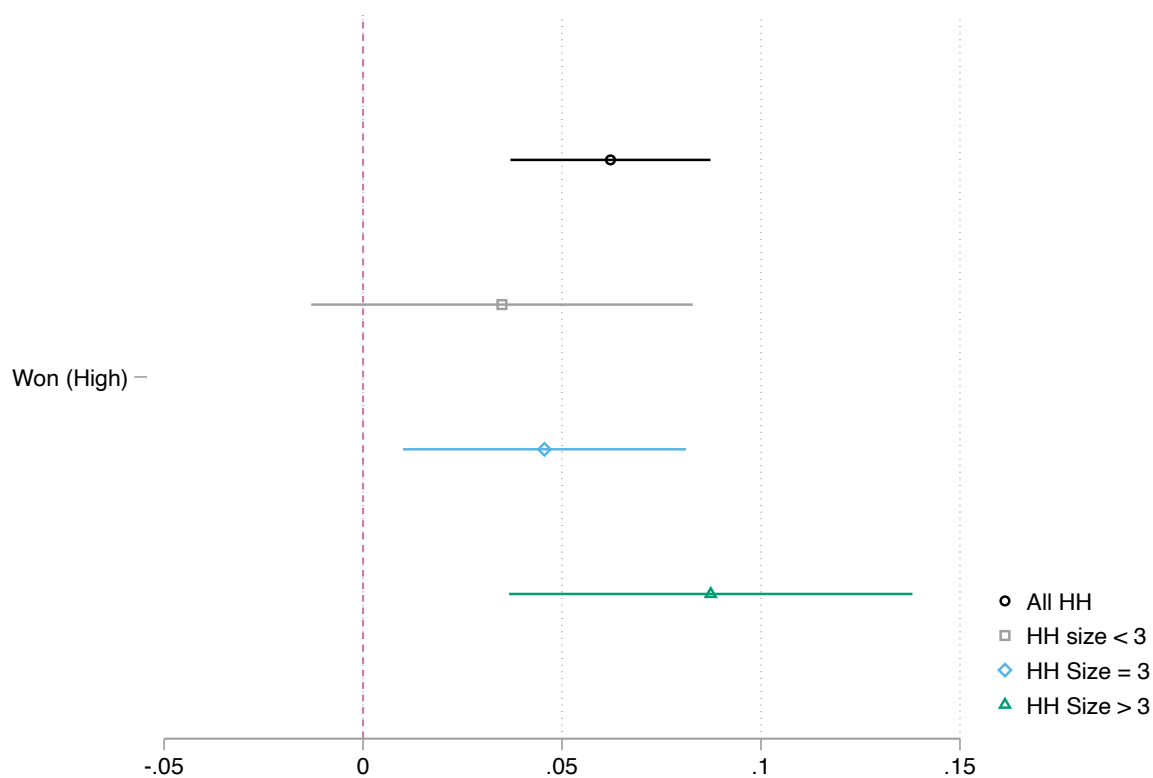
Notes: This figure shows the result of the estimate of the ITT estimate from the regression discontinuity for some of the achievement outcomes, pooling years after the application. Outcomes include standardized grades (A), class rank percentile (B), average test scores (C), probability of not being retained (D), indicator of not having dropped out of school (E), and chronic absenteeism (F). Standard errors are clustered at the applicant level, and regression considers scores using the IMSE-optimal bandwidth. Controls include year and gender-by-age fixed effects. ***, **, * indicates significance at 1, 5, and 10%.

Figure 1.6: Heterogeity by Age at Application: Grades



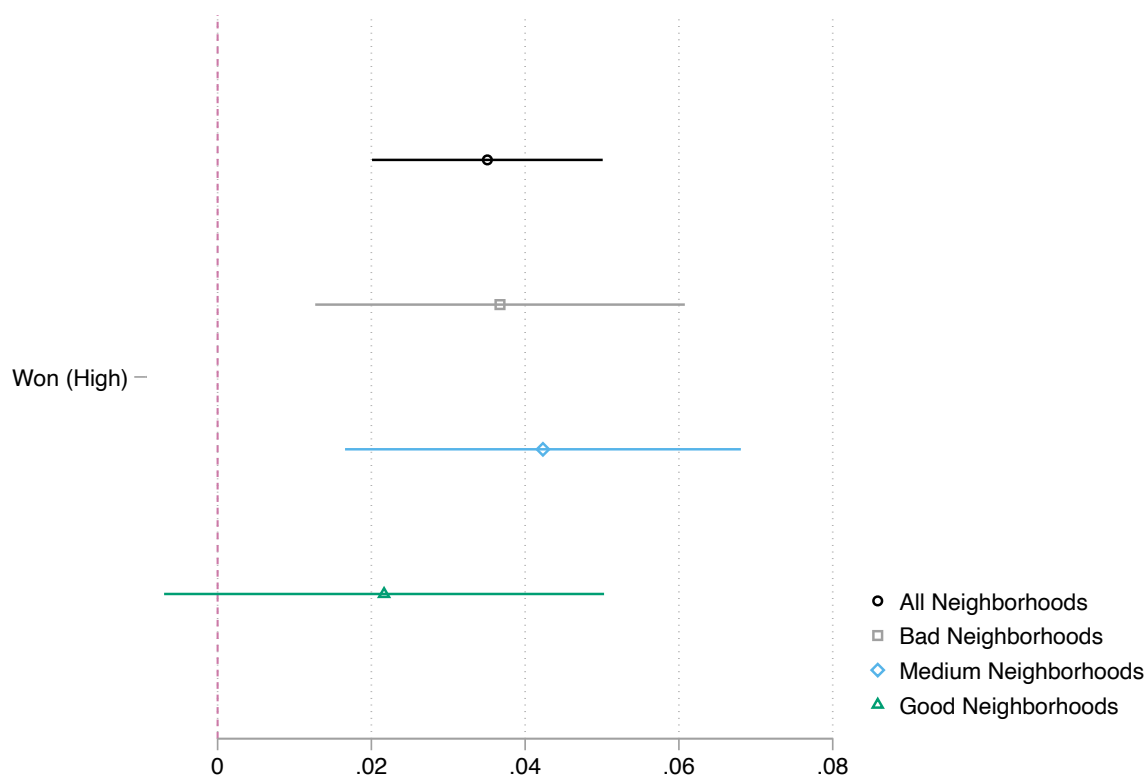
Notes: This figure displays the heterogeneous effects depending on age at application for the ITT estimates on grades. Each horizontal bar reports the coefficient of the jump at the cutoff and the confidence interval of the estimate at the 10% significance level, using Equation 1.1 and fitting a linear polynomial at each side of the cutoff. All estimates include year and gender by age fixed effects and consider scores the IMSE-optimal bandwidth. Standard errors are clustered at the applicant level.

Figure 1.7: Heterogeity by Household Size: Grades



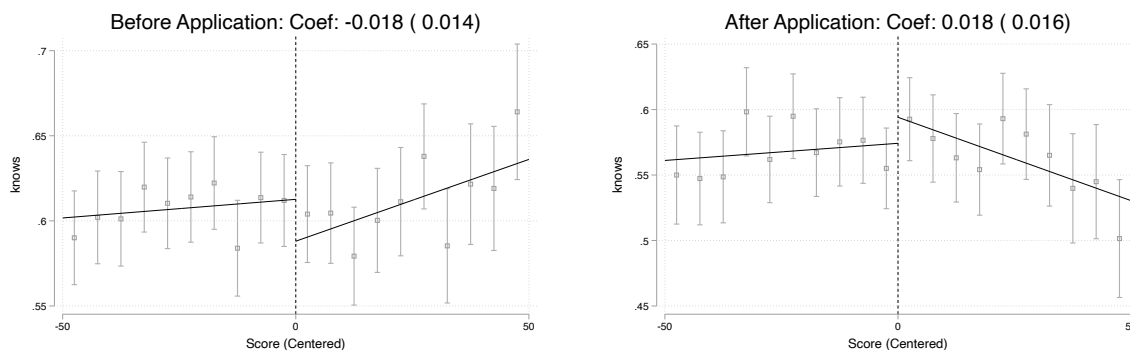
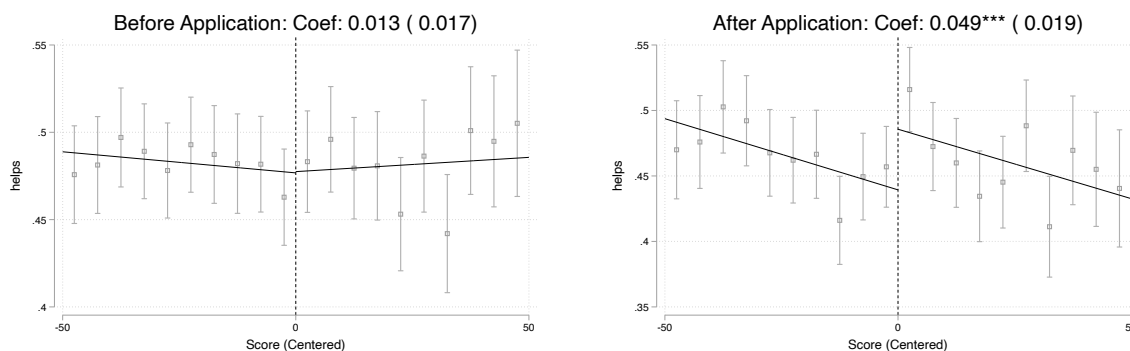
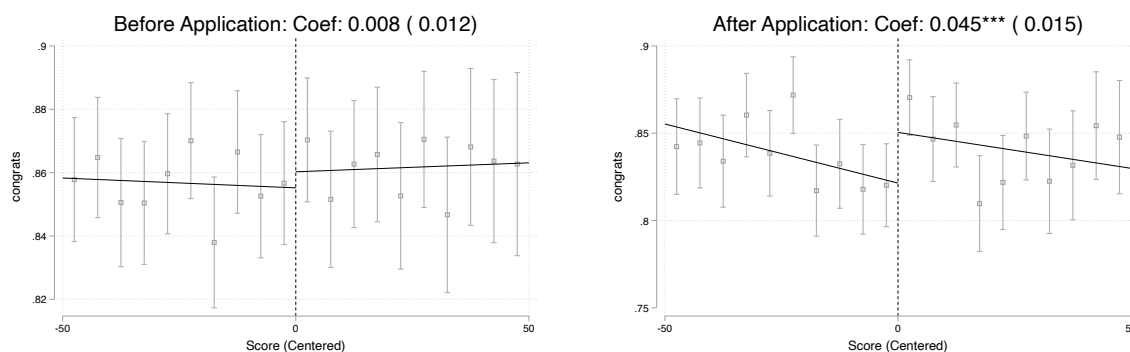
Notes: This figure displays the heterogeneous effects depending on household size at application for the ITT estimates on grades. Each horizontal bar reports the coefficient of the jump at the cutoff and the confidence interval of the estimate at the 10% significance level, using Equation 1.1 and fitting a linear polynomial at each side of the cutoff. All estimates include year and gender by age fixed effects and consider scores the IMSE-optimal bandwidth. Standard errors are clustered at the applicant level.

Figure 1.8: Heterogeneity by Neighborhood of Origin: Grades



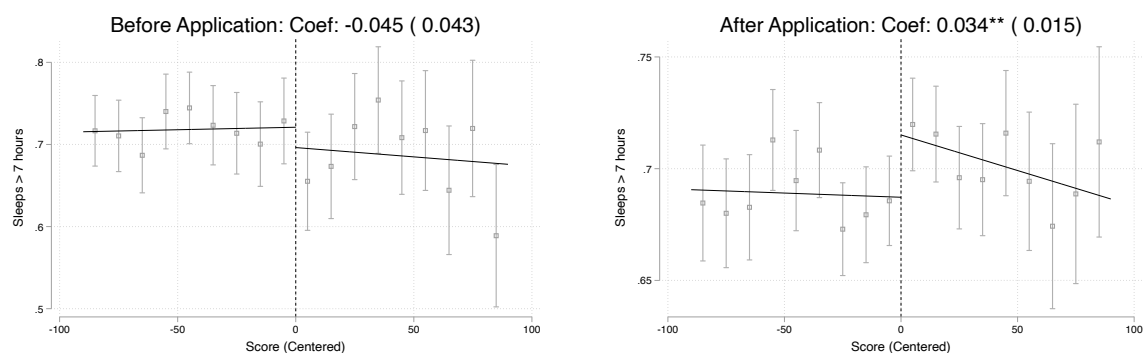
Notes: This figure displays the heterogeneous effects depending on neighborhood quality of origin for the ITT estimates on grades, using average years of schooling at the census tract level. Each horizontal bar reports the coefficient of the jump at the cutoff and the confidence interval of the estimate at the 10% significance level, using Equation 1.1 and fitting a linear polynomial at each side of the cutoff. All estimates include year and gender by age fixed effects and consider scores the IMSE-optimal bandwidth. Standard errors are clustered at the applicant level.

Figure 1.9: RD of parents' behavior

Panel A: Parents know children's grades**Panel B: Parents help children study****Panel C: Parents congratulate children's good grades**

Notes: This figure shows the result of the ITT estimate of the regression discontinuity regarding parental responses in relation to their children's studies. In Panel A, the outcome variable indicates whether the parent knows their children's grades. Panel B indicates whether the parent helps their children study, and Panel C whether they congratulate their children when they get good grades. Standard errors are clustered at the applicant level, and scores are considered using the IMSE-optimal bandwidth. Controls include year and gender-by-age fixed effects. ***, **, * indicates significance at 1, 5, and 10%.

Figure 1.10: RD on hours of sleep
Do you (the child) sleep 7 or more hours?



Notes: This figure shows the result of the ITT estimate of the regression discontinuity regarding children's hours of sleep, considering the fraction that report sleeping seven or more hours per night. The left-hand panel shows the difference before application, and the right-hand panel after the application. Standard errors are clustered at the applicant level, and scores are considered using the IMSE-optimal bandwidth. Controls include year and gender-by-age fixed effects. ***, **, * indicates significance at 1, 5, and 10%.

Tables

Table 1.1: Differences in Demographics at application

Dep. Variable	Control Mean	ITT	Observations	Bandwidth
Panel A: Application Characteristics (Applicant)				
Applicant's Age	36.363	0.205 (0.184)	38,859	55.373
Applicant's Gender	0.920	-0.001 (0.006)	41,725	60.751
Married = 1 (Applicant)	0.305	0.012 (0.012)	38,642	54.984
Single-Parent	0.621	-0.013 (0.014)	34,601	47.990
Household Size	3.005	0.007 (0.018)	38,162	54.121
Children at School	1.426	-0.002 (0.020)	41,479	60.290
Former Applications	1.177	0.035 (0.042)	42,930	63.128
Self-reported Income	11.401	0.157 (0.135)	35,528	49.570
NH quality (ed)	9.930	0.028 (0.032)	48,056	73.660
Panel B: Children Characteristics				
Age	11.895	0.041 (0.056)	53,886	86.933
Ind. Female	0.495	-0.008 (0.009)	50,031	78.023
Grade	5.979	0.024 (0.053)	56,781	94.746

Notes: This table shows the differences in demographic characteristics at the time of application. Panel A presents the characteristics of the application, while Panel B shows the characteristics of the children. Columns (1), (3), and (4) display the mean of the variable for non-awarded households, the number of applicants, and the optimal IMSE-bandwidth to perform the RD, respectively. Column (2) shows the ITT estimate of the regression discontinuity using Equation 1.1 by regressing each demographic characteristic on the score, corresponding to the jump at the cutoff. Standard errors are clustered at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Table 1.2: Differences in Outcomes at application

Dep. Variable	Control Mean	ITT	Observations	Bandwidth
Panel A: Achievement Outcomes				
Grades	5.685	0.009 (0.013)	34,762	55.138
Grades (sd)	-0.151	0.017 (0.022)	35,337	56.325
Percentile	0.485	0.005 (0.006)	36,729	59.245
Ind. Percentile > 0.5	0.473	0.009 (0.011)	35,904	57.543
Score Verbal	-0.205	0.012 (0.036)	11,682	60.264
Score Math	-0.246	0.022 (0.038)	10,042	71.152
Av. Score	-0.241	0.018 (0.035)	11,791	58.728
Ind. Progression	0.830	0.007 (0.008)	41,759	65.298
Ind. at School	0.956	0.003 (0.004)	47,089	77.336
Ind. Dropout (> 14)	0.005	0.003 (0.002)	15,008	82.984
Ind. Repeated	0.044	-0.002 (0.004)	49,180	81.940
Absenteeism	0.100	-0.002 (0.002)	44,032	75.813
Chronic Absenteeism	0.362	-0.001 (0.009)	47,707	82.954
Panel B: Other Outcomes				
Ind. Public School	0.341	-0.001 (0.009)	53,294	93.040
Ind. Private	0.004	0.001 (0.001)	38,537	58.744
School Quality	0.444	0.005 (0.005)	44,569	75.500
Class Size	36.012	0.028 (0.162)	43,983	74.000
Ind. Change School	0.119	-0.000 (0.007)	43,169	72.071
Ind. Change Comuna	0.026	-0.000 (0.002)	41,025	74.975
Ind. Priority Student	0.612	-0.001 (0.012)	32,879	50.385

Notes: This table shows the differences in the main primary and secondary school outcomes at the time of application. Panel A presents the main achievement outcomes, while Panel B shows additional variables related to school quality. Columns (1), (3), and (4) display the mean of the variable for non-awarded households, the number of applicants, and the optimal IMSE-bandwidth to perform the RD, respectively. Column (2) shows the ITT estimate of the regression discontinuity using Equation 1.1 by regressing each demographic characteristic on the score, corresponding to the jump at the cutoff. Standard errors are clustered at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Table 1.3: Achievement outcomes

	Control Mean	ITT	ATE	Observations	Bandwidth
Dep. Variable	(1)	(2)	(3)	(4)	(5)
Grades	5.685	0.035*** (0.009)	0.163*** (0.044)	374,813	55.075
Grades (sd)	-0.151	0.059*** (0.014)	0.273*** (0.070)	381,350	56.270
Percentile	0.485	0.017*** (0.004)	0.079*** (0.020)	396,337	59.219
Ind. Percentile > 0.5	0.473	0.026*** (0.007)	0.122*** (0.032)	387,852	57.537
Av. Score	-0.240	0.058*** (0.021)	0.321*** (0.121)	56,408	58.371
Score Math	-0.247	0.052*** (0.019)	0.272** (0.107)	55,622	70.234
Score Verbal	-0.205	0.046** (0.021)	0.264** (0.122)	55,269	60.003
Ind. Progression	0.865	0.012** (0.005)	0.059** (0.025)	394,492	58.882
Ind. at School	0.956	0.004** (0.002)	0.019** (0.009)	501,755	77.336
Ind. Dropout	0.003	-0.000 (0.000)	-0.000 (0.001)	513,710	85.808
Ind. Repeated	0.033	-0.002* (0.001)	-0.009* (0.005)	465,443	73.655
Absenteeism	0.100	-0.001 (0.001)	-0.004 (0.005)	473,524	75.553
Chronic Absenteeism	0.366	-0.009* (0.005)	-0.041* (0.022)	506,994	84.025
Controls	-	Yes	Yes	-	-

Notes: This table shows the differences in the main primary and secondary school achievement outcomes after application. Columns (1), (4), and (5) display the mean of the variable for non-awarded households, the number of applicants, and the optimal IMSE-bandwidth to perform the RD, respectively. Column (2) shows the ITT estimate of the regression discontinuity using Equation 1.1 by regressing each demographic characteristic on the score, corresponding to the jump at the cutoff. Column (3) presents the LATE estimates, rescaling the ITT estimates by the first stage in Equation 3.6 and presenting the effects of using the voucher to buy a house. All estimates include year and gender by age fixed effects and clustered standard errors at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Table 1.4: Achievement outcomes: Robustness

	Mean	Preferred	No Control	Cont t-1	Av. BW	No Weight	Poly 2	No COVID
Dep. Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grades	5.685	0.035*** (0.009)	0.035*** (0.009)	0.027*** (0.010)	0.035*** (0.009)	0.032*** (0.008)	0.039*** (0.013)	0.044*** (0.010)
Grades (sd)	-0.151	0.059*** (0.014)	0.059*** (0.014)	0.040*** (0.014)	0.058*** (0.014)	0.056*** (0.013)	0.065*** (0.021)	0.075*** (0.017)
Percentile	0.485	0.017*** (0.004)	0.017*** (0.004)	0.007* (0.004)	0.017*** (0.004)	0.016*** (0.004)	0.020*** (0.006)	0.020*** (0.005)
Ind. Percentile > 0.5	0.473	0.026*** (0.007)	0.026*** (0.007)	0.012* (0.007)	0.026*** (0.006)	0.023*** (0.006)	0.034*** (0.010)	0.031*** (0.008)
Av. Score	-0.241	0.057*** (0.021)	0.057*** (0.021)	0.055** (0.022)	0.057*** (0.021)	0.057*** (0.019)	0.065** (0.031)	0.075*** (0.022)
Score Math	-0.246	0.050*** (0.019)	0.050*** (0.019)	0.050** (0.021)	0.051** (0.021)	0.051*** (0.017)	0.052* (0.029)	0.067*** (0.020)
Score Verbal	-0.205	0.046** (0.021)	0.046** (0.021)	0.039* (0.023)	0.045** (0.021)	0.050*** (0.019)	0.051 (0.031)	0.060*** (0.022)
Ind. Progression	0.830	0.015*** (0.005)	0.015*** (0.005)	0.007 (0.006)	0.015*** (0.006)	0.015*** (0.005)	0.015* (0.008)	0.019*** (0.006)
Ind. at School	0.956	0.004** (0.002)	0.004** (0.002)	0.005* (0.003)	0.004* (0.002)	0.003* (0.002)	0.004 (0.003)	0.005** (0.002)
Ind. Dropout	0.010	-0.000 (0.001)	-0.000 (0.001)	-0.001 (0.001)	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)	-0.001 (0.001)
Ind. Repeated	0.044	-0.003** (0.001)	-0.003** (0.001)	-0.003* (0.002)	-0.003** (0.001)	-0.002** (0.001)	-0.003* (0.002)	-0.006*** (0.002)
Absenteeism	10.020	-0.105 (0.112)	-0.105 (0.112)	-0.009 (0.129)	-0.143 (0.126)	-0.059 (0.099)	-0.204 (0.168)	-0.329*** (0.121)
Chronic Absenteeism	0.362	-0.009* (0.005)	-0.009* (0.005)	-0.006 (0.005)	-0.011** (0.006)	-0.005 (0.004)	-0.012* (0.007)	-0.012** (0.006)
Observations		375,205	375,205	183,857	399,831	375,205	375,205	201,675
Children		62,636	62,636	31,315	66,793	62,636	62,636	32,555
Controls		Yes	No	Yes	Yes	Yes	Yes	Yes

Notes: Each cell reports the coefficient of the jump at the cutoff, using [Equation 1.1](#). In column (1), we present our preferred specification, controlling for age by gender and using the IMSE-optimal bandwidth. Column (3) uses no controls, and column (4) controls for the outcome at application. Column (5) uses the average optimal bandwidth across outcomes, and column (6) uses no weights. Column (7) presents the results using a polynomial degree 2, and column (8) restricts the sample to the years before COVID (before 2020). Standard errors clustered at the applicant level. ***, **, * indicates significance at 1, 5, and 10%.

Table 1.5: Achievement outcomes by Gender

Dep. Variable	Male		Female	
	ITT	ATE	ITT	ATE
	(1)	(2)	(3)	(4)
Grades (sd)	0.071*** (0.019)	0.344*** (0.095)	0.046** (0.019)	0.200** (0.090)
Percentile	0.020*** (0.006)	0.099*** (0.027)	0.014** (0.006)	0.058** (0.027)
Av. Score	0.085*** (0.029)	0.511*** (0.177)	0.029 (0.029)	0.134 (0.155)
Ind. Progression	0.021*** (0.008)	0.102*** (0.038)	0.009 (0.007)	0.039 (0.032)
Absenteeism	-0.226 (0.148)	-1.026 (0.696)	0.018 (0.144)	0.164 (0.671)
Mean Grades (sd)	-0.284		-0.028	
Mean Percentile	0.450		0.518	
Mean Av. Score	-0.268		-0.209	
Mean Ind. Progression	0.797		0.856	
Mean Absenteeism	10.173		10.023	
Bandwidth	Optimal		Optimal	
Observations	190,315		184,890	
Children	62,576		60,312	
Controls	Yes		Yes	

Notes: This table shows the effects on our main achievement outcomes, split by gender. Columns (1) and (3) display the ITT estimates of winning the subsidy, while (2) and (4) show the LATE estimates of using the voucher to buy a house. All estimates include year and age fixed effects and clustered standard errors at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Table 1.6: School Quality Outcomes

	Control Mean	ITT	ATE	Observations	Bandwidth
Dep. Variable	(1)	(2)	(3)	(4)	(5)
Ind. Public School	0.352	-0.000 (0.007)	0.002 (0.030)	523,481	88.615
Ind. Private	0.004	-0.001 (0.001)	-0.004 (0.005)	392,599	58.501
School Quality	0.445	0.004 (0.004)	0.015 (0.017)	471,465	75.319
Class Size	36.024	0.111 (0.097)	0.462 (0.457)	459,242	72.242
Ind. Change Comuna	0.012	-0.001 (0.001)	-0.005 (0.004)	447,901	76.160
Ind. Change School	0.117	-0.004 (0.003)	-0.019 (0.012)	452,356	70.710
Controls	-	Yes	Yes	-	-

Notes: This table shows the effects on our other primary and secondary education outcomes after application, focusing on school characteristics. Columns (1), (4), and (5) display the mean of the variable for non-awarded households, the number of applicants, and the optimal IMSE-bandwidth to perform the RD, respectively. Column (2) shows the ITT estimate of the regression discontinuity using Equation 1.1 by regressing each demographic characteristic on the score, corresponding to the jump at the cutoff. Column (3) presents the LATE estimates, rescaling the ITT estimates by the first stage in Equation 3.6 and presenting the effects of using the voucher to buy a house. All estimates include year and gender by age fixed effects and clustered standard errors at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Table 1.7: Labor Market outcomes for parents

Dep. Variable	Mean Control	ITT	ATE	Observations	Bandwidth
Employed	0.462	-0.009* (0.005)	-0.034* (0.019)	309,416	60
Months Worked	3.967	-0.009* (0.005)	-0.039* (0.022)	309,416	60
Months Worked - Permanent	2.664	-0.017** (0.007)	-0.060** (0.025)	309,416	60
Months Worked - Temporary	1.399	0.005 (0.010)	0.018 (0.036)	309,416	60
Yearly Earnings	63.200	-0.015** (0.007)	-0.061** (0.031)	309,416	60
Yearly Earnings - Permanent	45.770	-0.022** (0.009)	-0.080** (0.036)	309,416	60
Yearly Earnings - Temporary	17.430	-0.010 (0.016)	-0.045 (0.064)	309,416	60
Controls		Yes	Yes		

Notes: This table shows the differences in labor market outcomes for parents after application. It includes the applicant of the household and their partner, and we restrict the sample to households that have children in their schooling years. Columns (1), (4), and (5) display the mean of the variable for non-awarded households, the number of applicants, and the bandwidth selected to perform the RD, respectively. Column (2) shows the ITT estimate of the regression discontinuity using [Equation 1.1](#) by regressing each demographic characteristic on the score, corresponding to the jump at the cutoff. Column (3) presents the LATE estimates, rescaling the ITT estimates by the first stage in [Equation 3.6](#) and presenting the effects of using the voucher to buy a house. All estimates include year and gender by age fixed effects and clustered standard errors at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Table 1.8: Day-care and Preschool Attendance

	Control Mean	ITT	ATE	Observations	Bandwidth
Dep. Variable	(1)	(2)	(3)	(4)	(5)
In School (0/1)	0.484	0.022*** (0.008)	0.094** (0.045)	92,711	89.177
In School SC (<i>Sala Cuna</i>)	0.235	0.012 (0.013)	0.099 (0.128)	17,538	99.983
In School NM- (<i>Nivel Medio Inferior</i>)	0.404	0.046*** (0.015)	0.213** (0.087)	15,558	68.317
In School NM+ (<i>Nivel Medio Superior</i>)	0.428	0.024** (0.012)	0.086 (0.062)	23,205	76.791
In School PK (<i>Pre-Kinder</i>)	0.698	0.014 (0.010)	0.056 (0.050)	29,580	78.024
In School KN (<i>Kindergarten</i>)	0.824	-0.000 (0.007)	-0.001 (0.033)	37,897	91.673
Controls	-	Yes	Yes	-	-

Notes: This table shows the differences in daycare and preschool attendance after application. Columns (1), (4), and (5) display the mean of the variable for non-awarded households, the number of applicants, and the optimal IMSE-bandwidth to perform the RD, respectively. Column (2) shows the ITT estimate of the regression discontinuity using Equation 1.1 by regressing each demographic characteristic on the score, corresponding to the jump at the cutoff. Column (3) presents the LATE estimates, rescaling the ITT estimates by the first stage in Equation 3.6 and presenting the effects of using the voucher to buy a house. All estimates include year and gender by age fixed effects and clustered standard errors at the application level. ***, **, * indicates significance at 1, 5, and 10%. SC (+) corresponds to *Sala Cuna* (children aged 0-2), while NM- (*Nivel Medio Inferior*) and NM+ (*Nivel Medio superior*) are for children aged 3 and 4, respectively. PK (*Pre-Kinder*) and KN (*Kindergarten*) include children aged 5 and 6, respectively. In School (0/1) combines all children aged 1-5, only excluding Kindergarten enrollment, which is currently mandatory in Chile.

Table 1.9: High School and Post-Secondary outcomes

	Control Mean	ITT	ATE	Observations	Bandwidth
Dep. Variable	(1)	(2)	(3)	(4)	(5)
Ind. graduated HS	0.853	0.014** (0.006)	0.058* (0.030)	32,046	71.881
Grades HS	5.614	0.038*** (0.014)	0.169** (0.070)	21,157	60.681
Ind. PSU	0.829	0.007 (0.010)	0.029 (0.045)	25,790	66.421
Av. Score	-0.316	0.054** (0.022)	0.254** (0.109)	17,057	59.040
Score Math	-0.283	0.043** (0.021)	0.191* (0.099)	17,536	63.629
Score Verbal	-0.311	0.058** (0.023)	0.290** (0.114)	16,316	56.086
Score History	-0.302	0.054* (0.030)	0.224* (0.127)	13,420	82.606
Score Science	-0.301	0.062* (0.036)	0.335* (0.197)	9,736	59.801
Ind. college	0.592	0.018** (0.009)	0.083** (0.040)	27,579	73.233
Years of college	1.727	0.119*** (0.044)	0.530** (0.211)	27,136	71.422
Ind. dropout	0.354	0.019 (0.014)	0.091 (0.063)	17,125	77.025
Ind. graduated	0.444	-0.006 (0.024)	-0.018 (0.109)	7,008	61.470
Ind. University	0.398	0.031** (0.014)	0.143** (0.068)	17,729	82.961
Years Accredited	5.600	-0.062 (0.042)	-0.287 (0.201)	15,413	75.440

Notes: This table shows the differences in high school completion and post-secondary outcomes after application. Columns (1), (4), and (5) display the mean of the variable for non-awarded households, the number of applicants, and the optimal IMSE-bandwidth to perform the RD, respectively. Column (2) shows the ITT estimate of the regression discontinuity using Equation 1.1 by regressing each demographic characteristic on the score, corresponding to the jump at the cutoff. Column (3) presents the LATE estimates, rescaling the ITT estimates by the first stage in Equation 3.6 and presenting the effects of using the voucher to buy a house. All estimates include year and gender by age fixed effects and clustered standard errors at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Table 1.10: Labor Market Outcomes for Children Not in College

Dep. Variable	Mean Control	ITT	ATE	Observations	Bandwidth
Employed	0.536	0.012** (0.006)	0.049* (0.024)	32,251	60
Months Worked	4.696	0.029* (0.018)	0.121* (0.069)	32,251	60
Months Worked - Permanent	3.248	0.033 (0.037)	0.135 (0.151)	32,251	60
Months Worked - Temporary	1.449	0.026* (0.013)	0.110* (0.054)	32,251	60
Yearly Earnings	44.888	0.028 (0.027)	0.121 (0.114)	32,251	60
Yearly Earnings - Permanent	33.482	0.031 (0.050)	0.113 (0.210)	32,251	60
Yearly Earnings - Temporary	10.832	0.012 (0.019)	0.052 (0.092)	32,251	60
Controls		Yes	Yes		

Notes: This table shows the differences in labor market outcomes for children who were not currently enrolled in post-secondary education. Columns (1), (4), and (5) display the mean of the variable for non-awarded households, the number of applicants, and the bandwidth selected to perform the RD, respectively. Column (2) shows the ITT estimate of the regression discontinuity using Equation 1.1 by regressing each demographic characteristic on the score, corresponding to the jump at the cutoff. Column (3) presents the LATE estimates, rescaling the ITT estimates by the first stage in Equation 3.6 and presenting the effects of using the voucher to buy a house. All estimates include year and gender by age fixed effects and clustered standard errors at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Chapter 2

The Effects of Widespread Online Education on Market Structure and Enrollment

2.1 Introduction

Over the past two decades, rapid advances in digital technology have transformed numerous industries, expanding their offerings to include the online delivery of goods and services. Sectors traditionally reliant on physical, face-to-face interactions—such as healthcare and education—are increasingly adopting hybrid or fully virtual models of operation. This shift has been particularly pronounced in higher education, where colleges have increased their online offerings to meet the growing demand for more affordable and flexible learning (Deming et al., 2012; Aucejo et al., 2024). In 2019, around 15% of all U.S. undergraduate students were enrolled exclusively in distance education, a figure that rose to 24% by 2022, accelerated by the COVID-19 pandemic (NCES, 2022). Expanding education—or any service—to online formats presents both opportunities and risks. On the one hand, online education can increase access for previously excluded students, democratizing access to higher education (Barrow et al., 2024). On the other hand, the shift to remote delivery can fundamentally change the nature of these services, potentially compromising their quality (Bettinger et al., 2017; Garrett et al., 2022). As a result, online education may boost overall enrollment in higher education but also risks diverting students from high-quality, in-person instruction to potentially inferior online alternatives. This is particularly problematic in higher education, where quality is difficult to assess.

These concerns are amplified when considering the equilibrium effects of online education. The entry of online programs can increase the competitive pressure on traditional in-person programs (Deming et al., 2016a), initially benefiting students through lower tuition costs (Deming et al., 2015). However, because in-person programs operate with substantial fixed

costs, sustained declines in enrollment and revenue can lead to program closures, ultimately reducing the availability of high-quality educational options. This risk is especially problematic in markets with limited in-person institutions, where students may be left with lower-quality online alternatives as their only choice.

This paper examines the impact of online education expansion in Brazil, the world's largest market for online higher education. We leverage plausibly exogenous variation in the differential entry of online programs across regions and fields of study to assess how the introduction of online undergraduate degrees impacts market expansion, student diversion, and in-person program availability. We integrate this analysis with an equilibrium model of college education to quantify the overall effects of online education and evaluate alternative policy designs.

Brazil provides an ideal setting for studying the expansion of online education for two reasons. First, analyzing equilibrium effects requires a setting where online education constitutes a significant share of the education market. In Brazil, fully remote programs have grown rapidly, comprising 17% of all new undergraduate enrollments in 2010, 44% by 2019, and 65% post-COVID. Second, to isolate the causal impact of online education on local market outcomes from nationwide time trends, we need variation in online penetration across local markets. Brazilian regulation provides this variation by requiring online programs to establish local physical hubs, which students must attend periodically, generating geographic differences in online market entry. Additionally, by prohibiting online education in specific majors, the regulation introduces further variation across fields of study.

For our analysis, we use several administrative datasets. First, we draw on detailed data from the Brazilian Higher Education Census to assess market shares and track the entry and exit of degree programs. We also collect tuition fee data from various sources to examine colleges' pricing strategies. Additionally, we combine university entrance exam data with the Higher Education Census and matched employer-employee records to estimate the labor market returns of specific degrees, allowing us to compute the value added of online and in-person programs. We thus have a rich window into student demand for in-person and online degrees, colleges' behavior, and overall degree quality.

Our analysis focuses on the private sector, which accounts for 82% of incoming students between 2010 and 2019, and nearly all online programs. During this period, private sector enrollment grew from 1.7 million to 3 million students, with 89% of this growth driven by online programs, particularly at for-profit institutions. Two main factors have characterized the growth of online education. First, it has improved access for older, lower-income adults seeking affordable and flexible learning options. While the average incoming in-person student is 24 years old, the average online student is 29 years old. Second, there has been a shift away from traditional in-person programs. Since 2010, in-person enrollment has stagnated, with significant declines in business and education programs—fields that have seen the largest increases in online enrollment.

We begin our analysis by comparing online degree programs to traditional in-person programs in terms of duration, tuition fees, dropout rates, and value added. To do this, we

examine equivalent programs offered by the same institutions, differing only in delivery mode. Our findings show no significant differences in program duration, consistent with regulations mandating that both formats follow the same curriculum. We also document that tuition fees for online degrees are 57% lower than those for in-person programs. Regarding quality, online programs show lower dropout rates but are associated with a 38% lower value added, measured by students' gains in labor market outcomes after accounting for their entrance exam scores.

To examine the causal effects of online program expansion in Brazil, we use a linear model that regresses changes in outcomes—such as student enrollment, market structure, and tuition—on the change in the number of online degrees from 2010 to 2019. We define our unit of analysis as the interaction of a commuting zone and a field of study. We begin by estimating the model using OLS. The validity of this approach relies on a parallel trends assumption, which requires that outcomes in regions and fields with lower online degree entry would have followed the same trend as those in higher online degree entry areas had they experienced similar online degree entry. To address potential bias from unobserved shocks, we also implement a shift-share instrumental variable (SSIV) approach (Bartik, 1991b; Goldsmith-Pinkham et al., 2020), combining predetermined institution headquarter locations (shares) with online sector growth (shift). Our instrument builds on the observation that institutions tend to expand their online programs in regions closer to their headquarters and assumes that a region's proximity to an institution is not correlated with unobserved region-specific shocks in fields where the institution offers in-person degrees (i.e., exogenous shares).

Both the OLS and SSIV approaches yield qualitatively consistent results: expanding online education increases online enrollment, reduces in-person enrollment, and raises overall college enrollment. This highlights the dual effect of online education. On one hand, it expands access for students who might not otherwise have attended college. On the other hand, it pulls students away from in-person programs, often toward lower-quality online alternatives. This shift heightens competition, forcing local in-person institutions to lower prices and reduce profits. As competition intensifies, in-person programs are less likely to persist in the market, accelerating the shift toward online degrees. As a result, the value added attained by the average student declines. We then present an analysis for different age cohorts, showing that this expansion primarily benefits older students by increasing their college enrollment and value added. Younger students, who are more likely to enroll in in-person programs initially, experience a stronger diversion to online options and a reduction in value added.

Estimating the linear model provides a transparent method for recovering marginal causal effects, but it rests on a strong no-interference assumption, which requires that changes in the number of online degrees in a given region and field only affect outcomes of that specific field of study. This assumption is violated if degrees across fields are substitutes. Moreover, the linear structure may not capture out-of-sample counterfactuals effectively, where competition and supply-side responses, such as price changes and entry or exit decisions, can lead to

significant non-linear effects. To address these limitations, we develop a supply and demand model for college education that incorporates rich substitution patterns and accounts for equilibrium responses, allowing us to assess the impact of online education expansion under different counterfactual scenarios.

Our equilibrium model consists of students and educational institutions. On the demand side, students decide whether to attend college and in which degree to enroll. On the supply side, institutions decide whether to enter a particular market, which degrees to offer, and what prices to charge. To operate in a market, institutions must establish either a campus for in-person degrees or a hub for online degrees. Their decision-making occurs in two stages. In the first stage, institutions simultaneously choose in which regions to operate and which degrees to offer, after observing their fixed entry costs and taking into account pre-existing offerings (Seim, 2006; Atal et al., 2025). They form expectations about their competitors' entry decisions and adjust their strategies accordingly, potentially opting out of opening an in-person campus to avoid fixed costs if they anticipate competitors will expand their online offerings. In the second stage, institutions compete on prices.

We estimate demand by leveraging different variations in the data. To estimate substitution between degrees in different fields of study and delivery modes, we use market-level differences in degree availability induced by the shift-share instrument. Our estimates suggest that in-person and online degrees are close substitutes: when an in-person degree closes, 67% of students switch to another in-person option, 18% move online, and 14% opt out of college. To estimate price elasticities, we use contemporaneous prices of the same degree in other regions as proxies for cost-shifters. These leave-one-out mean prices serve as an instrument for prices (Hausman et al., 1994). The mechanics behind this instrument are that firm-level cost variations influence prices in all markets where a degree is offered, and the key identification assumption is that, although costs for degrees from the same institution are correlated across markets, demand shocks are not. We estimate median own-price elasticities of approximately -2.9 for in-person degrees and -1.1 for online degrees, which aligns with findings from the literature (Dobbin et al., 2021; Armona and Cao, 2024).

For the supply model, we recover entry elasticities with respect to profits using two instruments that affect profits but do not influence institutions' fixed costs. First, we exploit regional variation over time in internet penetration as a demand shifter for online education. Second, we use differences in competitors' distance to various regions, which creates varying levels of competition that impact profits without affecting institutions' fixed costs. We estimate a median entry elasticity with respect to own profits of 1.2 and find that in-person campuses are 2.3 times more expensive to open than online hubs.

We use our estimated model to quantify the impact of online education expansion on student enrollment, market structure, tuition fees, value added, and consumers' expenditure and surplus. To examine how supply-side equilibrium effects shape these outcomes, we simulate three progressively more flexible counterfactuals where we remove the supply of online education. We benchmark each of these scenarios against a baseline counterfactual that reflects the status quo, where online education exists.

In the first counterfactual scenario, we examine the effects of removing online education without accounting for supply-side responses (i.e., fixing degree offerings and prices). Under this scenario, two-thirds of students enrolled in online programs would switch to in-person programs, while the remaining third would exit the market. This shift results in a 14% decline in total enrollment. Because online degrees provide greater value-added than the outside option, the loss of students leaving the market outweighs the gains from those switching to in-person programs, leading to a net decline in total value-added. However, given the significant tuition gap between online and in-person programs, the increase in in-person enrollment raises total expenditures on higher education.

Under the second counterfactual, we allow institutions to respond by adjusting the prices of their degrees. As a result, tuition fees for in-person degrees increase by 13.6%, causing a further reduction in enrollment and total value added, and increased expenditure.

Finally, under the third counterfactual, we allow institutions to adjust both tuition fees and degree offerings. Relative to the previous scenario, the supply of in-person degrees under this scenario expands by 17.3%, attracting new students and increasing total enrollment by 4.3%. The total value added under this counterfactual is 3% higher than under the first counterfactual, underscoring the importance of accounting for equilibrium responses. Compared to the status quo—where online education is available—total value added in the absence of online education is 3.4% higher. However, a revealed-preference analysis assuming full rationality shows that consumer surplus declines by 11.9%. This suggests that students may strongly prefer more affordable and flexible options, even if they offer lower value added.

We then examine the distributional consequences of online education expansion by evaluating its effect on value added across different student age cohorts. Under the first counterfactual, which removes online education while ignoring supply-side responses, no cohort experiences significant value-added gains, and older students (ages 35-45) see a decline in value added. When supply-side responses are fully accounted for under the third counterfactual, younger students benefit from increased access to in-person programs that were previously limited due to the presence of online options. However, older students, who have strong preferences for online degrees, are worse off and tend to exit the market.

Our findings highlight the uneven effects of online education. While its expansion broadens access for older students, increasing their opportunities for higher education, it reduces in-person options, pushing younger students into lower-quality alternatives. Using this insight, we explore potential government policies aimed at aligning online education with those who benefit most. Specifically, we examine a policy where online education is available only to students above the age of 25. Our results show that this approach would increase value added for all cohorts compared to the baseline scenario, where online education is available to all age groups.

Our findings shed light on how disruptive technologies like online services can reshape competition and market dynamics. While they expand choice and lower costs, benefits may be unevenly distributed. In markets with imperfect information, some consumers may unknowingly switch from higher-quality options, leaving them worse off—especially as de-

clining demand erodes traditional alternatives. To mitigate these risks, policymakers could ensure access to established options for affected groups while allowing others to adopt new technology, balancing innovation with quality preservation.

This paper adds to the growing literature studying the effects of introducing more accessible, lower-quality options in educational markets. Research on community colleges shows they democratize access to higher education but also divert some students from four-year institutions (Rouse, 1995, 1998; Mountjoy, 2022). Likewise, studies on online education underscore its potential to expand access or draw students away from higher-quality alternatives (Deming et al., 2012; Goodman et al., 2019). Additional research also emphasizes the competitive pressure online degrees exert on tuition prices of traditional in-person programs (Deming et al., 2015). Our paper advances this line of work by developing an equilibrium framework that accounts for market expansion, market diversion, price changes, and endogenous degree offerings.

Our research is also related to the literature that examines the effects of online forms of education on student learning and academic progression (Figlio et al., 2013; Bettinger et al., 2017; Kofoed et al., 2024), and on labor market outcomes (Deming et al., 2016b; Hoxby, 2018; Fabregas and Navarro-Sola, 2024). Our findings are consistent with this literature, highlighting the role of online education as a preferable alternative to no education, albeit less favorable than in-person options. Furthermore, our study contributes to the expanding literature on the market effects of online services beyond education, particularly telemedicine. Zeltzer et al. (2023) show that telemedicine can reduce overall healthcare spending without compromising diagnostic accuracy or outcomes. Additionally, evidence suggests that online healthcare services can enhance efficiency by offering faster and shorter consultations and improving the matching of doctors and patients (Dahlstrand et al., 2024; Dahlstrand, 2024).

We also contribute to the broader literature that analyzes education policy using equilibrium models of imperfect competition. These models have been used to study the effects of educational policy on pricing and quality in secondary schools (Neilson et al., 2013; Allende, 2019) and colleges (Dobbin et al., 2021; Barahona et al., 2023a; Armona and Cao, 2024), on instructional levels (Bau, 2022), and on institutions' decisions to participate in voucher programs (Sanchez, 2023). More closely related to our work, three papers assess the impact of competition on market structure. All three papers focus on the impact of improved public sector offerings on private institutions' entry and exit decisions. Bodéré (2023) explores the effects of higher-quality public preschools in Pennsylvania, while Dinerstein and Smith (2021) assesses the impact of increased public school funding in New York. Similarly, Dinerstein et al. (2023) investigates the expansion of public schools in the Dominican Republic. Our contribution to this literature is to provide evidence of increased competition driven by the private sector by introducing a new delivery format into the market.

The paper is organized as follows. Section 2.2 provides institutional background on Brazil's higher education sector, discusses the data, and presents descriptive statistics on the growth of online education. Section 2.3 presents the results from the linear model used to estimate the causal effects of expanding online education on various outcomes. In Section

2.4, we introduce and estimate the equilibrium model, and in Section 2.5, we use it for counterfactual analysis. Finally, Section 2.6 concludes the paper.

2.2 Setting and data

In this section, we describe the regulatory landscape of Brazilian higher education and online education. We then describe the data sources and provide several descriptive statistics about the expansion of online education.

Online higher education landscape in Brazil

Brazil’s higher education system has experienced significant expansion over the past decade, with new undergraduate enrollment growing from approximately 2.2 million students in 2010 to 3.6 million in 2022. A crucial factor underlying this growth has been the rising popularity of online degree programs, predominantly offered by for-profit private institutions.¹ The shift towards online education has been dramatic: in 2010, 17% of new students chose online programs, rising to 44% by 2019, and soaring to 65% after the COVID-19 pandemic.²

Online degree programs in Brazil, referred to as “Educação a Distância”, offer remote versions of traditional in-person undergraduate diplomas and are required to adhere to the same curriculum and duration standards.³ Diplomas make no distinction between whether a degree was earned online or in person, theoretically placing both modes on equal footing. However, despite their lower cost, online programs are often perceived as being of inferior quality, a concern that continues to challenge Brazilian policymakers (Bertolin et al., 2023).

Online programs are required to include in-person sessions for essential activities like assessments and laboratory work, which must be conducted either at the institution’s main campus or at designated local hubs. These hubs are decentralized centers created to support the face-to-face elements of online education, placing geographic limits on the reach of these programs. Despite the in-person requirements, all instruction remains fully remote. Most institutions provide live, synchronous classes to facilitate real-time interaction between students and instructors. In addition, 78% of institutions offer asynchronous resources—such as pre-recorded lectures, reading materials, and interactive quizzes—giving students more flexibility to engage with course content on their own schedule (ABED, 2018).

¹Private institutions make up 95% of the total online education market, with 79% of that share held by for-profit institutions. Furthermore, this market is heavily concentrated, with seven institutions dominating 73% of the entire online education market.

²Brazil’s share of fully remote undergraduate enrollments is significantly high compared to other countries. Although reliable cross-country data on fully online degrees is generally scarce, official data from 2019, prior to the COVID-19 pandemic, indicates fully remote enrollment shares of approximately 13% in Australia, 14% in India, 17% in Mexico, 8% in the United Kingdom, and 15% in the United States.

³Students have the option to transfer credits between online and in-person formats within the same institution if they change modalities.

Several key factors have driven the growth of online education. First, there's growing demand stemming from the flexibility these programs offer, enabling more students—particularly an older demographic—to pursue higher education while balancing other responsibilities (El Galad et al., 2024). In 2019, 71% of new online students were over the age of 24, compared to just 32% in traditional in-person programs. Second, the widespread improvement in internet infrastructure across Brazil has significantly facilitated access to online degrees, even in previously underserved regions. In 2010, just 40% of Brazilian households had internet access (IBGE, 2010), but by 2019, this figure surged to 83% (PNAD, 2019). Lastly, government reforms introduced in 2016 have streamlined the accreditation process for new online programs and granted institutions greater autonomy to establish new hubs. This regulatory evolution has made it considerably easier for educational institutions to manage, expand, and diversify their online offerings, contributing to the sector's overall growth.⁴

Some fields of study face restrictions on being offered online. Specifically, Law, Medicine, and Psychology require special authorization from regulatory bodies such as the National Bar Association and the National Health Council. To date, no online programs in these fields have been approved. In contrast, disciplines like Business and Education have seen significant growth in online education. In 2019, these two fields accounted for 77% of total online enrollment, in contrast to their 31% share in in-person programs.

Data

Higher education census

This dataset encompasses several layers of information. First, it includes institution-level details, such as ownership status and the parent firm. Second, it captures program characteristics, including detailed categories of the field of study, the delivery mode (online or in-person), the required number of hours for graduation, and the year the degree was introduced. Additionally, it provides information on the hubs associated with each online program. Third, it contains student-level data for all enrolled students, along with their demographics, enabling us to track each student's educational path.

Tuition fees

Universities are not required to report tuition fees to the supervising authority, so we rely on four distinct data sources to gather this information. The first two sources come from Brazil's government fellowship and loan programs, PROUNI and FIES. We utilize administrative records from the National Education Fund (FNDE), which track the payments made by the government to students in these programs, allowing us to estimate the tuition fees at

⁴See [Resolução CNE/CES N1, de 11 de março de 2016](#), [Decreto 9.057, de 25 de maio de 2017](#), and [Portaria Normativa 11, de 20 de junho de 2017](#).

participating institutions. The third source is a nationally representative survey conducted by Hoper, a consultancy specializing in higher education. The fourth source is administrative data from QueroBolsa, Brazil’s largest degree search platform. In Appendix B.2, we outline the methodology used to combine these sources into a unified tuition price for each degree program. As a result, we can recover year-specific tuition prices for approximately 95.5% of degree program-years, covering 98.5% of total enrollment.

Test scores

We have access to detailed data for all students who took ENEM, Brazil’s university entrance exam. This standardized test is high stakes, as it determines eligibility for financial aid and is used for admissions by several public universities. The data include scores for each section of the exam, along with responses to a comprehensive socioeconomic background questionnaire. College applicants who take the entrance exam vary significantly in age; for example, in 2010, about 62% had finished high school more than a year before taking the test, and 25% were 25 or older.

Matched employer-employee records

Finally, we integrate the previously mentioned data sources with matched employer-employee annual administrative records (RAIS) from the Ministry of Labor. This dataset includes detailed worker and firm-level variables, such as salaries, contracted hours, hiring and firing dates, and occupation of Brazil’s formal labor market. We use these data to construct earning profiles for each program and student, covering both online and in-person formats. These profiles allow us to calculate value-added measures for each program in the system.

Additional auxiliary sources

We use data from the 2010 Brazilian Population Census to estimate the size of various age cohorts within each region, allowing us to define the potential market for higher education students. We also use administrative data from DSCOM (“Dados do Setor de Comunicações”) from the Ministry of Science and Technology to calculate internet penetration rates across different regions over time.

Data definitions

Next, we provide several definitions for the units of analysis and samples that we use throughout the article.

Regions and markets

A central aspect of our analysis is defining regions that allow us to segment educational markets. Brazil consists of 5,568 municipalities, which we group into 137 meso regions—an administrative division from the Brazilian National Bureau of Statistics (IBGE) that clusters municipalities based on proximity and common features. These meso regions serve as our definition of local markets, which we refer to as “regions” throughout the paper. We define a “market” as the intersection of a region and a year. The market size is determined by the number of 18- to 45-year-old residents without a college degree living in the region for that year.

Firms

We define a firm as a company that may own multiple universities, and we use the terms “firms” and “institutions” interchangeably. In our data, we observe two types of firms: those that are expanding their services into other regions and others that operate in the same locations every year. We refer to the former as *expanding institutions* and the latter as *local institutions*. Most local institutions operate in a single location, with a few of them operating in two to five different locations.

Degrees

Throughout this draft, we use two different definitions of “degrees.” The first is *degree programs*, which are distinct undergraduate programs sharing the same administrative code. These programs are offered by the same university, have identical curricula, and are delivered in the same format—either in person or online. For in-person degree programs, the only variation may be the schedule, such as morning, evening, day, or full-time. For online degree programs, those sharing the same administrative code may be associated with different physical hubs, allowing them to be available in multiple regions. The dataset includes 35,527 unique degree programs.

To reduce the data’s dimensionality, we introduce a second definition of degrees by aggregating similar degree programs within the same institution. We define a *degree* as the combination of all degree programs offered by the same institution, within the same field of study, and delivered in the same format.⁵ For example, in our dataset, a degree might include all degree programs in the field of “Business” (such as Administration, Accounting, Marketing, and Economics), offered by Anhanguera Educacional, a for-profit educational company in Brazil, and taught in person. A degree can be offered in several regions and years.

⁵In total, we have 11 fields of study based on the International Standard Classification of Education (ISCED) codes. Details on the categorization are provided in Online Appendix B.1, Table B1.

Sample

Our analysis centers on the private sector, which accounts for nearly all online programs and roughly 82% of total enrollment, including both online and in-person students. This sector is predominantly non-selective and is often perceived as lower in quality compared to the public sector, which is higher quality and highly selective ([Barahona et al., 2023b](#)). Consequently, there is limited substitution between the two sectors. Public institutions, which rely entirely on government funding—whether federal or state—are also less influenced by market forces.

We also limit our analysis to the years 2010, when microdata first became available, and 2019, the final year before the onset of the COVID-19 pandemic. To avoid including very small markets in our analysis, we apply the following restrictions: we exclude 27 regions with fewer than 50,000 residents aged 18–45, institutions with fewer than 500 students nationwide in any given year, and degree-region pairs with fewer than five students in any year. After these exclusions, our final sample consists of 110 regions over a 10-year period, covering 474 institutions, 4,119 unique degrees, 15,206 degree-region pairs, and 92,625 degree-region-years. 340 institutions are local and 134 are expanding, out of which 93 expanded online.

Value added

An important component of our analysis is computing a quality measure for *degree* programs. We approximate program quality using value added measures derived from data on student enrollment, test scores, a comprehensive set of covariates, and administrative wage records. To calculate value added, we track all students taking the university entrance exam (ENEM), assigning them to specific degree programs based on their initial college enrollment or to an outside option if they do not enroll. We then follow these individuals in the labor market, recording their highest salary reported in the administrative wage records (RAIS) several years post-exam.

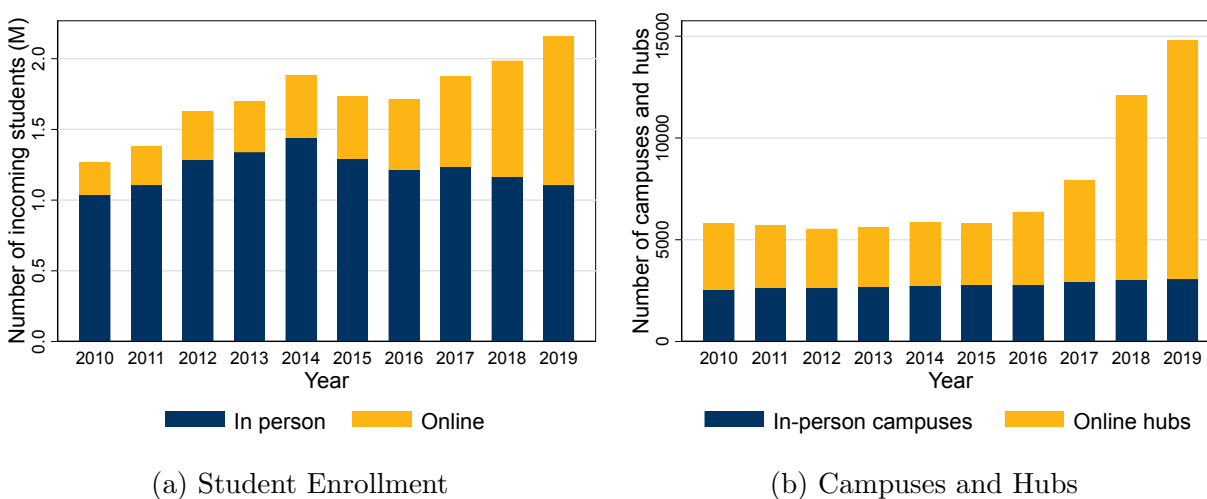
Our approach employs a standard selection-on-observables model ([Rothstein, 2010](#); [Angrist et al., 2017](#); [Ainsworth et al., 2023](#)), which compares labor market earnings across degree programs while controlling for test scores and a wide range of student characteristics. Our methodology for calculating value added, outlined in Appendix B.2, incorporates three key features. First, we normalize all measures relative to the outside option of not attending college. Because we care about students’ outside options in their specific local markets, we calculate value added separately by region. As a result, degree programs’ value added can differ across regions if local market conditions favor some programs over others or if the outside option of not attending college is different. Second, we analyze value-added outcomes using 2010 test takers and follow them seven years post-application (using ENEM 2010 to RAIS 2017 data), providing a stable, time-invariant measure. Finally, following [Angrist et al. \(2023\)](#), we use empirical Bayes methods to improve the precision of estimates for smaller programs.

Descriptives

Trends in the Brazilian private-sector higher education sector

We present the trends of the overall expansion of online education from 2010 to 2019 in Figure 2.1. Panel 2.1 shows the enrollment trends of incoming students for both online and in-person education. Between 2010 and 2014, college enrollment steadily increased, with the share of online education remaining relatively stable, representing approximately 23% of total enrollment by 2014. However, following the policy reforms around 2016 that liberalized access, described in Section 2.2, a notable shift occurred: online education saw significant growth while in-person enrollment began to decline. By 2019, over half of incoming students at private universities were enrolled in online education. Panel 2.1 illustrates similar trends in the number of in-person campuses and online hubs in Brazil. Between 2010 and 2015, the number of campuses and hubs remained comparable, with a marked rise in the number of online hubs following the policy reforms.

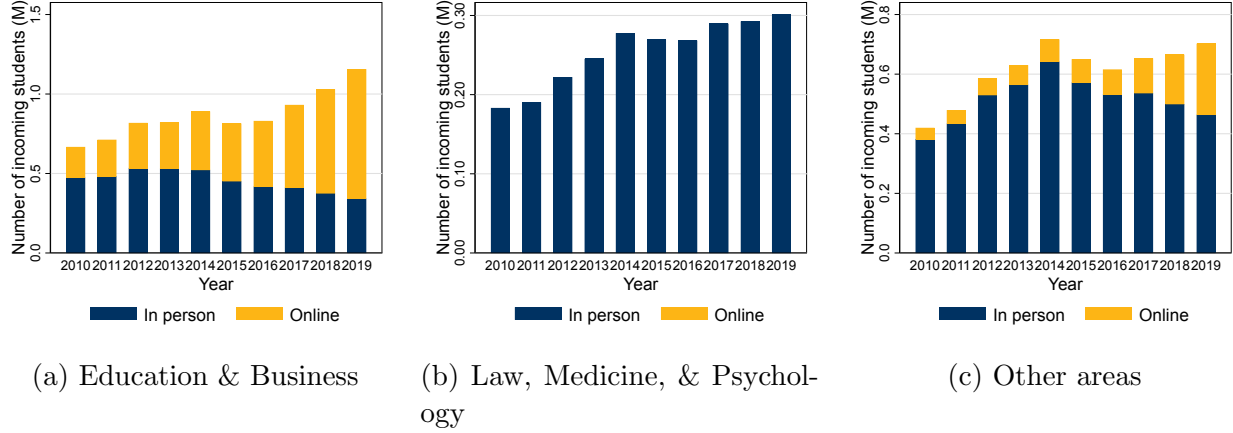
Figure 2.1: Expansion of online education



Notes: This figure presents the Brazilian private-sector higher-education sector trends. Panel (a) shows the number of incoming students in private institutions for in-person (blue) and online (yellow) degrees across years. Panel (b) shows the number of open in-person campuses (blue) and online hubs (yellow) across all regions in Brazil.

The expansion of the online education sector varied across fields of study. As shown in Figure 2.2, the most substantial growth occurred in programs related to Business and Education, partly due to their suitability for online delivery. In contrast, fields such as Law, Medicine, and Psychology saw no comparable expansion, as legal restrictions prevent institutions from offering these programs online. Moreover, in-person enrollment declined in fields that experienced substantial online growth but continued to rise in those where online education is restricted.

Figure 2.2: Expansion of online education enrollment by categories of field of study



Notes: This figure presents enrollment trends of the Brazilian private-sector higher-education sector across fields of study. Panel (a) shows the number of incoming students in private institutions for in-person (blue) and online (yellow) degrees in Education and Business, Panel (b) for degrees in Law, Medicine, and Psychology, and Panel (c) for degrees in other fields of study.

Finally, we highlight that online degree programs are especially popular among older students. There is a noticeable difference in the age distribution of students enrolled in online versus in-person programs. In Figure 2.3, we show online and in-person enrollment segmented by age group for 2010 and 2019.⁶ While in-person programs are dominated by younger students, online programs attract a majority of students aged 26 and older. By 2019, around 64% of online students in these fields were aged 26 or older, compared to just 28% in in-person programs.

Comparison between online and in-person programs

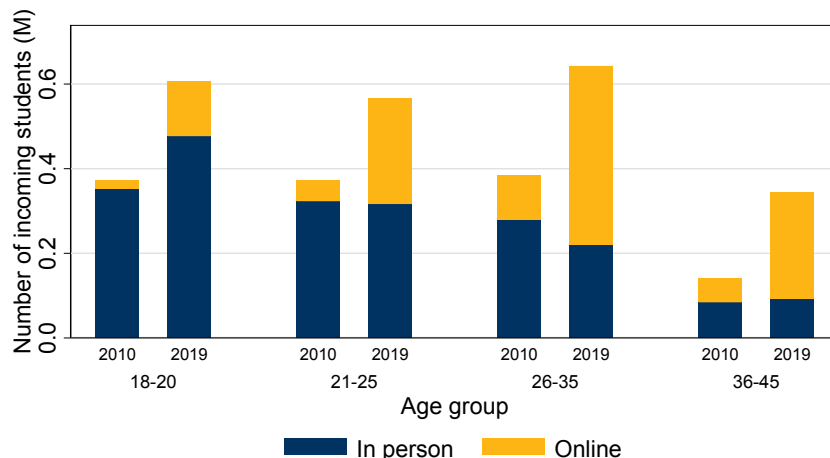
To examine the practical differences between online and in-person programs, we compare equivalent *degree* programs offered by the same university, differing only in their mode of delivery. We estimate the following regression:

$$Y_{jrt} = \beta \cdot o_j + \delta_{m(j)t} + \delta_r + \varepsilon_j, \quad (2.1)$$

where Y_{jrt} is an outcome in degree program j in region r in year t , $o_j \in \{0, 1\}$ indexes whether the program is online, $\delta_{m(j)t}$ is a university-major and year-specific fixed effect (e.g., Economics at the Universidade Norte do Paraná in 2011), and δ_r is a region fixed effect. We

⁶Throughout our analysis, we split students into four age groups: 18-20, 21-25, 26-35, and 36-45. In 2010, each of the first three groups represented 30% of total college enrollment, and the last group represented the remaining 10%. Regarding the total population, the first group represents 12%, the second 24%, the third 38% and the fourth 24%.

Figure 2.3: Expansion of online education enrollment by age group



Notes: This figure presents enrollment levels across age groups in the Brazilian private-sector higher-education sector in 2010 and 2019. For each age group, the left and right bars show enrollment in private institutions for in-person (blue) and online (yellow) degrees in 2010 and 2019, respectively.

examine four primary outcomes: (1) the log of the number of hours required to complete the program, (2) the log of the tuition price, (3) first-year dropout rates, and (4) value added as defined in Section 2.2. Since our value-added measure does not vary over time, we estimate the corresponding regression using a cross-section of the data for $t = 2011$ (corresponding to 2010 ENEM takers). All regressions are weighted by the number of students enrolled in program j .

The results are presented in Table 2.1. Column (1) reports the findings for the log of the required number of hours to complete the degree, showing that both online and in-person programs require a similar number of hours, which is expected as it is mandated by law. In column (2), we observe that online programs are 0.6 log-points less expensive than their in-person counterparts. Column (3) shows that online programs have dropout rates that are 0.021 percentage points lower (over a base of 0.20) than their in-person counterparts. Lastly, column (4) reveals that the value added of online programs is, on average, 0.057 log-points lower than that of in-person programs (over a base of 0.14).

These findings shed light on the potential trade-offs between in-person and online education. Online education, for example, can be offered at a significantly lower cost, making tuition more affordable. Its flexibility and reduced fees may contribute to slightly higher first-year persistence rates. However, it generally provides a lower value added compared to in-person programs. Nonetheless, both online and in-person degrees, on average, deliver higher value added than the alternative of not attending college (see Online Appendix B.1, Figure B1).

Table 2.1: Comparison between online and in-person programs

	log total hours (1)	log prices (2)	dropout rate (3)	value added (4)
Online	-0.003 (0.005)	-0.573 (0.030)	-0.021 (0.011)	-0.057 (0.007)
Obs.	544,920	549,704	549,704	17,540
Mean dep var. (levels)	3220	4430	0.20	0.14
Region FE	Yes	Yes	Yes	Yes
University-major-year FE	Yes	Yes	Yes	Yes

Notes: This table reports the coefficient of a linear regression of each outcome against an indicator variable for whether the degree is online or not, reported in Equation (2.1). All regressions are weighted by the degrees' number of students and control for region fixed effects and university $\times$ program name $\times$ year fixed effects. Column (1) uses the log-number of hours to complete the program as an outcome, Column (2) uses the log of the tuition price, Column (3) shows first-year dropout rates, and Column (4) uses value added. To calculate value added, we run a regression of log income on a rich set of student characteristics and degree fixed effects. We use the estimate of the degree fixed effect as the measure of value added (see Section 2.2). Columns (1), (2), and (3) use all years. Column (4) uses only 2011, which is the year for which we estimate value added.

2.3 The Effects of Online Entry on Market Outcomes

In this section, we estimate the effects of introducing an additional online degree on various market outcomes using a linear model. To achieve this, we compare changes in outcomes between 2010 and 2019 across regions and fields of study with varying levels of exposure to the growth of online degrees. Throughout this section, we use the “degree” definition as outlined in Section 2.2.

Specifically, we estimate the following structural equation:

$$\Delta y_{ra} = \phi \Delta N_{ra}^o + \varepsilon_{ra}, \quad (2.2)$$

where ΔN_{ra}^o denotes the change in the number of online degrees offered in region r and field of study a between 2010 and 2019, and Δy_{ra} represents the change during the same period in the following outcomes of interest: (i) the number of online students relative to market size, (ii) the number of in-person students relative to market size, (iii) the total number of in-person degrees, (iv) and the average price of in-person degrees. The error term ε_{ra} captures unobserved shocks at the region-field level that influence the trend of y_{ra} . The coefficient ϕ is the key parameter of interest.⁷

⁷Equation (2.2) can be derived by taking differences between 2019 and 2010 of the following structural equation: $y_{rat} = \phi N_{rat}^o + \delta_{ra} + \delta_t + \varepsilon_{rat}$, where y_{rat} and N_{rat}^o are the outcome of interest and the number of online degrees in region r , field of study a , and year t , and δ_{ra} , δ_t , and ε_{rat} are region-field of study fixed effects, year fixed effects, and region-field-year specific shocks.

Estimating a linear model presents both advantages and limitations. On the one hand, it offers simplicity and transparency. On the other hand, it relies on strong assumptions. In particular, it imposes a no-interference assumption implying that changes in the number of online degrees offered in region r and field of study a only impact the outcomes in that particular region and field of study, without influencing outcomes in other regions or fields—a condition known as the Stable Unit Treatment Value Assumption (SUTVA). This assumption may be violated if degrees across different fields are close substitutes. In Section 2.4, we extend our model to account for potential market-level interactions.

We propose two alternative estimation strategies for the linear model, each based on different assumptions, identifying the parameter of interest. First, we outline the assumptions required for a causal interpretation of ϕ when estimating Equation (2.2) using OLS. Next, we introduce a shift-share instrumental variable approach to address potential threats to the identification in the OLS regression. Finally, we present and compare the results from both estimation strategies.

Ordinary Least-Squares Regression

The parameters estimated through OLS can be interpreted as causal under the conditional independence assumption $\mathbb{E}[\varepsilon_{ra} | \Delta N_{ra}^o] = 0$. Since Equation (2.2) is written in differences, this assumption is implied by a parallel trends assumption, which states that the trajectory of outcomes for regions and fields of study with lower online degree growth must represent the outcomes in higher-growth regions and fields had they experienced lower online growth (Callaway et al., 2024). Unfortunately, standard tests for this assumption—based on evaluations of pre-trends—are not feasible, as online education was already widespread and growing at the start of our sample. To address this concern, we implement a shift-share instrumental variable approach.

Shift-share instrumental variables

We tackle the issue of endogeneity in institutions’ entry decisions by exploiting quasi-random variation in regions’ exposure to online entry. Following Goldsmith-Pinkham et al. (2020), we implement a shift-share instrumental variables (SSIV) design—or “Bartik instruments”—with exogenous shares.⁸ Our shares variable is constructed from a combination of three factors: (i) differences in regions’ exposure to potential entrants based on the distance between the regions and institutions’ headquarters (ii) institutions’ specialization across different fields of study based on their 2010 in-person offerings, and (iii) an indicator variable that

⁸Goldsmith-Pinkham et al. (2020) implement SSIV under the assumption that the shares are exogenous. In contrast, Borusyak et al. (2021) and Borusyak and Hull (2023) provide a framework for SSIV where identification is achieved through an exogenous shift.

captures whether online education is allowed in a given field of study.⁹ For the shift, we use the rapid expansion of online degree programs, driven by a combination of factors, including growing demand, advancements in online technology, and the policy changes mentioned in Section 2.2.

Building on this, we define the following shift-share instrument:

$$z_{ra} = \sum_f \underbrace{\Delta N_f^o}_{\text{Shift}} \cdot \underbrace{(z_{fr} z_{fa} z_a)}_{\text{Share}}, \quad (2.3)$$

where ΔN_f^o represents the shift, capturing the total number of online degrees introduced by institution f between 2010 and 2019. We allow this shift to be correlated with the distribution of shocks ε_{ra} . The variable $z_{fr} z_{fa} z_a$ corresponds to the exogenous shares that predict the region r and field of study a where institution f is likely to expand. As described above, these shares are derived from three sources: (i) exposure to potential entrants, z_{fr} , (ii) the institutions' propensity to expand in different fields of study, z_{fa} , and (iii) regulatory constraints within fields of study, z_a . We begin by detailing the shift, followed by an explanation of each component of the share.

Figure 2.4 provides a visualization of each component. In each panel, we order institutions by the location of their headquarters from northwest to southeast, based on the IBGE encoding system. We explain each panel in detail as we introduce the different components of Equation (2.3).

Institutions' expansion of online degrees (ΔN_f^o)

The first component of the instrument in Equation (2.3) is the shift term. We estimate ΔN_f^o by calculating the total number of online degrees that each institution opened in any region or field of study between 2010 and 2019:

$$\Delta N_f^o = \sum_r \sum_a (N_{fra,2019}^o - N_{fra,2010}^o), \quad (2.4)$$

where $N_{fra,t}^o$ takes the value of 1 if institution f offers a degree in field of study a in region r in year t .¹⁰ In total, we have 474 institutions, out of which 93 expanded their online offerings between 2010 and 2019.

Figure 2.4 illustrates the variation in overall online degree expansion across institutions. On the x-axis, each column represents one of the 93 institutions that expanded online between 2010 and 2019, organized by their headquarters' location region. There is significant heterogeneity across institutions. Our instrument predicts more intense online expansion in regions located near institutions that expanded more aggressively.

⁹As discussed in Section 2.2, regulation prohibits the fields of Law, Medicine, and Psychology from offering remote education.

¹⁰Our results are robust to using a leave-one-out estimate of ΔN_f^o . Because our identification strategy relies on having exogenous shares, a leave-one-out estimator is not necessary in this setting.

Figure 2.4: Expansion of online education enrollment by categories of field of study



(a) Online degrees by firm (N_f^o) (b) Prob. of online regional entry by firm (z_{fr}) (c) Share of degrees across field of study by firm (z_{fa})

Notes: This figure illustrates the main patterns in the data driving the different components of the shift-share instrument. Panel (a) shows the total number of online degrees offered by each of the 93 institutions that expanded their online presence between 2010 and 2019 in the years 2010 and 2019 (N_f^o). Based on the IBGE encoding system, institutions are arranged by their headquarters' location region from northwest to southeast. ΔN_f^o corresponds to the difference between the 2010 and 2019 bars. Panel (b) is a heat map representing the exposure of regions to potential entrants (z_{fr}). Each row in the vertical axis is one of the 110 regions in our data where institutions can enter and are arranged from northwest to southeast, following Panel (a). Each row in the horizontal axis is one of the 93 institutions that expanded their online presence between 2010 and 2019 and are arranged from northwest to southeast based on their headquarters' location region. The black lines separate Brazil into five macro regions, such that all cells in the diagonal blocks represent region-institution pairs that are located in the same macro region. Each region-institution pair is shaded in blue, with lighter shades indicating lower exposure, or a lower likelihood that institution f will enter region r , and darker shades indicating higher exposure. Panel (c) shows the share of in-person degrees offered by each institution in 2010 (z_{fa}). Each bar represents one of the 93 institutions that expanded their online presence between 2010 and 2019, and are arranged from northwest to southeast based on their headquarters' location region.

Exposure to potential entrants (z_{fr})

We leverage the differential exposure of regions to institutions' online expansion, driven by their distance from each institution's headquarters. Our empirical analysis shows that institutions are more likely to establish online hubs—and consequently offer degrees—in regions closer to their headquarters, likely due to the lower costs of launching and maintaining nearby operations.¹¹ We document this pattern through the following regression analysis, using data from 2010:

$$\text{Entered}_{fr} = g(d_{fr})'\gamma + \delta_f + \delta_r + \eta_{fr}, \quad (2.5)$$

¹¹These lower costs might be attributed to factors such as easier management and coordination enabled by proximity, reduced setup and operational expenses, and a deeper understanding of local needs and challenges near the headquarters.

where Entered_{fr} equals 1 if institution f had entered region r by 2010 and 0 otherwise; d_{fr} represents the distance between the headquarters of institution f and region r .¹² We define the vector $g(d_{fr}) = [\log(1 + d_{fr}), H_{fr}]'$, where $H_{fr} \in \{0, 1\}$ indicates whether the headquarters of institution f are located in region r (i.e., $d_{fr} = 0$), to account for cases where the distance is zero. Our results show a negative and significant relationship between distance and the probability of entry. Specifically, a 10% increase in the distance between a region and an institution's headquarters reduces the probability of entry by 4%. More detailed results can be found in Online Appendix B.1, Table B2.

We use the estimates from Equation (2.5) to determine the likelihood that institution f will open an online hub in region r , based solely on the distance between the institution and the region. This likelihood is calculated by predicting entry using distance while excluding fixed effects, thereby removing potentially endogenous variation from unobserved regional characteristics. We then normalize the likelihood to sum to one across all regions.¹³ Our measure of exposure is defined as:

$$z_{fr} = \frac{g(d_{fr})'\hat{\gamma}}{\sum_r g(d_{fr})'\hat{\gamma}}. \quad (2.6)$$

Figure 2.4 illustrates the exposure instrument, z_{fr} . The y-axis represents each of the 110 regions in our sample, organized by the five macro-regions defined by the IBGE, with black lines separating each macro-region. Based on the IBGE encoding system, regions are arranged from northwest to southeast within each macro-region. On the x-axis, each column corresponds to one of the 93 institutions that expanded their online presence between 2010 and 2019, ordered by the location of their headquarters, using the same strategy as for the regions and Figure 2.4. Each region-institution pair is shaded in blue, with lighter shades indicating lower exposure, or a lower likelihood that institution f will enter region r , and darker shades indicating higher exposure. We observe that exposure tends to be higher for region-institution pairs that are geographically close and within the same macro-region.

Institutions' propensity to expand in different fields of study (z_{fa})

Section 2.3 focuses on estimating the likelihood that each institution enters a given region. However, different institutions might have different likelihoods of expanding into different fields of study. To address this, we use institutions' 2010 offerings to predict their likely fields of expansion, assuming that institutions initially specializing in certain areas may find it easier to expand within those fields. Since many institutions offered few or no online degrees in 2010, we use the intensity of their in-person instruction to predict where they are

¹²Our results are robust to estimating exposure using a different year.

¹³Equation (2.6) can be interpreted as the probability that institution f opens an online hub in region r , assuming it opens exactly one hub, with the probability depending only on the distance between the institution's headquarters and the region.

most likely to expand online. We estimate the likelihood of expansion in each field based on the share of in-person degrees offered in that field in 2010. Specifically, we calculate:

$$z_{fa} = \frac{\sum_r N_{fra,2010}^\ell}{\sum_a \sum_r N_{fra,2010}^\ell}, \quad (2.7)$$

where $N_{fra,2010}^\ell$ takes the value of 1 if institution f offers an in-person degree from field of study a in region r in 2010.

Figure 2.4 illustrates the resulting shares by institution, aggregated into broader groups of study areas. On the x-axis, each column represents one of the 93 institutions that expanded their online presence between 2010 and 2019, ordered by the location of their headquarters, as in Figures 2.4 and 2.4. We observe that some institutions concentrated more on fields like Education and Business, while others prioritized areas such as Engineering or Math. Our instrument predicts more intense online expansion in field of study a in regions with nearby institutions that were already specializing in that area of study in 2010.

Areas of study regulatory constraints (z_a)

Throughout our sample, regulations prohibit online education in Law, Medicine, and Psychology. We use this constraint to build an indicator variable, z_a , which equals 1 for areas of study allowed to expand online and 0 for those that are not. We incorporate z_a into Equation (2.3), ensuring that our instrument predicts zero online growth in areas of study such as Law, Medicine, and Psychology.

Identification

As is standard in linear models using instrumental variables, two assumptions must hold for the correct interpretation of causal effects. First, the instrument must be relevant, meaning it must have predictive power over the endogenous variable. We test this by conducting a first-stage regression between the SSIV, z_{ra} , and the endogenous variable, ΔN_{ra}^o —the change in the number of online degrees offered in region r and field a —yielding an F-statistic of 201. The coefficient from this regression is presented in Table 2.2, Panel B, Column (1). Second, the exclusion restriction must hold. Specifically, the exposure shares, z_{fra} , must be uncorrelated with the structural error term, ε_{ra} . Formally, this requires that $\mathbb{E}[z_{fr} z_{fa} z_a \varepsilon_{ra}] = 0$ for all expanding firms (i.e., $\forall f$ where $\Delta N_f^o \neq 0$).

To build intuition around the identification assumption, it helps to define $\bar{\varepsilon}_{fr} = \mathbb{E}[z_{fa} z_a \varepsilon_{ra} | fr]$, which represents the propensity of institution f to expand in region r due to anticipated region-field-specific shocks, ε_{ra} . For example, if institution f specializes in business degrees in 2010, and region r is expected to see an increase in demand for online business degrees, institution f might be more likely to offer online degrees in that region, making $\bar{\varepsilon}_{fr}$ larger. Using the law of iterated expectations, the identification assumption can be reformulated as $\mathbb{E}[z_{fr} z_{fa} z_a \varepsilon_{ra}] = \mathbb{E}[z_{fr} \bar{\varepsilon}_{fr}] = 0$ for all expanding firms. This

condition holds if the distance between regions and a given institution is uncorrelated with region-specific unobserved demand shocks, ε_{ra} , related to the field of study in which the institution specializes in in-person education.

Results

Table 2.2 presents the results for both identification strategies. Panel A displays the OLS results, while Panel B shows the SSIV results. We find that the OLS and SSIV strategies yield qualitatively similar results, though the IV estimates are larger in magnitude. We interpret this as evidence that endogeneity concerns about online entry are small, as institutions might have difficulty anticipating future demand shocks when deciding to expand their online portfolio. This suggests that most online expansion decisions are driven by other factors, such as costs—which are captured by our shift-share instrument—and further influenced by the 2016 policy reforms that facilitated rapid expansion to close locations that might have lower expansion costs. Given the similarity of the results, we focus our discussion on the SSIV, though the conclusions are qualitatively the same for OLS.

Average effects

We begin by examining the impact of introducing an additional online degree in a specific field and region on both online and in-person enrollment, with results presented in Columns (2) and (3), respectively. These outcomes are reported relative to market size, defined as the number of individuals aged 18 to 45 without a college degree in that region. Column (2) indicates that each additional online degree introduced between 2010 and 2019 in a given field increased online enrollment by 0.391 students per 1,000 individuals in the market. This is equivalent to a 14% increase in total enrollment when the number of online degrees in that field is raised by 50%. In contrast, Column (3) shows that this increase in online degrees led to a reduction in in-person enrollment in that field of study by 0.2 students per 1,000 individuals, which corresponds to a 7% decline in total enrollment under the same 50% increase in online degree availability.

These results reveal two opposing forces: online degrees expand the market by attracting new students to college while simultaneously diverting students from in-person programs. Our findings show that for each additional online student, 52% are new to higher education, while 48% would have otherwise enrolled in an in-person degree.¹⁴ These forces create an ambiguous effect on total value added. Market expansion increases value added, as newly enrolled students enter programs with positive value compared to the outside option of no college, which we normalize to have zero value added. However, market diversion reduces value added by shifting students from higher value-added in-person programs to lower value-added online alternatives. Based on the average value added differences between online and

¹⁴Note that the linear model does not allow for cross-field effects, thereby excluding the possibility of expansion coming from other fields of study.

in-person degrees, a back-of-the-envelope calculation suggests that increasing the number of online degrees by 50% raises value added by 7.6% due to the increase in online enrollment, while the corresponding decline in in-person enrollment reduces value added by 7.7%. These effects largely cancel out, resulting in a net reduction in total value added of 0.04%.¹⁵

Table 2.2: Effects of introducing an additional online degree

	Δ in online degrees (1)	Δ in online students (2)	Δ in in-person students (3)	Δ in in-person degrees (4)	Δ in log-price of in-person degrees (5)
Panel A: OLS regression					
Δ in online degrees		0.344 (0.033)	-0.190 (0.026)	-0.148 (0.031)	-0.009 (0.002)
Panel B: IV regression					
shift-share instrument	1.723 (0.122)				
Δ in online degrees		0.391 (0.035)	-0.187 (0.033)	-0.184 (0.037)	-0.014 (0.003)
Panel C: Average value of the dependant variable in levels in 2010 and 2019					
2010	2.82	0.72	3.10	10.98	5.63
2019	9.48	2.79	2.94	12.19	6.21
Obs.	1,210	1,210	1,210	1,210	957

Notes: This table shows the results from estimating the linear model from Equation (2.2), $\Delta y_{ra} = \phi \Delta N_{ra}^o + \varepsilon_{ra}$, for different outcomes, y_{ra} , using two different identification strategies. Panel A shows the results estimating the model via OLS regressions. Panel B estimates the model via 2SLS using the shift-share design described in Section 2.3. Panel C shows the average value of the outcome, y_{ra} , for years 2010 and 2019. Column (1) shows the 2SLS first-stage regression of the total number of online degrees against the shift-share instrument. Columns (2)-(5) present the regression coefficients for the number of online students relative to market size, the number of in-person students relative to market size, the total number of in-person degrees, and the average price of in-person degrees. Columns (1)-(4) are estimated using all region-field of study pairs, while column (5) uses the 954 region-field of study pairs for which there is at least one in-person degree in 2010 and 2019.

We next analyze the consequences of the online expansion on the availability and pricing of in-person degrees. In Column (4), we show that the number of in-person degrees decreased in regions and fields with larger online growth. Specifically, for each additional online degree, there are 0.184 fewer in-person degrees relative to a 2010 baseline of 10.98. These effects stem

¹⁵We calculate changes in total value added using the formula: $\Delta VA = \frac{\Delta s^o \cdot VA^o + \Delta s^i \cdot VA^i}{s^o \cdot VA^o + s^i \cdot VA^i}$. In this expression, Δs^o and Δs^i represent the coefficients from Columns (2) and (3) of Table 2.2, Panel B. The terms s^o and s^i denote the average number of online and in-person students attending college in 2010 for every 1,000 individuals aged 18 to 45 living in that region, as reported in Columns (2) and (3) of Table 2.2, Panel C. Additionally, VA^o and VA^i represent the average value added of online and in-person degrees in 2010, given by 0.078 and 0.164, respectively.

from both an increase in degree exit and entry deterrence, which exacerbates the diversion from in-person degrees to online alternatives. As a result, even students who prefer to enroll in a specific in-person program may be forced to switch to an online alternative when their preferred in-person option exits the market. Finally, in Column (5), we show that the average price of in-person degrees dropped by 1.4% for each additional online degree introduced in that region and field of study. This supports the idea that online degrees intensify local competition, driving down prices and deterring new in-person program entry.

Effects by age groups

The expansion of online education may affect age groups differently. Evidence shows that online programs tend to attract older individuals, who often face greater challenges and costs in attending in-person classes (Goodman et al., 2019; Aucejo et al., 2024). In our context, by 2010, college enrollees aged 18-20 were 12 times more likely to pursue an in-person degree over an online degree than students aged 36-45. Differences in preferences across age cohorts can lead to heterogeneous patterns of market expansion and diversion that we analyze next.

In Table 2.3, we estimate enrollment changes across different age cohorts. We find high rates of market diversion from in-person to online degrees among younger cohorts. For students aged 18-21, each additional online enrollment leads to 0.67 students leaving the in-person sector (ratio of the coefficients in Column 5 to Column 1, Panel B), indicating that online education acts as a substitute for in-person alternatives. For students aged 21-25, 0.52 students leave in-person programs for every new online student (ratio of the coefficients in Column 6 to Column 2, Panel B). For older cohorts (ages 26-35 and 36-45), this substitution drops to 0.43 and 0.26 students, respectively (calculated similarly using Panel B coefficients). A back-of-the-envelope calculation suggests that increasing the number of online degrees by 50% decreases value added by 1.6% among students aged 18-20 but increases value added by 8.9% among students aged 36-45.

Overall, the results from Tables 2.2 and 2.3 indicate that online education has the potential to expand access to higher education, especially among older cohorts. However, online education also diverts students from in-person alternatives. Increased competition drives down in-person tuition but reduces the number of in-person degrees. Consequently, the average value added in the economy declines, and students who would prefer in-person learning may find themselves pushed into online alternatives as their preferred programs disappear.

This section's analysis relies on a linear model that assumes no interaction between degrees in different fields within the same region. This limitation is important because it prevents the model from distinguishing between overall market growth driven by new students entering college and shifts caused by students switching from other fields of study. Moreover, the linear approach is unsuitable for out-of-sample counterfactuals, where competition, price changes, and entry/exit decisions can have nonlinear effects.¹⁶ We address these concerns in the next section.

¹⁶Nonlinearity is crucial when considering entry and exit decisions. While small increases in online

Table 2.3: Effects of introducing an additional online degree on enrollment by age groups

	Δ in online students				Δ in in-person students			
	18-20	21-25	26-35	36-45	18-20	21-25	26-35	36-45
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: OLS regressions by cohort								
Δ in online degrees	0.333 (0.044)	0.414 (0.043)	0.370 (0.036)	0.263 (0.024)	-0.332 (0.058)	-0.235 (0.037)	-0.172 (0.030)	-0.081 (0.012)
Panel B: IV regressions by cohort								
Δ in online degrees	0.393 (0.054)	0.472 (0.045)	0.419 (0.038)	0.295 (0.025)	-0.262 (0.079)	-0.247 (0.044)	-0.181 (0.034)	-0.076 (0.016)
Panel C: Average value of the dependant variable in levels in 2010 and 2019								
2010	0.67	0.75	0.96	0.81	6.35	2.70	1.34	0.63
2019	3.50	3.08	3.05	2.55	8.20	2.59	1.01	0.55
Obs.	1,210	1,210	1,210	1,210	1,210	1,210	1,210	1,210

Notes: This table shows the results from estimating the linear model from Equation (2.2), $\Delta y_{ra} = \phi \Delta N_{ra}^o + \varepsilon_{ra}$, for the change in the number of in-person and online students from different age groups, using two different identification strategies. Panel A shows the results estimating the model via OLS regressions. Panel B estimates the model via 2SLS using the shift-share design described in Section 2.3. Panel C shows the average value of the outcome, y_{ra} , for years 2010 and 2019. Columns (1)-(4) show the regression coefficients for the number of online students relative to market size for age cohorts 18-20, 21-25, 26-35, and 36-45. Columns (5)-(8) show the regression coefficients for the number of in-person students relative to market size for the same age cohorts.

2.4 Model

We develop and estimate an equilibrium model of the Brazilian college education market, covering both in-person and online formats. This model overcomes two key limitations of linear models: first, by explicitly modeling demand, it flexibly accounts for substitution across areas of study; second, by modeling supply, it supports counterfactual analyses that incorporate price adjustments and entry/exit decisions in equilibrium. Using this framework, we assess the impact of expanded online education on value added and explore targeted policies that restrict online education to older cohorts.

Setup

The model consists of institutions (i.e., firms) $f \in \mathcal{F}$ that offer undergraduate college degrees. Students can choose to major in a particular field of study $a \in \mathcal{A}$ by enrolling in either an

competition may have minimal impact, extrapolating this effect to larger scales could misleadingly imply that substantial increases in competition would also have no effect. Therefore, simple linear extrapolation cannot accurately represent the relationship between the number of online competitors and market structure outcomes.

in-person or online degree. A degree is defined as the combination of the firm offering it, the field of study it belongs to, and whether it is in-person or online, and is represented by $j \in \{\mathcal{F} \times \mathcal{A} \times \{0, 1\}\}$. Geographical regions are denoted by $r \in \mathcal{R}$ and years by $t \in \mathcal{T}$. Markets are the intersection of regions and years. A single degree j can be available in multiple markets, and a market may offer multiple degrees. Each firm offers a product bundle (i.e., a combination of degrees) $\mathcal{J}_{f,rt} \in \mathcal{B}_f$ in each market, where $\mathcal{J}_{f,rt} = \emptyset$ represents the option of not offering any degree and is available for every firm.¹⁷

The model has two stages. In the first stage, institutions decide whether to operate in a given market and select a bundle to offer by comparing expected profits to bundle-specific entry fixed costs. At this stage, institutions know the pre-existing market structure, $\mathcal{J}_{f,rt_0}$, all relevant characteristics of demand and marginal and fixed costs of potential entrants up to idiosyncratic shocks to demand and marginal costs, and their own private-information fixed-costs shocks. Institutions' entry choices determine the market structure. After entry decisions are made, demand and marginal cost shocks are realized. In the second stage, institutions compete a la Bertrand by setting tuition prices for the degrees in their chosen bundle, and demand is realized.

In the following sections, we present the components of the model in reverse order. First, we outline the demand model. Second, we discuss the institutions' pricing decisions. Third, we detail the institutions' optimal bundle choice.

Demand

In each year t and region r , a potential student, $i \in \mathcal{I}_{rt}$, decides whether to enroll in a degree $j \in \mathcal{J}_{rt}$, where $\mathcal{J}_{rt} = \cup_f \mathcal{J}_{f,rt}$ are the degrees available in year t and region r , or not to enroll at all. A degree j is characterized by the institution f that offers it, the field of study a it belongs to, and a vector of characteristics $x_j = [x_j^{(1)}, x_j^{(2)}]$ that we define below.

The utility student i gets from enrolling in degree j in region r in year t is:

$$u_{ijrt} = -\alpha_i p_{jrt} + x_j^{(1)} \beta_i + \psi w_{jrt} + \delta_{jrt} + \epsilon_{ijrt}, \quad (2.8)$$

where p_{jrt} denote the tuition price, $x_j^{(1)}$ is a vector that indicates whether the degree is online or in-person (i.e., $x_j^{(1)} \in \{[1, 0], [0, 1]\}$), w_{jrt} is a demand shifter based on region r 's internet penetration in year t interacted with whether degree j is in person or online, and δ_{jrt} captures degree-market-specific characteristics constant across individuals. We allow individuals to have heterogenous preferences over tuition prices and delivery mode, assuming $[\log(\alpha_i), \beta_i]' \sim \mathcal{N}(\mu_{b(i)}, \Sigma)$, where $\mu_{b(i)}$ depends on the student's age bin $b(i)$. The term ϵ_{ijrt} represents a consumer-specific demand shock following a generalized extreme value distribution, consistent with a nested logit model. The nests are defined by the degree's area of study, and the intra-nest correlation is denoted by ρ .

¹⁷The set of available bundles, $\mathcal{B}_f$, is firm specific to allow for firm specialization. For instance, some institutions may lack the necessary technology to offer degrees in certain fields, such as Medicine.

We further decompose the degree-market-specific utility as follows:

$$\delta_{jrt} = x_j^{(2)}\delta + \delta_j + \delta_{ra} + \delta_{ta} + \delta_{toj} + \xi_{jrt}, \quad (2.9)$$

where $x_j^{(2)}$ include a constant, the age of the program, the average score of incoming students, the average wages of graduate students, the length of the degree (in number of hours required), and the degree's stem load, calculated as the share of degree programs comprising the degree that are stem. The term δ_j captures degree-specific components of utility, δ_{ra} captures area of study and region-specific factors, δ_{ta} represents area of study and year-specific components, δ_{toj} account for online-year factors, allowing for differential yearly demand shifts between online and in-person education. We refer to these terms as the mean utility components of δ_{jrt} . Finally, ξ_{jrt} denotes a degree-region-year-specific idiosyncratic demand shock.

We denote s_{jrt} as the share of potential students from region r in year t who choose to enroll in degree j , and is calculated as:

$$s_{jrt}(\mathbf{p}_{rt}) = \int_{i \in \Theta_{rt}} di, \quad (2.10)$$

where $\mathbf{p}_{rt}$ denotes the market vector of degree prices, and $\Theta_{rt} = \{i \in \mathcal{I}_{rt} : u_{ijrt} \geq u_{ikrt}, \forall k \in \mathcal{J}_{rt}\}$ captures the set of potential students in region r and year t who choose to enroll in degree j .

Identification and estimation

We estimate the demand model with yearly data from 2010 to 2019 using the generalized method of moments introduced by [Berry et al. \(1995\)](#), combining both instrumental variables and micro-moments to identify the model parameters.¹⁸

Instruments for prices. A key challenge in demand estimation is price endogeneity, as institutions may set prices in response to unobserved demand shocks ξ_{jrt} . Ideally, we would rely on cost-shifters that affect pricing decisions but are independent of demand shocks. When these shifters are difficult to observe, proxies serve as a practical alternative. Following [Hausman et al. \(1994\)](#), we use the contemporaneous prices of the same degree in other regions as a proxy, defined as $z_{jrt}^p = \frac{1}{|\mathcal{R}(j)|-1} \sum_{r' \neq r} p_{jr't}$, which represents the average price of degree j in year t in regions other than r . The intuition behind this instrument is that variation in firm-level costs impacts prices across all markets where the degree is offered. The primary identification assumption is that while costs for degree j are correlated across regions, demand shocks are not.

Instruments for substitution patterns. The demand model incorporates individual heterogeneity through random coefficients on α_i and β_i and the nesting parameter ρ . To identify

¹⁸For a comprehensive review of related literature, see [Berry and Haile \(2016\)](#), and for a detailed guide on best practices, including those we adopt, refer to [Conlon and Gortmaker \(2020\)](#).

these parameters, we use instruments that shift the choice set, $\mathcal{J}_{rt}$, but are unrelated to demand. To construct these instruments, we employ shift-share type instruments similar to the one outlined in Section 2.3. Specifically, we build two instruments designed to predict the number of online degrees, z_{rta}^o , and the number of in-person degrees, z_{rta}^l , offered from a given field of study in each market. Each instrument is calculated as follows:

$$z_{rta}^o = \sum_f z_{fr} z_{fa} z_a N_{ft}^o \quad , \quad z_{rta}^l = \sum_f z_{fr} z_{fa} N_{ft}^l$$

where z_{fr} represents the likelihood of institution f expanding to region r , based on the distance between the institution's headquarters and the region as defined in Section 2.3; z_{fa} captures the institution's propensity to expand into different fields of study as outlined in Section 2.3, z_a is the indicator variable (defined in Section 2.3) that takes a value of zero for fields of study where online instruction is not allowed, and N_{ft}^o and N_{ft}^l represent the total number of online and in-person degrees offered by institution f in year t across all regions and fields of study.

These instruments shift the total number of online and in-person degrees available in region r , year t and field of study a , respectively, providing variation to identify the variance of β_i , the random coefficients on the two indicator variables for online or in-person degrees, and the nesting parameter ρ . Finally, following Gandhi and Houde (2019), we construct differentiation instruments as functions of the distance between degree j 's predicted price and other degrees' predicted prices given by $z_{jrt}^{gh1} = \frac{1}{|\mathcal{J}_{rt}|-1} \sum_{k \in \mathcal{J}_{rt}} |\hat{p}_{jrt} - \hat{p}_{krt}|$ and $z_{jrt}^{gh2} = \frac{1}{|\mathcal{J}_{rt}|-1} \sum_{k \in \mathcal{J}_{rt}} \mathbb{I}\{|\hat{p}_{jrt} - \hat{p}_{krt}| < \sigma_{\hat{p}}\}$, where $\hat{p}_{jrt}$ is the predicted price based the reduced-form pricing equation constructed from the price instrument and fixed effects and $\sigma_{\hat{p}}$ is the standard deviation of $\hat{p}_{jrt}$.¹⁹ These instruments generate variation that helps identify the variance of the random coefficient on price, α_i , and the nesting parameter ρ .

Micro-moments for age heterogeneity. We incorporate additional micro-moments to discipline the age heterogeneity parameters, $\mu_{b(i)}$, by matching moments predicted by the model with their empirical counterparts. We use eight moments defined by the probability that students from each of four age bins defined above choose to enroll in person or online.²⁰

Estimator. The estimator proposed by Berry et al. (1995) produces consistent estimates of $\mu_{b(i)}$, Σ , ψ , and δ_{jrt} , which together recover the distribution of α_i , β_i , market shares, and price elasticities. The estimator treats the mean utility components of δ_{jrt} from Equation (2.9) as nuisance parameters that might be unprecisely estimated. This is typically not an

¹⁹The reduced form regression is given by $p_{jrt} = \zeta_p z_{jrt}^p + \zeta_w w_{jrt} + \zeta_j + \zeta_{ra} + \zeta_{ta} + \zeta_{to_j} + \varepsilon_{jrt}$, where z_{jrt}^p is the instrument for price, and ζ_j , ζ_{ra} , ζ_{ta} , and ζ_{to_j} are degree, region-field of study, year-field of study, and year-online fixed effects.

²⁰Formally, the moments are defined as $\mathbb{E}[\sum_{j \in \mathcal{J}_{rt}} s_{ijrt} x_j^{(1)}(k) | i \in b]$ for $k \in \{1, 2\}$, where q_{ijrt} is the probability that individual i chooses degree j , $x_j^{(1)}(k)$ is the k -th element of $x_j^{(1)}$ (i.e., online or in-person), and b represents age bins corresponding to the age groups 18-20, 21-25, 26-35, and 36-45.

issue because market shares and price elasticities depend on δ_{jrt} and not on the mean utility components of δ_{jrt} . However, in our setting, we are interested in recovering demand for degrees that are not offered in certain markets in the data and for which we don't have estimates of δ_{jrt} . This requires reliable estimates of the mean utility components of δ_{jrt} .

To estimate the demand model, we proceed in two steps. In the first step, we follow [Berry et al. \(1995\)](#) and [Petrin \(2002\)](#), and estimate the parameters from Equation (2.8) using the generalized method of moments. We absorb the mean utility components of δ_{jrt} via fixed effects. Due to collinearity, $x_j^{(2)}$ is absorbed by δ_j during the estimation. We use a total of thirteen moments, five moments defined by $\mathbb{E}[\xi_{jrt}|Z_{jrt}] = 0$, where ξ_{jrt} is the demand shock from Equation (2.9), and Z_{jrt} are the instruments z_{jrt}^p , z_{rt}^o , z_{rt}^l , z_{rta}^a , and z_{jrt}^{gh} , and the eight micro-moments defined above.

In the second step, we impose additional structure on Equation (2.9) and estimate a mixed-effects Bayesian hierarchical model to recover each of the mean utility components of δ_{jrt} . We assume that $\delta_j \sim \mathcal{N}(0, \sigma_j^2)$, $\delta_{ra} \sim \mathcal{N}(0, \sigma_{ra}^2)$, $\delta_{ta} \sim \mathcal{N}(0, \sigma_{ta}^2)$, $\delta_{to} \sim \mathcal{N}(0, \sigma_{to}^2)$, and $\xi_{jrt} \sim \mathcal{N}(0, \sigma_\xi^2)$ are random coefficients, and δ is a non-random parameter that allows for flexible correlation between δ_{jrt} and $x_j^{(2)}$. We estimate the model via maximum likelihood to recover posterior means of the mean utility components of δ_{jrt} that we use to project utility into degree-markets, regardless of whether they are offered in the data or not.²¹ This second step is in the spirit of [Abdulkadiroğlu et al. \(2020\)](#) and [Andrews et al. \(2024\)](#), and serves to shrink imprecise estimates, reducing noise in the profit function and improving the estimation of the entry game.

Results

We present the estimated parameters in Online Appendix B.1, Table B3. To summarize our results, we present median own-price elasticities and diversion ratios for 2010 and 2019 in Table 2.4. We estimate median own-price elasticities of approximately -2.9 for in-person degrees and -1.1 for online degrees, which aligns with findings from the literature that estimates demand for in-person degrees in the U.S. ([Armona and Cao, 2024](#)) and Brazil ([Dobbin et al., 2021](#)).²² We find that, by 2019, when the price of an in-person degree marginally increases, 67% of students who leave that degree would enroll in another in-person program, 18% would switch to an online program, and 14% would exit higher education altogether. Additionally, we find that older cohorts show stronger preferences for attending online degrees (see Online Appendix B.1, Table B4). Finally, our findings indicate that higher internet penetration in-

²¹In a standard fixed-effects model, estimating δ when x_j and δ_j are collinear is problematic because δ_j would be treated as a fixed parameter, absorbing variation in x_j . However, in a mixed-effects framework, estimation is possible. The mixed model is estimated using maximum likelihood, where the key idea is that δ_j is integrated out in the likelihood function, leaving only the fixed parameters δ and variance components of the random effects to be estimated. Intuitively, within- j variation in δ_{jrt} helps estimate δ , while between- j variation helps estimate σ_j^2 .

²²To the best of our knowledge, no studies have reported price elasticities for online degrees.

creases demand for both online and in-person degrees, which we use as an additional source of variation to estimate entry fixed cost in the next subsections.

Table 2.4: Elasticity and diversion ratios

	$t = 2010$		$t = 2019$	
	In-person	Online	In-person	Online
Median own-price elasticity:	-2.70	-1.10	-2.91	-1.14
Median diversion ratios:				
To in-person:	0.78	0.34	0.67	0.16
To online:	0.04	0.49	0.18	0.70
To outside good:	0.15	0.15	0.14	0.14

Notes: The table reports median own-price elasticities and diversion ratios across products and markets for 2010 and 2019 for in-person and online degrees. We calculate diversion ratios as the share of students who decide to stop attending degree j upon an increase in tuition price that would switch to either an in-person degree, an online degree, or the outside option. Formally, we calculate diversion ratios as $D_{j\mathcal{K}} = \left(\left| \frac{\partial s_j}{\partial p_j} \right| \right)^{-1} \left(\sum_{k \in \mathcal{K}/j} \frac{\partial s_k}{\partial p_j} \right)$, where $\mathcal{K}$ the set of all degrees that are either in-person, online, or the outside option. For example, the diversion ratio from in-person degrees to online degrees in 2010 is 0.04, which is given by the median of $D_{j\mathcal{K}}$ across all degrees j that are in-person and for $\mathcal{K}$ the set of all online degrees.

Institutions' pricing decision

The variable profits of institution f in region r and year t , given its choice of degree offerings $\mathcal{J}_{frt}$, are expressed as:

$$\pi_{frt}(\mathcal{J}_{frt}) = \max_{\{p_{jrt}\}_{j \in \mathcal{J}_{frt}}} \sum_{j \in \mathcal{J}_{frt}} (p_{jrt} - c_{jrt}) \cdot s_{jrt}(\mathbf{p}_{rt}). \quad (2.11)$$

The log of marginal cost c_{jrt} is given by:

$$\log(c_{jrt}) = g(d_{jrt})' \gamma_g + \gamma_z z_{jrt}^p + \underbrace{x_j^{(2)} \gamma_x + \gamma_j + \gamma_{ra} + \gamma_{ta} + \gamma_{to}}_{\gamma_{jrt}} + \omega_{jrt}, \quad (2.12)$$

where the vector $g(d_{fr}) = [\log(1 + d_{fr}), H_{fr}]'$ captures the log-distance between institution f 's headquarters and region r , as well as an indicator for whether region r corresponds to the headquarter's location, as in Equation (2.5). The term z_{jrt}^p represents the price instrument used to estimate demand, $x_j^{(2)}$ denotes the degree characteristics from Equation (2.9), γ_j

captures degree-specific components of the cost function, γ_{ta} represent area of study-year specific components, γ_{to} reflects online-year specific components that allow for differential trends in demand for online and in-person education. Lastly, ω_{jrt} is a degree-region-year specific idiosyncratic supply shock. Institutions compete a la Bertrand.

Estimation

We estimate the cost parameters in three steps. First, we recover c_{jrt} for all degrees offered in the data by inverting the firms' first-order conditions. Second, we estimate γ_g and γ_z via OLS and absorb the elements of γ_{jrt} as fixed effects. Analogous to the demand model estimation, the components of γ_{jrt} are treated as nuisance parameters and are not estimated precisely. Because we need to estimate costs for degrees that may not be offered in certain markets in the data, we require reliable estimates for each component of γ_{jrt} . We do this in a third step, where we estimate a mixed-effects Bayesian hierarchical model where

$$\gamma_{jrt} = x_j^{(2)}\gamma + \gamma_j + \gamma_{ra} + \gamma_{ta} + \gamma_{to} + \omega_{jrt}, \quad (2.13)$$

with $\gamma_j \sim \mathcal{N}(0, \varsigma_j^2)$, $\gamma_{ra} \sim \mathcal{N}(0, \varsigma_{ra}^2)$, $\gamma_{ta} \sim \mathcal{N}(0, \varsigma_{ta}^2)$, $\gamma_{to} \sim \mathcal{N}(0, \varsigma_{to}^2)$, and $\omega_{jrt} \sim \mathcal{N}(0, \varsigma_{\omega}^2)$ random coefficients, and γ a non-random parameter that allows for flexible correlation between γ_{jrt} and $x_j^{(2)}$. We estimate the model via maximum likelihood to recover posterior means of each component of γ_{jrt} that we use to estimate marginal costs for all degree-markets, regardless of whether they are offered in the data or not.

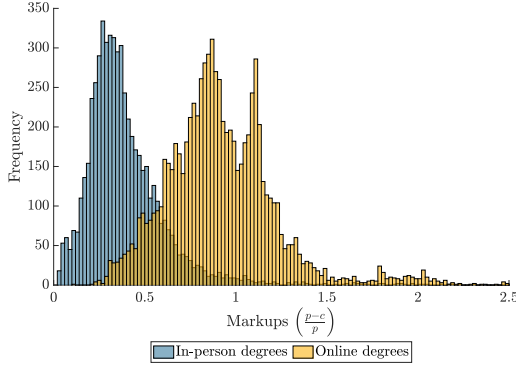
Results

We present the distribution of markups—defined as the ratio of price minus marginal cost to price—for in-person and online degrees in Figure 2.5. We present the corresponding estimated parameters in Online Appendix B.1, Table B5. We find that online degrees have higher markups than in-person degrees, consistent with institutions pricing their online alternatives at a more inelastic part of the demand curve. However, despite having larger markups, online degrees' prices are lower due to lower marginal costs. The average price of in-person and online degrees is \$5650 and \$1821 per year, respectively, and the average estimated marginal cost is \$3966 and \$303 per year, respectively.

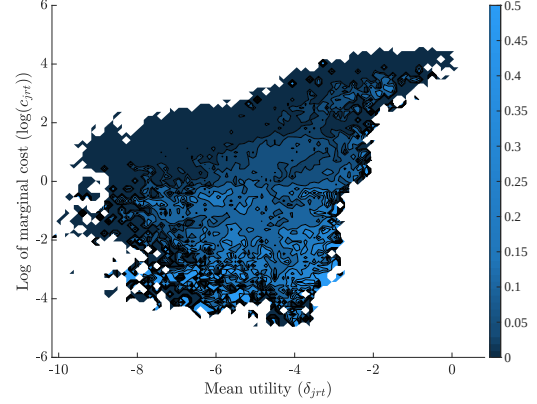
We also find that degrees that face higher demand—through a higher value of the mean utility, δ_{jrt} —and a lower marginal cost, are offered in more regions in 2019. Figure 2.5 shows a heatmap with the relationship between degrees' demand, marginal cost, and their empirical probability of being offered in 2019. Higher probabilities are more prevalent in the south-east region of the map. This finding suggests that, in the data, institutions are choosing more profitable degrees when deciding which bundles to offer, which is consistent with economic theory.²³

²³This result is not mechanically driven from the model. Entry probabilities are estimated directly from

Figure 2.5: Estimated parameters from demand and pricing decision



(a) Distribution of markups



(b) Probability of a degree being offered in each region in 2019

Notes: The figure summarizes the estimated demand and marginal cost functions. Panel (a) shows the distribution of markups, defined as the ratio between the difference between prices and marginal costs and prices, for in-person and online degrees. In-person degrees are depicted in blue and online degrees in yellow. Panel (b) shows the relationship between demand and marginal costs determinants and the empirical probability that a given degree is offered in each market. The x-axis represents the mean utility components of δ_{jrt} of each degree, and the y-axis the logarithm of their predicted marginal cost. The shaded area represents the empirical probability that degrees with given characteristics are offered in each region in 2019. Lighter colors reflect larger probabilities. Uncolored areas are combinations of mean utility and marginal costs for which no potential degree exists. Note that entry probabilities are directly estimated from the data and are agnostic about the underlying entry model.

Institutions' choice of degrees' offerings

In this subsection, we model institutions' offerings decisions in 2019, denoted by $\mathcal{J}_{f,rt}$, based on the existing degree offerings in 2010, denoted by $\mathcal{J}_{f,rt_0}$. We take a static approach to the problem for two reasons. First, there are adjustment frictions that make the timing of institutions' responses to economic shocks uncertain. We avoid modeling such frictions by comparing medium-run changes in market structure over a span of nine years. Second, in our setting, many institutions can offer multiple degrees in each market, which makes the state space highly dimensional and the problem intractable.²⁴ Our static approach captures the main trade-offs institutions face when choosing offerings in a parsimonious way.

the data, while demand and marginal costs are estimated from market shares and prices from existing degrees.

²⁴Bodéré (2023) develops an approximation method for high-dimensional dynamic games with single-product firms. Our setting features multi-product firms, which complicates the approximation of the state space due to the need to account for interactions across multiple products.

Institution f 's fixed cost of offering a given bundle, $\mathcal{J}_{f,rt}$, depends on three key components: (i) the composition of the bundle and whether its degrees were part of the previous bundle, $\mathcal{J}_{f,rt_0}$, (ii) the infrastructure investments required to offer the degrees in the bundle, which depend on the need for an in-person campus or online hub, the existence of prior infrastructure, and the distance from f 's headquarters to region r , and (iii) a private-information cost shock only known by institution f . The fixed cost function is parametrized as follows:

$$FC_{fr}(\mathcal{J}_{f,rt}|\mathcal{J}_{f,rt_0}) = \text{Degrees}(\mathcal{J}_{f,rt}|\mathcal{J}_{f,rt_0}) + \text{Infrastructure}(\mathcal{J}_{f,rt}|\mathcal{J}_{f,rt_0}) + \sigma_\varepsilon \varepsilon_{fr\mathcal{J}_{f,rt}}, \quad (2.14)$$

where $\varepsilon_{fr\mathcal{J}_{f,rt}}$ is a firm-region-bundle specific idiosyncratic shock that is only observed by institution f and that we assume to follow an extreme value type I distribution. These shocks, unobserved by the researcher, help to rationalize firms' bundle choices observed in the data, and can include firms' private information about the economic returns of offering certain degrees or expanding into specific regions.

The first component of the fixed cost equation represents the cost associated with maintaining existing degrees or opening new ones:

$$\text{Degrees}(\mathcal{J}_{f,rt}|\mathcal{J}_{f,rt_0}) = \mathbb{1}\{\mathcal{J}_{f,rt} \neq \emptyset\} \cdot \kappa_0 + \sum_{k \in \mathcal{J}_{f,rt} \cap \mathcal{J}_{f,rt_0}} \kappa_{a(k)o(k)}^{\text{old}} + \sum_{k \in \mathcal{J}_{f,rt} \setminus \mathcal{J}_{f,rt_0}} \kappa_{a(k)o(k)}^{\text{new}}, \quad (2.15)$$

where κ_0 captures an operational fixed cost that turns on when at least one degree is offered, $\kappa_{a(k)o(k)}^{\text{old}}$ represents the cost of maintaining an existing degree, and $\kappa_{a(k)o(k)}^{\text{new}}$ denotes the cost of opening a new degree using the existing infrastructure. Note that these costs only vary at the field of study level, indexed by $a(k)$ for each degree k , and whether the degree is in person or online, indexed by $o(k)$. We restrict κ_0 , κ_{ao}^{old} , and κ_{ao}^{new} to be non-negative so that offering degrees is more expensive than not offering them.

The second component reflects the cost associated with opening a new campus or hub necessary to host a given degree:

$$\text{Infrastructure}(\mathcal{J}_{f,rt}|\mathcal{J}_{f,rt_0}) = \sum_{k \in \{\text{campus}, \text{hub}\}} \mathbb{1}\{\text{New}_{frk}\} (\chi_0^k + g(d_{fr})' \chi^k), \quad (2.16)$$

where $\mathbb{1}\{\text{New}_{frk}\}$, for $k \in \{\text{campus}, \text{hub}\}$, is an indicator variable that equals one if $\mathcal{J}_{f,rt}$ includes an in-person or online degree (respectively), while $\mathcal{J}_{f,rt_0}$ does not. This variable captures the transition in a firm's offerings in region r , reflecting the need to invest in building the required campus infrastructure for in-person degrees if the firm previously did not offer any in-person programs, or in developing online hubs if the firm had not previously provided online degrees. The vector $g(d_{fr})$ is the distance function between the firm's headquarters and the region, as defined in Equation (2.5). To account for differences in cost, we allow χ_0^k and χ^k to vary between building a new campus or a new online hub. As before, we impose the restriction that $(\chi_0^k + g(d_{fr})' \chi^k)$ is non-negative to ensure that building new infrastructure is more costly than utilizing existing infrastructure.

Firms decide which degree bundles to offer after observing the region-bundle-specific idiosyncratic shocks, $\varepsilon_{fr\mathcal{J}_{frt}}$, but before the demand and supply shocks, ξ_{jrt} and ω_{jrt} , are realized. Their objective is to maximize expected profits, as defined by the firm's problem:

$$\Pi_{frt} = \max_{\mathcal{J}_{frt} \in \mathcal{B}_f} \mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})] - FC_{fr}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}), \quad (2.17)$$

where $\pi_{frt}(\mathcal{J}_{frt})$ is given by Equation (2.11) and the expectation is taken over the distribution of all other firms' idiosyncratic shocks, $\varepsilon_{f'r\mathcal{J}'_{f'rt}}$, and over the demand and supply shocks, ξ_{jrt} and ω_{jrt} . We assume the players' strategies form a Bayesian Nash equilibrium.

Identification and estimation

To identify the fixed cost parameters, we exploit variation in the firms' bundle choices across different markets. Nonparametric identification of the fixed costs requires instruments that shift firms' expected profits but do not enter the fixed cost function directly. We use two sets of profit shifters. The first set of instruments corresponds to the demand shifters, w_{jrt} , based on region r 's internet penetration in year t interacted with whether degree j is in person or online as outlined in Section 2.4. In Online Appendix B.3, we show that higher internet penetration growth is associated with higher online entry. The identification assumption is that while greater internet penetration may increase demand for online programs, it does not affect firms' fixed cost of offering different degrees. The second set of instruments is the vector of distances between region r and institution f 's competitors. Firm f 's expected profits will be lower in regions that are close to the headquarters of other firms that offer degrees similar to firm f . However, we only allow fixed costs to depend on firm f 's distance to region r and not that of its competitors.²⁵

Our identification follows the framework for inference in incomplete-information games from Aradillas-López (2020). In their framework, point identification can be obtained by assuming that the underlying equilibrium selection mechanism is degenerate (i.e., the data come from a single equilibrium), without having to assume which equilibrium is chosen (Seim, 2006; Bajari et al., 2010; Atal et al., 2025). We use the firm's maximization problem to recover bundle choice probabilities. From Equation (2.17), the probability that firm f chooses bundle $\mathcal{J}_{frt}$ is given by

$$\phi_{\mathcal{J}_{frt}} = \frac{\exp(\frac{1}{\sigma_\varepsilon} \mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})] - \text{Degrees}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) - \text{Infrastructure}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}))}{\sum_{\mathcal{J} \in \mathcal{B}_f} \exp(\frac{1}{\sigma_\varepsilon} \mathbb{E}[\pi_{frt}(\mathcal{J})] - \text{Degrees}(\mathcal{J}|\mathcal{J}_{frt_0}) - \text{Infrastructure}(\mathcal{J}|\mathcal{J}_{frt_0}))}. \quad (2.18)$$

If expected profits were observed and did not depend on competitors' strategies, estimation of the fixed cost parameters via maximum likelihood would be straightforward. However, the expected profits, $\mathbb{E}[\pi_{frt}(\mathcal{J})]$, on the right-hand side of Equation (2.18) are a function of

²⁵Our identification assumption could be violated if competitors' entry into markets affects firms' fixed costs by changing local input prices (e.g., by increasing the cost of acquiring new online hubs).

the choice probabilities, $\phi_{\mathcal{J}_{frt}}$, on the left-hand side of the same equation. Because of that, a common concern related to discrete games with incomplete information is the potential for multiple equilibria (i.e., multiple solutions of Equation (2.18)), which complicates both estimation and counterfactuals.²⁶ To avoid this issue in the estimation stage, we follow [Sweeting \(2009\)](#) and estimate the parameters in two steps.

In the first step, we compute choice probabilities directly from the data. Specifically, we estimate a multinomial choice logit model using all own and competitors' drivers of variable profits, which include the demand mean utilities, δ_{jrt} , evaluated at $\xi_{jrt} = 0$, the demand shifters, w_{jrt} , the marginal costs, c_{jrt} , evaluated at $\omega_{jrt} = 0$, and the determinants of fixed costs from Equation (2.14), to predict the probability that each bundle is offered in the data. We denote the estimated probabilities by $\hat{\phi}_{\mathcal{J}_{frt}}$. We provide more details about the prediction model in Online Appendix B.3.

We use the estimated probabilities, $\hat{\phi}_{\mathcal{J}_{frt}}$, to estimate the expected variable profits, $\mathbb{E}[\pi_{frt}(\mathcal{J})]$. To do that, we randomly draw own and competitors' bundle choices, $\{\mathcal{J}_{frt}, \mathcal{J}_{f'rt}\}$, using the estimated probabilities, $\hat{\phi}_{\mathcal{J}_{frt}}$, and demand and supply shocks, ξ_{jrt} and ω_{jrt} from their respective empirical distributions. For each draw, we compute variable profits solving for the static equilibrium game given by Equation (2.11). We do this 10,000 times per market. We compute the expected variable profits for each bundle choice by integrating over all simulation draws. Finally, we fit a random forest prediction model to reduce noise and get stable predictions for bundles with low probability of being chosen. This gives us viable estimators of $\mathbb{E}[\pi_{frt}(\mathcal{J})]$ for all bundles $\mathcal{J} \in \mathcal{B}_f$ for all firms f . We denote the estimated expected variable profit by $\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J})]$.

In the second step, we use the estimated expected variable profits, $\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J})]$, and the equilibrium conditions on entry probabilities from Equation (2.18) to recover the fixed cost parameters via maximum likelihood, replacing $\mathbb{E}[\pi_{frt}(\cdot)]$ with $\hat{\mathbb{E}}[\pi_{frt}(\cdot)]$ in Equation (2.18).

Results

We present the estimated parameters in Online Appendix B.1, Table B6. We estimate a median entry elasticity with respect to own profits of 1.2, meaning that the probability of offering a given bundle increases by 1.2% for each percentage increase in expected profits. We also find that adding a new degree to the bundle costs more than continuing to offer an existing degree. For many areas of study, offering an existing degree does not have an additional fixed cost (i.e., $\kappa_{ao}^{\text{old}} = 0$), which explains why, conditional on keeping their campus or hub open, institutions keep offering degrees in 2019 that they were already offering in 2010. We also find that the costs of building new infrastructure are increasing with the distance between regions and firms' headquarters. The cost of building a new in-person campus or online hub in a region 1200 kilometers away from the institution's headquarters—the median

²⁶Multiple equilibria are more prevalent in complete-information games, where assuming a degenerate equilibrium selection mechanism is not enough for identification ([Aradillas-López, 2020](#)). Note that counterfactual analysis still requires assumptions about which equilibrium is selected.

distance in the data—is \$42 and \$18 per person living in the market, respectively. We show the full distribution of fixed costs associated with opening a new campus or hub in every region as a function of distance in Online Appendix B.1, Figure B2.

2.5 Counterfactuals

We use the model to assess the impact of introducing online education, isolating the effects of demand, pricing, and program offerings. Additionally, we explore policies that restrict online education for specific groups, examining how these measures could allow in-person programs to remain in the market while ensuring online access for those who benefit most.

The effects of the online education expansion: the role of supply and demand

Using our model, we compare the 2019 status quo, referred to as *Baseline* (BL), where online education exists, with counterfactuals where online education does not exist. To isolate the supply and demand effect, we consider increasingly flexible counterfactuals. We describe these below and summarize them in Table 2.5.

1. *No supply-side responses (NS)*: we remove online degrees while restricting institutions from adjusting tuition or program offerings. Value added under NS depends on two factors: (i) the value added differences between online degrees, in-person degrees, and the outside option, and (ii) the extent to which students switch from online degrees to in-person programs or the outside option.
2. *Price responses (PR)*: institutions are allowed to set prices optimally, though their program offerings stay constant. This scenario helps measure the effects of reduced competition, which may lead to increased prices and potentially displace students from high-quality in-person degrees.
3. *Equilibrium (EQ)*: In the final counterfactual, institutions have the flexibility to decide which degrees to offer in each market, corresponding to the equilibrium model outlined in Section 2.4. Under EQ, the availability of in-person degrees is expected to expand, potentially boosting enrollment in these programs and increasing total value added. We summarize the counterfactuals in Table 2.5. We present the algorithm we use to find the market equilibrium of each counterfactual in Online Appendix B.3.

We examine six different outcome categories: (i) the number of online and in-person degrees, (ii) total college enrollment levels, separated by in-person and online enrollment, (iii) the total value added produced in the economy, (iv) total expenditure in higher education, (v) the average price of in-person degrees, (vi) total variable profits, and (vii) revealed-preferences consumer surplus.²⁷ Detailed formulas for constructing the last five outcome measures are provided in Online Appendix B.3.

²⁷Our consumer surplus measure assumes that students make informed decisions and enroll in degrees

Table 2.5: Counterfactual exercises

Counterfactual	Description
Baseline (BL)	Free entry of online education
No supply-side responses (NS)	No online education & no supply-side responses
Price responses (PR)	(NS) + firms choose prices
Equilibrium (EQ)	(NS) + (PR) + firms choose degree offerings

Notes: This table summarizes the main counterfactuals simulated in Section 2.5.

Results are presented in Table 2.6. Under the Baseline scenario (BL) with the presence of online education, an average market has 47.8 online degrees and 41.7 in-person degrees offered. Total college enrollment reaches 5.9%, with 57.4% of students attending in person and 42.6% attending online. The total value added per enrolled student under BL is 0.137 log-points, and students spend an average of \$4211 in tuition. On the firm side, the average price of in-person degrees is approximately US\$6500 per year, and average industry profits per enrolled student are US\$1742. Finally, consumer surplus per enrolled student is \$9604 higher than in a world without higher education. These numbers compare well to the data reported in Table 2.6, Column (0). Differences arise from the fact that our counterfactuals take an average over 10,000 simulations, while the observed data correspond to a single realization of shocks.

When online degrees are removed from the market, and institutions are unable to adjust prices or alter their degree offerings (i.e., under the NS scenario), total enrollment declines by 14% (dropping from 5.9% to 5.1%). This decline underscores the role of online education in broadening access for students who might otherwise not attend college. Additionally, because online programs can attract students who might have chosen in-person degrees, restricting the supply of online education leads to a 49.8% increase in in-person enrollment (rising from 3.4% to 5.1%). This drop in overall enrollment is accompanied by a corresponding 5.6% increase and 13.6% decrease in consumer expenditure and firms' profits, respectively.

The impact on value added is ambiguous when online education is restricted. On one hand, since online degrees provide higher value added than not attending college, the reduction in total enrollment in higher education could lower overall value added. On the other hand, because online degrees generally have lower value added than in-person programs, the shift toward greater in-person enrollment could raise total value added. Overall, we observe a small increase in value added of 0.4%. Consumer surplus, measured by revealed preferences, decreases by 14% due to the lower number of available options.

When institutions are allowed to respond to the absence of online degrees by adjusting the prices of their in-person programs (i.e., under the PR scenario), tuition prices increase

that maximize their welfare. This assumption contrasts with policymakers' concerns regarding the challenges students may face in identifying low-quality degrees.

Table 2.6: The effects of the online education expansion

	DT (0)	BL (1)	NS (2)	PR (3)	EQ (4)	$\Delta_{EQ/BL}$ (5)
Number of online degrees:	72.1	47.8	0	0	0	—
Number of in-person degrees:	52.3	41.7	41.7	41.7	48.9	17.3%
Total college enrollment:	6.3%	5.9%	5.1%	5%	5.2%	-11.9%
In-person enrollment:	3.2%	3.4%	5.1%	5%	5.2%	53.4%
Total value-added:	0.131	0.137	0.138	0.136	0.142	3.4%
Total expenditure:	4134	4211	4448	4502	4590	9%
Value-added per dollar spent:	0.033	0.033	0.031	0.03	0.031	-6.3%
Av. price of in-person degrees:	5375	6530	6530	7420	6721	2.9%
Total profits:	—	1742	1504	1589	1618	-7.1%
Consumer surplus:	—	9604	8261	8107	8464	-11.9%

Notes: This table reports the value of different outcomes for each counterfactual. We report the average value across markets, weighted by market size, for seven different outcomes: the number of online and in-person degrees, total college enrollment and in-person enrollment, the total value added produced in the economy, the average price of in-person degrees, and total variable profits. Detailed formulas for constructing the last three outcome measures are provided in Online Appendix B.3. Column (0) reports the values in the original data. Columns (1)-(4) present the value for each counterfactual from Table 2.5, averaged across 1,000 simulations. Column (5) reports the average value for the percentual changes between the *Equilibrium* counterfactual and the *Baseline* counterfactual, given by $\Delta_{EQ/BL} = \frac{EQ-BL}{BL}$.

by 13.6%. This rise in prices pushes students away from higher-quality degrees, resulting in a decrease in total value added of 1%.

When institutions are allowed to adjust their degree offerings (i.e., under the EQ scenario), they expand the supply of in-person options by 17.3%, increasing from an average of 41.7 degrees to 48.9 degrees per market. Moving from the baseline counterfactual to the full equilibrium scenario, total enrollment decreases by approximately 11.9% (from 5.9% to 5.2%), while in-person enrollment rises by 53.4% (from 3.4% to 5.2%). In terms of value added, the equilibrium scenario shows an improvement relative to the baseline, with an increase of 3.4%, reaching 0.142 log-points. However, total expenditure increases by 9%, decreasing the value added per dollar spent by 6.3%. Additionally, the average tuition price for in-person degrees rises by 2.9% relative to the baseline counterfactual. However, the increased markups are insufficient to fully offset the profit losses resulting from the absence of online degree offerings, leading to an average profit decrease of 7.1%. Finally, consumer

surplus, measured by revealed preferences, decreases by 11.9%.

These results highlight the benefits and risks associated with expanding online education. On the one hand, the availability of online options increases college enrollment, reduces total expenditure, and increases consumer surplus, defined as the area below the demand curve. On the other hand, the online expansion diverts students from high value-added in-person alternatives, reducing total value added. Even if students are not fully informed about the quality of the degrees they choose, online education's flexibility and cost savings may offset the decline in total value added.

Unpacking value-added changes by age groups

The overall changes in total value added from the previous subsection mask important heterogeneity between those who benefit and those who lose out. Older individuals, who might not attend college without online education, are the primary winners of its expansion. In contrast, younger individuals are more often diverted from higher-quality, in-person programs. Diversion from those degrees can be particularly detrimental if online education induces exit or deters entry of in-person degrees that young individuals would prefer, even in the presence of online options.

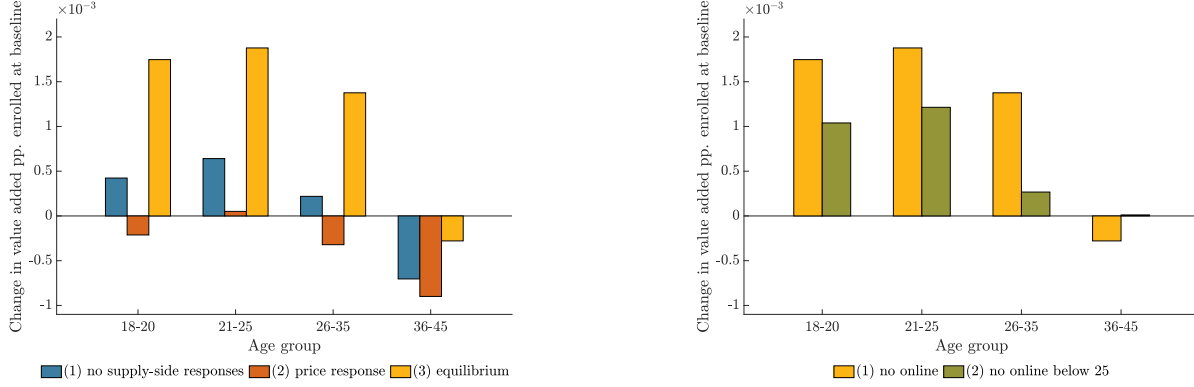
Figure 2.6 illustrates the heterogeneous effects on value added by cohort, showing changes in total value added for each age group relative to the baseline counterfactual. When online education is restricted and supply-side adjustments are not allowed, the 18-25 age cohort experiences a slight increase in value added, while the 36-45 cohort faces significant decreases. Allowing institutions to adjust prices results in value-added reductions across all cohorts. Only in the equilibrium counterfactual, where institutions can adjust their degree offerings, the 18-35 age group sees substantial gains in total value added.

In Online Appendix B.1, Figure B3, we show heterogeneous results for total enrollment, value added per dollar spent in higher education, and revealed-preferences consumer surplus. We find that the expansion of online education benefits all age groups in these three dimensions, with large benefits for individuals above 25 years old and small benefits for individuals 18-20 years old.

Targeted policies

Next, we explore hypothetical policies designed to harness the benefits of online education for older students while preserving value added for younger cohorts. Specifically, we examine a policy restricting online education access for students aged 18-25. By directing younger cohorts toward in-person programs, this policy leaves sufficient profits for in-person degrees, enabling them to coexist with online options. Consequently, younger students can pursue high-value-added in-person degrees, while older students continue to benefit from the expanded access offered by online education. Our main results, shown in Figure 2.6, indicate that, compared to the baseline counterfactual, this targeted policy raises value added across

Figure 2.6: Changes in total value added by age group under each counterfactual



(a) Value added changes decomposition

(b) Targeted ban to 18-25 year old students

Notes: This figure presents changes in total value added for different age groups under different counterfactuals. In panel (a), we show changes in value added relative to the *Baseline* counterfactual for all counterfactuals presented in Table 2.5. In panel (b) we show the changes in value added relative to the *Baseline* counterfactual for the *Equilibrium* counterfactual and an additional counterfactual in which we restrict access to online education for students aged 18-25. In both panels, the total value is divided by the number of enrolled students in the *Baseline* counterfactual.

all cohorts. In Online Appendix B.1, Figure B3, we show that restricting online education in this targeted way has a smaller negative effect on other outcomes of interest.

2.6 Conclusion and discussion

In this paper, we examine the equilibrium effects of the rapid growth of online college education in Brazil, highlighting two key findings. First, the expansion of online degrees creates a dual impact: it broadens access to higher education for new students but simultaneously diverts potential students away from in-person options. Second, the increased availability of online degrees intensifies competition, which drives down prices for in-person programs and deters new in-person entrants.

Using an equilibrium model of supply and demand, we quantify the impact of online education on total value added and find significant heterogeneity across age groups. A ban on online degrees would benefit younger cohorts by preserving access to higher-quality, in-person programs, while older cohorts, who benefit from the flexibility and availability of online options, would face reduced college access. A targeted policy restricting online education for younger students would increase overall value added across all cohorts compared to an unrestricted baseline.

Our analysis highlights how introducing lower-cost alternatives can reshape competition

and market structure, with potential adverse effects in settings with imperfect information. Consumers may unknowingly substitute lower-quality online degrees for traditional options, particularly as in-person providers become less available in the market. Policymakers could help counter these effects by restricting online education access for groups that would gain more from traditional degrees, thereby sustaining demand for incumbent high-quality institutions and preventing them from leaving the market.

While this paper addresses critical aspects of online education's recent expansion, several open questions remain. First, this study assumes that degree value added is fixed and policy invariant, though a surge in online degrees could saturate the labor market, reducing returns to college education. Modeling college education as an equilibrium of supply and demand interacting with labor market outcomes presents an exciting path for future research. Second, technological advancements could further enhance the cost-effectiveness and quality of online education, possibly altering some conclusions of this study. Lastly, non-pecuniary benefits—such as trust, networking, and social skills—are more effectively developed in in-person settings, where students engage in face-to-face interactions. These social benefits may be harder to replicate in online formats, suggesting that traditional college experiences offer unique non-monetary advantages, especially valuable for younger students, which could further support this study's findings.

Chapter 3

The Role of Housing Supply in Shaping Affordability: A Quasi-experimental Approach

3.1 Introduction

Housing has become increasingly unaffordable in many cities around the world (UN, 2023).¹ Between 2010 and 2022, real housing prices increased by an average of 27% globally, with some of the highest growth rates observed in Iceland (103%), Estonia (97%), New Zealand (97%), Chile (95%), Turkey (91%), and Canada (90%) (BIS, 2024).² Several major cities now exhibit exceptionally high ratios of prices to household income, including Hong Kong (16.7), Sydney (13.8), Vancouver (12.3), San Jose (11.9), and Los Angeles (10.9), rendering them “impossibly unaffordable” for their residents (Cox and Pavletich, 2024).

To curb the recent global housing affordability crisis, governments have increasingly responded by subsidizing housing. Common approaches include tax incentives, subsidies that lower annual mortgage payments, homeownership vouchers, tenant-based vouchers, and project-based assistance.³ Assessing the costs and benefits of these policies requires understanding housing fundamentals, particularly housing supply elasticity.

¹The UN estimates that 96,000 new affordable homes must be built each day to accommodate the estimated 3 billion people requiring adequate housing by 2030.

²The UN reports that the pandemic, along with geopolitical conflicts, increasing land demand, and other factors, has led to a global rise in housing prices, evictions, and housing deficits, making it increasingly difficult for households to afford basic necessities like housing.

³Some examples include: for tax incentives, there is the U.S. mortgage interest deduction (\$200 billion a year) and the Canadian tax credit for first-time home buyers; for mortgage subsidies, there is the Help-to-Buy program in the United Kingdom; for homeownership vouchers, there are *Mi Casa, Mi Vida* in Brazil (2009), *Subsidio Habitacional* in Chile (2011), and *Mi Casa Ya* in Colombia (2023); for tenant-based vouchers, there are Section 8 in the U.S. and the rental voucher in Chile (2014); and for project-based policies, there is the Low-Income Household Tax Credit in the U.S.

Measuring housing supply elasticity is challenging due to the difficulty in isolating supply-side from demand-side shifts. Additionally, there is significant variation in supply conditions across cities and neighborhoods, such as supply constraints, topography, and land availability. Recent literature has attempted to identify this elasticity by exploiting long-term labor demand shocks and geographical constraints as instruments for demand (Saiz, 2010; Baum-Snow and Han, 2024). However, labor demand shocks may be influenced by local supply factors, and supply constraints may correlate with demand, raising concerns about the validity of these instruments (Davidoff, 2016). Due to the lack of quasi-experimental evidence, the identification of housing supply elasticity remains unclear.

This paper addresses the aforementioned challenges of estimating housing supply elasticity by utilizing quasi-experimental variation in the case of Chile's *Subsidio Habitacional* program. Implemented in 2011, this policy was aimed at promoting homeownership among low- and middle-income households. The program enables eligible households to purchase properties up to a certain price cap, while the government provides a voucher to cover part of these costs. Depending on their income, applicants can apply to three different tiers, named Title 0, Title 1, and Title 2. The size of the subsidy is determined by the property price and the voucher type. Over its first 10 years, the program allocated 249,163 vouchers, generating a significant demand shock in Chile's housing market. By leveraging detailed administrative data on property transactions and the allocation of the subsidy, this study offers an identification strategy to estimate the housing supply elasticity.

Thus, this paper seeks to answer a central question: What is the short-term response of the housing supply to demand shocks? Relatedly, the paper also explores: (1) how fast do new constructions respond to an increase in prices? (2) How fast do rental units convert to homeownership when there is an increase in demand for owner-occupied housing? and (3) what is the effect of the homeownership voucher program on Chilean housing prices?

To implement the empirical analysis, we linked detailed information on 1,026,123 applicants and recipients of the voucher with novel administrative data of property transactions in Santiago de Chile from 1998 to 2020. Both datasets are georeferenced, so we linked them at the Census Zone (CZ) level.⁴ The resulting dataset provides comprehensive insights into the housing market in Santiago, including the residential location of voucher applicants and the evolution of prices and transactions at the local level over time. Figure C1a shows a map with a random sample of geolocated property transactions and the number of voucher recipients by CZ. Figure C1b shows the spatial correlation between transactions and the housing quality at the neighborhood level (measured by the ISMT index).⁵ These figures suggest that housing supply has grown more in supply-constrained markets (such as the city's downtown) compared to suburban areas, indicating a positive correlation between supply constraints and demand growth.

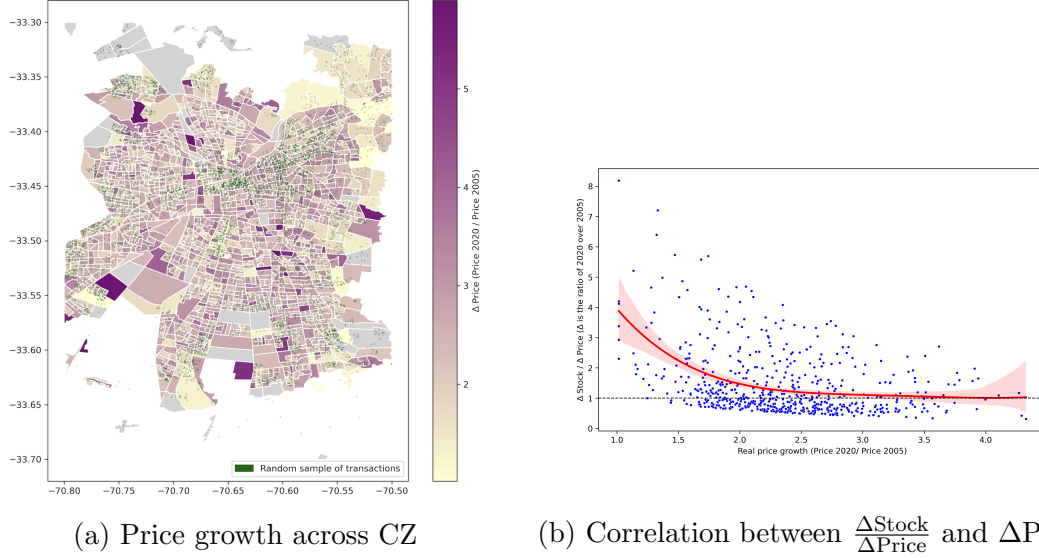
⁴The Metropolitan Region of Santiago is divided into 1,865 CZs, each with an average of 3,813 individuals.

⁵As explained in Section 3.2, the Socio-Territorial Index (ISMT) measures housing quality and socio-economic conditions.

The challenge in estimating housing supply elasticity is illustrated in Figure 3.1. Panel (a) shows that in Santiago transactions have been concentrated in the Central Business District (CBD) and high-income areas in the northwest neighborhoods of the city. However, the CZs with the largest price increases are more concentrated in suburban and intermediate regions. Panel (b), on the other hand, suggests that areas with significant price increases have experienced relatively small increases in housing stock, which is indicative of a low elasticity of supply. This leads to the question of how quickly the supply has responded to demand growth across neighborhoods, as well as why some suburban areas show a relatively high increase in prices.

In Section 3.4, we analyze how the housing supply responds to demand shocks. Using historical transaction data for different types of properties in Santiago, we build an index for local exposure to each voucher type. Additionally, using data on subsidy recipients from 2011 to 2019, we estimate the demand shock at the neighborhood level. These two metrics are then used to construct a shift-share instrument for housing demand. While OLS estimates appear to be biased by a combination of demand-side and supply-side effects, the 2SLS strategy estimates a housing supply elasticity of 0.543. This result is robust across other price measures, such as price per square foot and estimated appraisal value, with elasticities of 0.85 and 0.328, respectively. All measures are below 1, indicating a low supply response to demand shocks. Back-of-the-envelope calculations suggest an average price effect ranging from 6.29% to 10.13% as a consequence of the policy.

Figure 3.1: Growth in housing prices and stock across Santiago’s CZ (2005-2020)



Note: This figure displays the growth in housing prices and stock by Census Zone (CZ) in Santiago from 2005 to 2020. Panel (a) shows the average housing prices in real terms for 2005 and 2020 across CZs, with missing information colored in gray. A sample of property transactions between 2005 and 2020 is plotted in green. Panel (b) illustrates the correlation between the ratio of stock growth to price growth ($\frac{\Delta \text{Stock}}{\Delta \text{Price}}$) and price growth (ΔPrice). The dashed line represents an elasticity of 1, indicating proportional growth between stock and price. These are long-term elasticities obtained through a simple ratio and should not be interpreted as causal relationships. The polynomial fit is of the 4th degree, but a negative correlation is also observed in a linear polynomial fit.

We also find that new constructions are more than twice as responsive to price increases compared to rental units transitioning to owner-occupied housing. This is particularly relevant for low- and middle-income suburban neighborhoods, where the share of new construction is relatively small compared to central areas. This lack of responsiveness suggests that developers are not converting enough rental properties into owner-occupied homes, restricting opportunities for families aiming to become homeowners. Consequently, the demand for new constructions intensifies, further driving up prices. We argue four mechanisms could explain why developers focus on new construction in more central areas: higher profitability, significant variation in zoning policy in hyper-dense areas, better infrastructure closer to the CBD, and evidence of a high rate of redevelopment.

In Section 3.5, we slightly modify the main empirical strategy by implementing an Event Study of the vouchers’ exposure to housing prices. This exercise has three objectives: (i) to validate the empirical strategy from the previous section by testing for parallel trends, (ii) to directly estimate the effect of the voucher on housing prices, providing a robustness check on the housing supply elasticity estimate, and (iii) to assess heterogeneous effects of the policy. First, we find that the parallel trends assumption is generally satisfied, suggesting that confounders do not seem to correlate with the treatment. Second, we find a positive

and consistent effect on housing prices, with an average increase of 8.87% from 2011 to 2019. Coefficients of similar magnitude are found by [Carozzi et al. \(2024\)](#) in the context of homeownership subsidies in the United Kingdom. Furthermore, our estimates align with an average elasticity of 0.62, which is just slightly higher than the estimate in the previous section. These results suggest that a large share of the subsidies is capitalized into higher prices, benefiting previous owners and developers. Lastly, we find that neighborhoods with high population density, low share of new properties, and those farthest from CBD experience larger price effects. This aligns with the characteristics of neighborhoods more exposed to the policy and reflects the low elasticity of supply in low- and middle-income areas.

This study contributes to multiple strands of literature. Firstly, it adds to the body of work analyzing housing supply responses. Recent evidence in the U.S. shows that there is an important variation in the housing supply elasticity *between* and *within* cities ([Saiz, 2010](#); [Cosman et al., 2018](#); [Blouri et al., 2023](#); [Baum-Snow and Han, 2024](#)). [Saiz \(2010\)](#) finds elasticities ranging from 0.6 to 5.45, depending on a city’s supply constraints, while [Blouri et al. \(2023\)](#) estimates elasticities between 0.39 and 2.25 across U.S. counties. On the other hand, [Baum-Snow and Han \(2024\)](#) finds a housing price elasticity, in terms of new units, ranging from 0.2 to 0.9 in metropolitan areas. Understanding the spatial variation in housing supply elasticity is key to estimating the consequences of neighborhood-specific housing demand shocks ([Calabrese et al., 2011](#); [Couture et al., 2023](#)). Our estimate for Santiago (0.543) is closer to the bottom range of previous estimates, suggesting a low elasticity of supply. Nevertheless, our analysis reflects a short-term supply response, while the literature analyzes long-term variation. Additionally, we argue that neighborhoods with low constraints can still have low-supply responses when developers do not have enough incentives to build new housing. Lastly, this research contributes to the literature by exploring the supply response in a middle-income country, where housing market dynamics and policy impacts may differ significantly.

Secondly, this study contributes to the literature evaluating the effects of housing policy. In the context of tax incentives, there is extensive evidence suggesting that mortgage interest deductions (MID) in the U.S. disproportionately benefit high-income individuals and developers ([Sinai and Gyourko, 2004](#); [Poterba and Sinai, 2008](#); [Hilber and Turner, 2014](#); [Rappoport, 2016](#); [Davis, 2019](#)). On the side of owner-occupied subsidies, policies aimed at reducing mortgage payments (such as Help-to-Buy in the U.K.) seem ineffective in unaffordable areas where the housing supply is inelastic, leading to increased housing prices and higher profits for developers without stimulating new construction ([Carozzi et al., 2024](#)). The evidence we provide is consistent with this result, while in our case, we directly estimate the housing elasticity of supply. On the other hand, rental vouchers (e.g., Section 8 and MTO) have been shown to be valuable tools for reducing housing costs and providing long-term benefits for recipients ([Kling et al., 2007a](#); [Ludwig et al., 2013b](#); [Chetty et al., 2016b](#)). While rental vouchers do not appear to have a significant impact on housing prices on average, [Eriksen and Ross \(2015\)](#) find the largest price increase in cities with inelastic housing supply. Finally, project-based policies like the Low-Income Housing Tax Credit (LIHTC)

have been criticized for their cost-effectiveness, with around half of the subsidies captured by developers and a quarter dissipated in fixed costs (Soltas, 2024).

This study also contributes to ongoing policy debates. Our results suggest that housing subsidies may not be an efficient solution to address the housing affordability crisis and could even exacerbate the problem for unsubsidized households. The effectiveness of housing assistance largely depends on the responsiveness of the housing supply to demand shocks. Therefore, complementary measures, such as supply-side incentives for new construction in low- and middle-income areas, should be considered.

This paper is organized as follows. Section 3.2 describes the Chilean homeownership voucher. Section 3.3 presents the data used in the study. Section 3.4 outlines the empirical strategy and estimates the housing supply elasticity. Section 3.5 estimates an Event Study to validate the identification strategy and to directly estimate the effect of the policy on housing prices. Finally, Section 3.6 presents the main conclusions of the study.

3.2 The Chilean Homebuyer Voucher

This section provides an overview of the Chilean housing subsidy program, detailing its structure and evolution over time, which are essential for understanding the paper’s empirical strategy.

General context

In 2011, the Chilean government established a homeownership voucher program titled *Subsidio Habitacional* (SH or DS01). This housing subsidy was designed to finance the acquisition of new or used homes, either in urban or rural areas, to be used as housing for the recipient and their family.⁶ The government covers a percentage of the property’s value up to a certain price cap, but households are required to have savings before applying to the program.

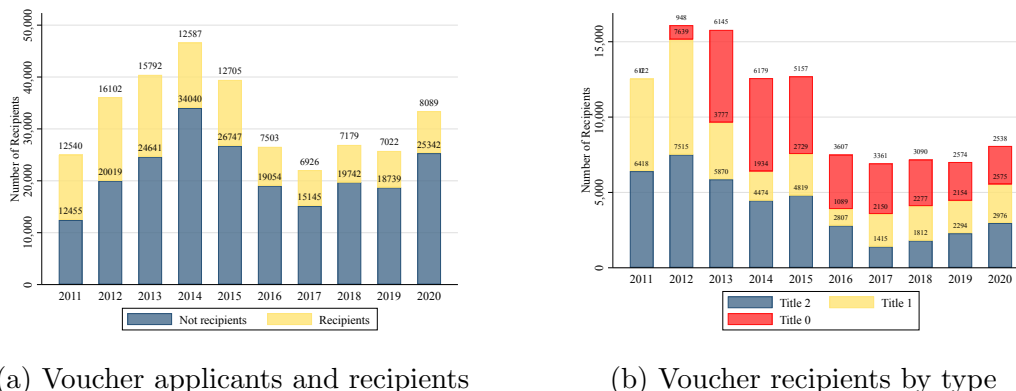
The homeownership voucher was introduced as a housing subsidy targeted at the “middle class.” In Chile, the homeownership rate stands at 66%, which is lower than the average of the OECD (71%) and the European Union (75%).⁷ As in many other countries, there is a strong correlation between homeownership and income. During the first ten years of the program, the government spent nearly \$526 million annually (0.2% of current GDP), distributing 249,163 vouchers. Assuming an average household size of four, this implies that 5.5% of the Chilean population was directly impacted by the policy. Figure 3.2 shows the

⁶Supreme Decree 01 (DS01), available at: <https://www.bcn.cl/leychile/navegar?idNorma=1026260>

⁷According to the Housing Tenures Report (2024) by the OECD, this figure is comparable to those observed in countries such as Mexico, Canada, the United Kingdom, the United States, France, the Netherlands, Sweden, and Korea, which range between 58% and 70%. The report is available at <https://www.oecd.org/els/family/HM1-3-Housing-tenures.pdf>.

voucher allocation over time. Panel (a) shows the number of applicants and recipients, while panel (b) indicates the recipients of each voucher type over time.

Figure 3.2: Voucher allocation over time



Note: The figure presents the distribution of voucher applicants and recipients over time, from 2011 to 2020. Panel (a) displays the number of voucher applicants, differentiating between those who received a voucher (“recipients”) and those who did not (“not recipients”). Panel (b) breaks down the recipients by the type of voucher received.

Voucher structure

Initially, the subsidy was divided into two tiers (Title 1 and Title 2), but at the end of 2012, the government introduced a new tier focused on low-income individuals, called “Title 0”. While there were no significant changes in the voucher structure over time, subsidies and property caps were increased for each voucher type in 2015. Table 3.1 outlines the main characteristics of the vouchers. We present the information for the year 2015 to explain each component more easily.

The first tier, called Title 0, focuses on low-income individuals with a maximum monthly income of \$938. This voucher enables the purchase of properties up to \$37,500, providing a subsidy of up to \$18,750, while requiring savings of at least \$1,125. For Title 1, the maximum income requirement is \$1,500. This voucher allows the purchase of properties up to \$52,500, with a maximum subsidy set at \$19,350, and requires savings of \$1,500. Lastly, for Title 2 applicants, the maximum income requirement is \$2,250. They can purchase a property worth up to \$82,500, and the subsidy cap is set at \$13,125. The saving requirement for this group is set at \$3,000. Additionally, it is worth noting that Title 1 and Title 2 require mortgage credit pre-approval from a financial entity, while Title 0 does not.

Table 3.1: Vouchers eligibility criteria and structure from 2015 (in US dollars)

	Title 0	Title 1	Title 2
Maximum income	\$938	\$1,500	\$2,250
Saving requirements	\$1,125	\$1,500	\$3,000
Credit requirement	No	Yes	Yes
Subsidy rule	$\min\{P; \$18,750\}$	$\$27,188 - 0.375 \times P$	$\$27,188 - 0.375 \times P$
Maximum subsidy	\$18,750	\$19,350	\$13,125
Minimum subsidy	-	\$7,500	\$4,500
Cap in property value	\$37,500	\$52,500	\$82,500

Note: The table outlines the eligibility criteria and subsidy rules for three voucher tiers based on income, savings, credit requirements, and property value caps. The subsidy rule indicates the relationship between subsidies and property prices, which can be visually examined in Figure 3.3b.

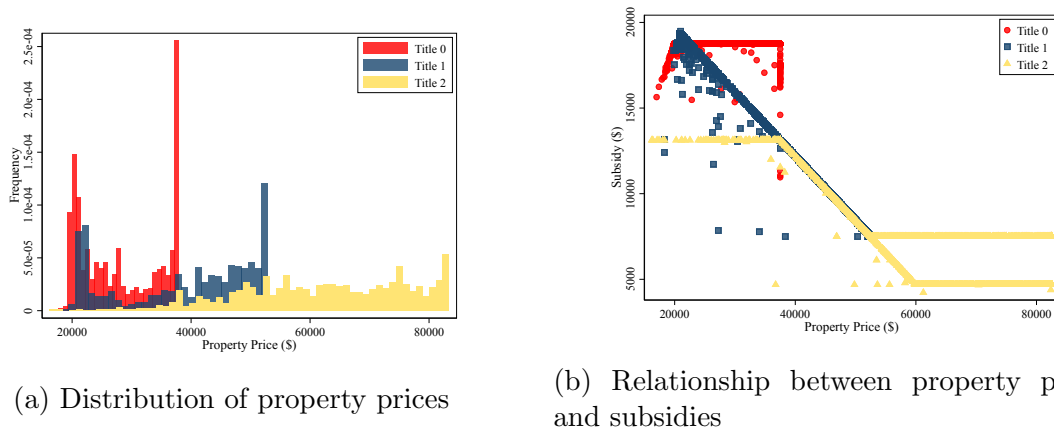
On top of that, the program stipulates that properties purchased in “priority” areas would receive a minimum subsidy of \$7,500.⁸ Figure 3.3 provides a clearer understanding of the subsidy structure. Panel (a) shows the distribution of property prices for each voucher type, while panel (b) illustrates the empirical relationship between subsidies and prices. For all three voucher types, there is significant bunching at the property price cap. It can be observed that subsidies are the largest for Title 0 and the lowest for Title 2, with a common minimum of \$7,500 for all vouchers (priority areas), and a specific minimum of \$4,500 for Title 2 when the property price exceeds \$60,000.

Thus, the monetary amount of each voucher was designed as a function of the value of the purchased property and the voucher type.⁹ The subsidy amount for Title 0 is the largest compared to the other voucher types, but it also restricts purchases to very low-priced homes. When the policy was implemented, property prices in Santiago averaged around \$70,000, meaning only recipients of Title 2 vouchers were able to purchase an “average-priced” property in Santiago.

⁸In Chile, priority areas of development are defined based on factors such as poverty levels, income disparities, economic productivity, infrastructure needs, and social equity. Regions with lower income, insufficient infrastructure, or potential for economic growth are targeted, with a focus on improving living standards, addressing inequality, and promoting sustainable development.

⁹This structure was consistent across all regions of the country, with the exception of a few isolated areas (e.g., Chiloé Island, Easter Island, Juan Fernández Island, and the Regions of Aysén and Magallanes), where the subsidies and property caps were relatively higher. For simplicity, we do not analyze these areas, which also represent less than 2% of the Chilean population.

Figure 3.3: Prices of purchased properties by voucher type and price-subsidy function



Note: This figure visualizes the structure of the subsidy. Panel (a) shows the distribution of prices of purchased properties by each voucher type. The right panel depicts the empirical relationship between prices and subsidies.

Application process and the use of the subsidy

To apply for the subsidy, households must be registered in the *Ficha de Protección Social* (FPS), which helps the Chilean Ministry of Social Development (MDS) identify eligible households for subsidies and social programs. Households are required to be below the 90th percentile of the FPS vulnerability score to be eligible to apply.¹⁰ Applicants must also demonstrate some borrowing capacity, as described in Table 3.1.

After the application process, the Chilean Housing Department (MINVU) rations vouchers through two main mechanisms. First, applicants are ranked according to a vulnerability score that considers factors such as household size, single-parent status, the presence of children or elderly individuals, the people-to-bedrooms ratio, and savings. Second, the central government allocates a budget to each of Chile's 16 administrative regions. Then, when demand for vouchers exceeds supply—which has occurred in every application call since the program's inception—each region distributes its budget based on the vulnerability ranking system until the funds are exhausted. As a result, in each application call and region, a threshold in the vulnerability index determines who receives a voucher.

When applicants receive the voucher, they take an average of two years to purchase a property. After getting it, recipients must reside in the property and retain ownership for at least five years (Clause No. 39 of the law). The prohibition against selling or renting the property is highly enforced. The property title is registered with the Conservador de

¹⁰The FPS score is based on household members' income-generating capacity and economic needs. The household is defined as those who share a home and a food budget.

Bienes Raíces (Chilean Property Registrar), and any violation of the law could result in the government reclaiming the subsidy or the property.¹¹

Chile’s 2014 rental voucher

In 2014, Chile introduced a rental voucher program named *Subsidio al Arriendo*, which was smaller in scale than the homeownership voucher. Over five years, the government spent \$350 million, 87% less than the annual spending on homeownership subsidies. The rental voucher is targeted at low-income individuals and provides \$7,500, which is 60% lower than the homeownership voucher for comparable individuals (nearly \$18,750 for Title 0). [Selman \(2022\)](#) found that the rental voucher did not affect applications for the homeownership voucher.

3.3 Data

To study the urban response to homeownership vouchers, we work with two main datasets. First, we use the administrative dataset of the applicants and users of the subsidy ([MINVU, 2024](#)), which allows us to understand the characteristics of the applicants and their residential addresses before and after using the subsidy. Secondly, we employ data on property transactions in Chile from 1998 to 2020, which gives information on the housing market before and after the implementation of the SH ([TocToc, 2024](#)).

The administrative data of the subsidy is administered by the MINVU. This dataset provides detailed information on all the applicants and users of the homeownership subsidy program from June 2011 to December 2020. The database includes 1,026,123 applicants in Chile’s 16 regions, where 322,369 (31,4%) belong to Santiago’s Metropolitan Region. From the applicants in Santiago, 106,445 individuals (33%) have been awarded the homeownership subsidy, representing \$1,511 million of monetary transfers through vouchers and 62,542 individuals who used the voucher to purchase a property in 10 years.¹²

The database has an ID for each applicant, and provides detailed information about them and their family at the moment of the application, such as gender, nationality, family’s vulnerability index, number of people living in their current home, residential address,

¹¹However, this is rare, as banks review property titles before issuing mortgages. Additionally, MINVU enforces compliance through random inspections and local community reports of misuse. In 2019, the government launched the “Te cache” program to identify misused subsidized properties. Reports mainly concerned illegal rentals or commercial use of subsidized homes ([MINVU, 2019](#)). Between 2014 and 2018, there were 2,309 reports of improper use, affecting 2% of voucher recipients.

¹²Some reasons that can explain why there is no full use of the subsidy are the following. Firstly, individuals have three years to decide where to purchase a property, which can generate a lag in its use. Secondly, they may apply for more than one type of subsidy in different years, so if they receive more than one, they can only use one. Finally, individuals may not have enough savings to purchase properties even though they receive the voucher, so they may decide not to purchase.

important financial information (declared income, savings, credits, among others), among others. It also includes the date of the application, the type of the subsidy (Title 0, 1, or 2), their score in each application, their results (win or lose), and if they used the voucher. In the case they have purchased a property with the voucher, the dataset reports the residential address of the new home, the property value, and the subsidy amount, among others.

The information on transactions was obtained by an agreement with the Chilean private company TocToc.¹³

3.4 Housing supply elasticity

What is the supply response to demand shocks? Do we expect to see higher prices after the implementation of a housing subsidy? In this section, we provide quasi-experimental evidence of the housing supply elasticity in the context of Santiago de Chile. To do so, we exploit the demand shock generated by the Chilean homeownership voucher (SH). We estimate a 2SLS panel regression with fixed effects, with a shift-share instrument to isolate supply-side and demand-side variation. The analysis relies on Santiago's transaction data for 1,779 CZs from 1998 to 2019, along with information on the number of voucher recipients by neighborhood and year.

Identification strategy

The housing supply elasticity is the change in housing stock when there is an increase in prices ($\eta = \frac{dQ}{dP} \frac{P}{Q}$). Identifying η requires isolating changes in supply from changes in demand, since equilibrium prices combine both demand-side and supply-side effects. Recent literature (Saiz (2010); Blouri et al. (2023); Baum-Snow and Han (2024)) has leveraged long-term labor demand shocks, also known as “Bartik shocks” (Bartik, 1991a), to isolate these effects. However, some evidence has pointed out that Bartik shocks may not be good instruments for housing demand as they are highly correlated with supply constraints (Davidoff, 2016). The identification strategy of this paper, in contrast, relies on short-term housing demand shocks caused by the introduction of the Chilean homeownership voucher. Therefore, this paper seeks to provide one of the first quasi-experimental studies of the housing supply elasticity.

Based on the previous literature, the inverse of the housing supply can be written as Equation 3.1:

$$\ln P = \frac{1}{\eta} \ln Q^s + \mathbf{X}'\gamma + \epsilon \quad (3.1)$$

where P is the housing price index, Q^s is the supply quantity (measured as housing stock), $\mathbf{X}$ are observable characteristics, and ϵ includes all the unobservables. Since $Q^s = Q^d$ in equilibrium, the goal is to find an instrument that captures only demand-side variation, so that η can accurately identify the response of housing stock to an increase in prices.

The approach in this paper is to use the structure of the subsidy to define the neighborhood's exposure to each type of voucher. By interacting these measures of exposure with the allocation of vouchers in each neighborhood, we estimate a shift-share instrument for the local demand for owner-occupied housing. This instrument enables us to isolate supply responses and therefore to estimate the short-term housing supply elasticity.

The main identification assumptions of the instrumental variable approach are the standard ones: (a) the instrument is relevant, and (b) it affects prices only through its impact on housing demand. In this section, we provide evidence supporting both assumptions. A potential threat to the identification is that the instrument may not be valid if the demand shift is correlated with unobserved factors that also affect supply. We partially deal with this by including neighborhood and time fixed effects in the main specification. These variables control, for example, for supply constraints that do not vary over time. In Section 3.5, we provide additional evidence of the exogeneity of the shock. By structuring the analysis into an Event Study framework, we test the existence of parallel trends between groups of different intensities of the treatment before the implementation of the policy.

We begin by describing the shift-share instrument, followed by an emphasis on the estimation and the results of the empirical strategy. Finally, we briefly explore the mechanisms that could explain the primary findings.

Shift-share instrument

To implement a shift-share framework, it is necessary to define which are the measures of exposure (also known as “shares”) and the shifts of housing demand.

To define neighborhoods' exposure to the policy, we use the fact that each of the vouchers specifies a set of properties that can be purchased based on their prices (see Table 3.1). This allows us to calculate the mass of properties in each neighborhood that are eligible for purchase with each voucher type. To achieve this, we construct an index of property prices at the neighborhood level just before the policy was introduced in 2011. This approach also prevents the issue of using the outcome variable to determine treatment exposure. We use transaction data from 1998 to 2010 to identify prices in each residential location. Leveraging transaction-level data, we run a regression of prices on housing characteristics and year fixed effects. Using these estimates, we predict the property price of all properties for the year 2010. In simple terms, this methodology adjusts all prices to be comparable at 2010 levels, accounting for housing market characteristics such as local real price growth. Figure C4 plots the evolution of average prices over time, and the estimated indicator for prices in 2010. Intuitively, the 2010 price index should be relatively constant over time. Additionally, Figure C5 reports the distribution of property prices in 2010 in Santiago's CZs and the average price at that moment (nearly \$70,000).

With the price index established, we calculate the number of properties eligible for purchase under each type of voucher in each CZ. The eligibility range is set as follows: \$18,750

to \$37,500 for Title 0, \$37,500 to \$52,500 for Title 1, and \$52,500 to \$82,500 for Title 2.¹⁴ Then, we estimate the shares (exposure) by dividing the number of eligible properties of type k by the total number of properties in each CZ. Equation 3.2 indicates the formula for estimating the shares for voucher-eligibility type k in each CZ i :

$$\omega_{ik}^{2010} = \frac{\sum_{n=1}^{N_i} \mathbb{1}\{\underline{P}^k < \hat{P}_n^{2010} < \overline{P}^k, i \in I\}}{\sum_{n=1}^{N_i} \mathbb{1}\{i \in I\}} \quad (3.2)$$

where the numerator sums all properties in neighborhood i whose adjusted 2010 prices fall within the bounds of eligibility for voucher k . The denominator sums all properties in the specific CZ. Figure C6 shows the CZ exposure to each voucher type (panel a) and the total exposure to the policy (panel b), which are a visual representation of the previous formula.

The next stage is defining the housing demand shocks. To do this, we use the administrative dataset of the subsidy allocation from 2010 to 2019. We calculate the number of voucher recipients living in each neighborhood k and year t .

We opt to use the original residential locations of voucher recipients rather than their new residences for multiple reasons. First, the number of recipients in each neighborhood-year is influenced by the quasi-random nature of the application process, particularly around the household vulnerability threshold (see rationing process in Section 3.2). In contrast, the use of the voucher is endogenous to market conditions, as individuals can choose *when* and *where* to purchase, which could bias the instrument. Second, housing demand shocks are not solely a reflection of current purchases but also encompass potential future demand. In other words, we assume the market responds in year t based on the exposure to the policy, which includes future potential buyers. Moreover, approximately 19% of voucher recipients purchase a property in the same Census Zone (CZ) where they previously resided, and nearly 51% remain within the same municipality. This strong correlation between origin and destination residential locations suggests the use of voucher allocation data is preferable, as it provides a larger sample size than voucher usage data.

Having established the housing demand shifts, we can now calculate the shift-share instrument. The instrument is a weighted summation of the housing demand shocks, where the weights are the exposures to each type of voucher. Equation 3.3 illustrates how the shift-share instrument is calculated.

$$S_{it} = \sum_{k=1}^3 \omega_{ik}^{2010} \times s_{kt} \quad (3.3)$$

¹⁴Even though Title 1 and Title 2 allow the purchase of properties starting from \$18,750, Figure 3.3 Panel (a) shows a strong correlation between voucher use and property prices. As a result, very few properties under \$52,500 are purchased with Title 2, and similarly, few properties below \$37,500 are bought with Title 1.

where s_{kt} represents the number of vouchers type k allocated in year t , ω_{ik}^{2010} is the share of eligible properties for voucher k in CZ i , and the summation goes over the three type of vouchers (Tile 0, 1 and 2).

The structure of the instrument suggests that when a neighborhood has a large share of eligible properties for voucher type k and experiences a demand shock for those properties, we should observe a corresponding shock in housing demand. For example, if a CZ has a large share of properties priced below \$37,500, we would expect a greater demand shock only if there is a significant number of recipients of Title 0 vouchers. In contrast, a comparable neighborhood with few or no Title 0 recipients should not experience a demand shock. Similarly, a neighborhood with many Title 0 voucher recipients but with property prices typically exceeding the \$37,500 cap would not see a housing demand shock either. The same logic applies to Title 1 and Title 2.

For the validity of the shift-share approach, we test relevance and exogeneity assumptions through different regressions and exercises. Related to the relevance condition, Table C3 (columns 4 and 5) and Figure C7 (panels d and e) show the correlation between the instrument and the logarithm of housing stock and property transactions. Those correlations are positive and significant at 10% and 1% of statistical significance, respectively. Additionally, the instrument reports a positive and significant correlation with different measures of prices, which represents the reduced form regressions. Regarding the exogeneity assumption, in this section, we only deal with it by including fixed effects at both the neighborhood and year levels. This strategy helps to control for differences in supply constraints, local amenities, productivity, and other factors, thereby reducing omitted variable bias. We assume throughout that any fundamental differences in supply constraints across neighborhoods are time-invariant and are therefore captured by the neighborhood fixed effects. In Section 3.5, we provide a more robust test by checking for pre-trends to support the exogeneity assumption.

Estimation and main results

To estimate the housing supply elasticity, we slightly modify Equation 3.1 into the following two-stage (2SLS) model:

$$\ln P_{i,t} = \beta \ln \hat{Q}_{i,t} + \mathbf{X}'_{i,t} \gamma + \mu_i + \lambda_t + \varepsilon_{i,t} \quad (3.4)$$

$$\ln Q_{i,t} = \theta \ln S_{i,t} + \mathbf{X}'_{i,t} \pi + \phi_i + \psi_t + v_{i,t} \quad (3.5)$$

where $\ln P_{i,t}$ denotes the logarithm of prices in CZ i at time t , and $\ln \hat{Q}_{i,t}$ is the predicted logarithm of the housing stock, instrumented using a shift-share variable. The model includes control variables $\mathbf{X}_{i,t}$ to account for other factors influencing prices, μ_i neighborhood fixed effects, and λ_t time fixed effects. The error term $\varepsilon_{i,t}$ captures the unexplained variation in prices, and it is clustered at the neighborhood level. The coefficient β provides an estimate of the inverse of the elasticity of supply, so $\eta = \frac{1}{\beta}$.

Regarding the first stage, $\ln Q_{i,t}$ is regressed on the shift-share instrument $\ln S_{i,t}$. The coefficient θ measures the strength of the relationship between the instrument and the endogenous variable. The equation also includes control variables $\mathbf{X}_{i,t}$, neighborhood fixed effects, ϕ_i , and time fixed effects, ψ_t . $v_{i,t}$ indicates the error, which is also clustered at the neighborhood level.

We start by showing the OLS results to get a sense of the effects when both demand and supply-side variation are present. This helps to understand the relevance of the instrument in our main estimation. Results are reported in Table C1. Columns (1) to (3) focus on the relationship between housing stock and prices (both in logarithms). The coefficients are positive and statistically significant across all specifications, ranging from 0.18 to 0.255. The inclusion of fixed effects and controls affects the magnitude of the coefficients and increases the R^2 . However, these estimates imply housing supply elasticities that are completely out of range (from 3.9 to 5.6). The explanation is that, as the supply effects are present, the coefficients are biased towards more negative values, which increase the estimated supply elasticity.

Columns (4) to (6) shift the focus from housing stock to transaction volumes. This variable helps to understand how different segments of the housing supply respond to rising prices, as we can distinguish between overall transactions, transactions for new constructions, and rental properties being sold (“rent-to-own” transactions).¹⁵ The high implied elasticities (ranging from 9.7 to 18.8) suggest a significant responsiveness of transaction volumes to price changes. These results will be contrasted with IV estimates to further refine the analysis. As expected, supply-side shifts are likely to be more pronounced for transactions than for stock, due to the greater variation in transaction activity across periods and census tracts. Figure C2 reports OLS estimates for another measure of prices, such as price per square foot and the property’s appraisal value. Results are consistent in terms of magnitude and significance.

Now we turn our attention to the instrumental variable approach. Results are reported in Table 3.2. Columns (1) through (3) focus on the relationship between the housing stock and various measures of prices: overall prices (column 1), price per square foot (column 2), and appraisal value (column 3). The IV estimates yield much larger coefficients for the effect of housing stock on prices compared to the OLS results, suggesting the presence of substantial bias in the OLS estimates. Specifically, in column (1), the coefficient of 1.843 indicates that a 1% increase in housing stock (driven by demand) leads to approximately a 1.84% increase in prices. The coefficients in columns (2) and (3) also demonstrate significant positive relationships, with magnitudes of 1.18 and 3.05, respectively. These results imply a

¹⁵We currently do not have specific data on traded properties that were previously rented, but we expect to obtain it soon. For now, to identify the likelihood of a property having been rented, we use tenure choice data at the municipality level from the CASEN survey (MDS, 2017). We then run a tenure choice model based on housing characteristics (e.g., size, apartment or house, price) and fixed effects. Using these estimates, we predict tenure status at the property level. Within each municipality, we classify properties as ‘rented’ if their predicted likelihood exceeds the municipality’s average.

very inelastic housing supply (ranging from 0.328 to 0.850), meaning that the market does not respond as strongly to changes in prices. Therefore, this supports the notion that supply constraints are binding in the housing market.

Regarding the relevance of the first stage, the *K-P F-statistic* consistently shows values above 20 across all estimations, indicating a reasonably strong instrument.

Columns (4) to (6) turn to analyse transactions: all, new constructions, and rent-to-own transactions. The results indicate a significant positive relationship between the number of transactions and prices, with the coefficient on overall transactions in column (4) being 0.574, indicating that a 1% increase in the number of transactions is associated with a 0.57% increase in prices. This implies an overall transaction-elasticity of 1.742. For transactions of new properties (column 5), the coefficient is 0.78, not far from the overall transactions estimation (implying an elasticity of nearly 1.3). Finally, the coefficient on rent-to-own transactions in column (6) is the largest at 1.946, which implies a rent-to-own transaction-elasticity of 0.515.

The low implied transaction elasticity for rent-to-own properties suggests that the market primarily responds to increasing demand for owner-occupied housing through new construction rather than by offering existing rental properties for purchase. The lack of response in rent-to-own transactions may mean that developers are converting few rental units into owner-occupied housing. This highlights a key supply-side limitation, especially in areas where there is little construction activity of low-price properties. To put this in context, Table 3.3 suggests that new constructions in Santiago tend to be concentrated on more central areas, leading to a geographic mismatch in housing availability. As a result, intermediate regions and suburban neighborhoods, where new construction is relatively scarce, may experience high pressure on housing prices and reduced opportunities for renters to transition into homeownership.

Table 3.3: Share of new constructions and CBD distance – Low price properties

Distance group (quintiles)	1	2	3	4	5
Share of new constructions (%)	58.0%	43.9%	21.6%	27.1%	31.1%
New constructions (normalized)	1	0.76	0.37	0.46	0.53
CBD distance (km)	2.03	5.51	9.31	12.81	17.13

Note: This table is constructed for low-price properties, defined as those valued at less than \$56,000 in 2010. It considers all new transactions from 2005 to 2019. The table shows the share of new constructions, the ratio of the new constructions in each group with respect to the first group (normalized to 1), and their corresponding distance to downtown across five distance groups.

This result challenges findings from the literature in developed countries, which show that housing supply tends to respond more strongly in neighborhoods that are further from the CBD (Saiz (2010); Carozzi et al. (2024); Baum-Snow and Han (2024)). These papers suggest

Table 3.2: IV estimates for housing supply

	(1) ln(Price)	(2) ln(Price sqft)	(3) ln(AV)	(4) ln(Price)	(5) ln(Price)	(6) ln(Price)
ln(Stock)	1.843* (0.962)	1.176* (0.634)	3.046* (1.637)			
ln(Transactions)				0.574*** (0.142)		
ln(NP Transactions)					0.780*** (0.200)	
ln(RtO Transactions)						1.946** (0.855)
Implied Elasticity	0.543	0.850	0.328	1.742	1.282	0.514
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,831	15,831	15,834	15,831	15,831	15,831
K-P F-Stat	21.78	21.78	20.61	352.4	46.87	20.85

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Each cell corresponds to the estimate from a 2SLS regression, where the dependent variables are the logarithms of price (in columns 1, 4, 5, and 6), price per square meter (column 2), and appraisal value (column 3). The main explanatory variables include the logarithms of stock, transaction volume, transactions of new properties, and rent-to-own transactions. Controls at the neighborhood level include the share of different property types, the average year of construction, and the total population. The regressions incorporate both year and neighborhood fixed effects to account for time and location-specific shocks. The table reports Kleibergen-Paap F-statistics (K-P F-stat) to assess the strength of the instruments used in the 2SLS estimation. Standard errors are clustered at the neighborhood level and are reported in parentheses.

that suburban development plays a key role in accommodating increased housing demand, as they have more available land for new construction. In contrast, our results suggest that suburban areas may also exhibit low elasticity of supply, particularly when developers concentrate construction efforts in more affluent areas and when there are rigidities in the conversion of rental units to homeownership. This highlights the importance of studying housing supply elasticities in middle- and low-income countries to better understand how the affordability crisis can be addressed in various contexts. In the next section, we present several mechanisms that could explain the misallocation of new constructions and the supply of owner-occupied housing.

Potential mechanisms

There are at least four key factors that could explain why housing supply may respond slowly even in areas without significant constraints. First, heterogeneity in zoning policies across neighborhoods can influence the pace of new developments (Glaeser et al. (2005); Ihlanfeldt (2007)). Second, higher profit margins for developers may concentrate new construction in more lucrative areas. Additionally, competition for land between real estate developers and farmers or agricultural users could limit the spatial size of the city (Brueckner (2000); Thisse and Fujita (2002)). Third, better infrastructure—such as transportation networks, utilities, and public services—can facilitate new housing developments with minimal additional investment (Duranton and Puga (2004); Asahi et al. (2023)). Finally, the rate of redevelopment of underutilized land can also contribute to faster supply responses (Glaeser and Gyourko (2005); Baum-Snow and Han (2024)). We will explore each of these mechanisms in the following paragraphs.

Zoning. Santiago de Chile has a zoning-by-neighborhood system¹⁶. Santiago’s downtown area, for example, has had a zoning policy in place since 1990, which has been modified 29 times. Even with these restrictions in place, the population in the CBD has doubled between the 2002 and 2017 censuses (Valenzuela, 2023), which is consistent with a relatively fast response of the housing supply. In fact, recent evidence suggests that developers have managed to circumvent these constraints through vertical expansion of buildings (Sheehan, 2024; Martinez and Lopez, 2018), a phenomenon increasingly relevant in Asia, Latin America, Oceania, and Europe (World Bank (2023)).

Profitability of new constructions. While we do not have direct information on developer profits from new constructions, we do have data on new builds across the city, which can serve as an indicator of profitability. Developers tend to build more when they expect higher returns. Given that zoning policies are often more stringent in downtown areas and suburban regions typically have greater land availability, this indicator can be seen as a lower bound for differences in profitability. Table 3.3 shows that suburban neighborhoods experience 47% less new construction compared to CZs closer to the CBD. Intermediate areas, located around 9.31 km from the CBD, witness an even lower rate of new construction (-67%). To support these estimates, using the subsidy data, we find that only 4.9% of Title 0 voucher recipients purchase new properties, whereas this figure rises to 29.2% for Title 1 recipients and 66.1% for Title 2 recipients. These pieces of evidence imply that developers are not creating many relevant options for low-income individuals in low-income neighborhoods.

Redevelopment. To test this hypothesis, we follow Baum-Snow and Han (2024) in building an indicator of the number of new properties developed on previously developed land. To do that, we use historical transaction data at the block level (smaller than the CZ level). Table 3.4 presents the results by CBD distance. The findings show that nearly 57%

¹⁶Each municipality establishes its own regulations, which include restrictions on the extensive margin (such as prohibiting certain types of infrastructure), the intensive margin (such as maximum building heights), and the type of land use (residential or commercial).

of new construction in the city's downtown occurs on blocks that already had properties, while the share of redevelopment decreases to 44% for blocks located 5.51 kilometers from downtown, and to around 30% for blocks more than 10 kilometers from the CBD (groups 3, 4 and 5). This evidence suggests a relatively high rate of redevelopment in central areas compared to the rest of the city, and is consistent with the vertical growth in the CBD. That being said, a more accurate method would involve using satellite imagery to capture developed land at the pixel level, which would provide a finer measure of redevelopment. We leave this exercise for future iterations of this research, so our current indicator likely represents an upper bound on the true extent of redevelopment.

Table 3.4: Share of redevelopment and CBD distance

Distance group (quintiles)	1	2	3	4	5
Share of redevelopment (%)	56.91%	43.79%	31.00%	26.76%	31.57%
CBD distance (km)	2.03	5.51	9.31	12.81	17.13

Note: This table reports the share of redevelopment (new properties built on previously developed blocks) across different distance quintiles from the CBD. The share of redevelopment decreases as the distance from the CBD increases, indicating higher redevelopment activity in areas closer to the city center.

Current infrastructure. Lastly, recent evidence from Chile ([Asahi et al., 2023](#)) underscores the significant role of highways and subway lines in increasing housing supply by stimulating demand. This implies that census zones (CZs) near the CBD can more easily respond to demand shocks, while areas farther from the city center face greater challenges in doing so.

Overall, these four mechanisms suggest that supply elasticity may be low not only in the CBD, but also in more distant areas. Therefore, housing assistance programs should consider how to encourage supply to respond to demand shocks beyond policy-restricted zones. Additionally, studying supply responses in different countries is crucial for understanding the broader impacts of housing assistance programs in various contexts.

Back-of-the-Envelope Calculation of Price Increases Induced by the Chilean Homeownership Voucher Program

To end this section, we use a simple supply-demand framework to estimate the effects on prices. The equation that enables us to estimate these effects is the following:

$$\Delta\%P = \frac{\Delta\%D}{\eta}$$

where $\Delta\%P$ is the percentage change in price, $\Delta\%D$ is the percentage change in demand, and η is the housing supply elasticity. As supply becomes less elastic (lower η), prices must

rise more to accommodate the additional demand. When supply is perfectly elastic, η is infinite, and prices remain stable despite the demand increase. To estimate the equation, we need to make an assumption about the increase in demand, which is not included in the previously mentioned results. We assume a 5.5% increase in demand for owner-occupied housing, which represents the percentage of Santiago's population that received the homeownership voucher from 2011 to 2019. As a second metric, we also use 3.4%, which represents the effective use of the subsidy¹⁷.

Based on the IV estimates of the housing supply elasticity, a 5.5% increase in demand for owner-occupied housing would imply an increase of 10.13% in average prices, 6.47% in prices per square foot, and 16.77% in appraisal values. On the other hand, a 3.4% increase in demand would imply a 6.29% increase in average prices, a 3.95% increase in prices per square foot, and a 10.41% increase in appraisal value.

The implications of these calculations are significant for understanding the impact of housing assistance in different contexts. In areas with tight housing supply, such programs may exacerbate affordability issues by driving up prices, reducing the intended benefit of the program. Conversely, in markets with more elastic supply, the program's price impact would be smaller, allowing for a more effective expansion of homeownership without significant inflationary pressures. This underscores the importance of local housing market conditions when designing and implementing demand-side interventions like homeownership vouchers.

Table 3.5: Implied effect on real prices (2011-2019)

Δ Demand	Δ Price	Δ Price per sq ft	Δ Appraisal value
5.5%	10.13%	6.37%	16.77%
3.4%	6.29%	3.95%	10.41%

Note: The table shows the estimated effect of a demand increase for owner-occupied housing on real prices, prices per square foot, and appraisal values, based on IV estimates of housing supply elasticity. A 5.5% increase in demand corresponds to the percentage of Santiago's population that received homeownership vouchers during this period, while 3.4% represents the population that actually uses them in the period.

In Section 3.5, we directly estimate the effects of the policy on housing prices in an Event Study framework. Results are of similar magnitude.

¹⁷We consider 62,542 recipients that have used the subsidy, which represents 250,168 individuals if we consider their families. This number divided by nearly 7.3 million people in Santiago is 3.4%.

3.5 Validation of the methodology and effects on housing prices: an Event Study approach

The previous section estimates Santiago's housing supply elasticity. The identification strategy relies on the assumptions of relevance and exogeneity. While the exogeneity assumption cannot be directly verified as relevant, some robustness tests can be conducted to support its validity. In this section, we slightly modify the empirical strategy to explore the potential presence of unobserved factors that may differentially affect the different treatments and control groups over time. Specifically, in this section, we implement an Event Study with a continuous treatment based on neighborhood exposure to different voucher types. This strategy enables us to test for parallel trends, check for the existence of anticipation effects prior to the policy, and estimate the impact of the policy (i.e., the implementation of the voucher system) on local housing prices. Thus, this section serves three main purposes: first, to validate the exogeneity assumption of the methodology proposed in Section 3.4; second, to provide a robust causal estimate of the policy's effect on housing prices; and third, to provide an analysis on the heterogeneous effect of the policy, which was not possible to implement in the instrumental variable approach.

Empirical strategy

The main strategy we follow in this section is the Event Study represented in Equation 3.6.

$$y_{i,t} = \alpha + \sum_{k=1}^3 \sum_{t \neq 2010} \beta_t^k D_t^k \times \text{Exp}_i^k + \sum_{k=1}^3 \sum_{t \neq 2010} \gamma_t^k D_t^k + \delta_k \text{Exp}_i^k + \mathbf{X}_{i,t}' \gamma + \lambda_i + \theta_t + \epsilon_{i,t} \quad (3.6)$$

where $y_{i,t}$ represents the outcome variable of interest, such as housing prices, for neighborhood i at year t . The coefficients β_k capture the interaction between the indicator of exposure (Exp_i^k) and the time binary variable for each time (D_t^k). We also include the binary variables and the exposure measure as a control. The vector $\mathbf{X}_{i,t}'$ includes control variables that vary by neighborhood and time, such as population, share of property type, and average year of construction of properties. The model also includes neighborhood fixed effects, λ_i , and year fixed effects, θ_t . The error term $\epsilon_{i,t}$ captures unobserved factors that affect the outcome variable. Standard errors are clustered at the neighborhood level.

The treatment indicator, Exp_i^k , is defined in the same way as the "shares" in Section 3.4 (see Equation 3.2). Thus, it can be described by Figures C5 and C6.

The identification strategy relies on three main assumptions. First, there must be parallel trends in housing prices, meaning that, prior to the implementation of the policy, neighborhoods with different levels of exposure to each voucher type followed similar price trends. Second, the policy change must be exogenous, implying that the reform was not implemented in response to unobserved characteristics of the housing market that differ across treatment

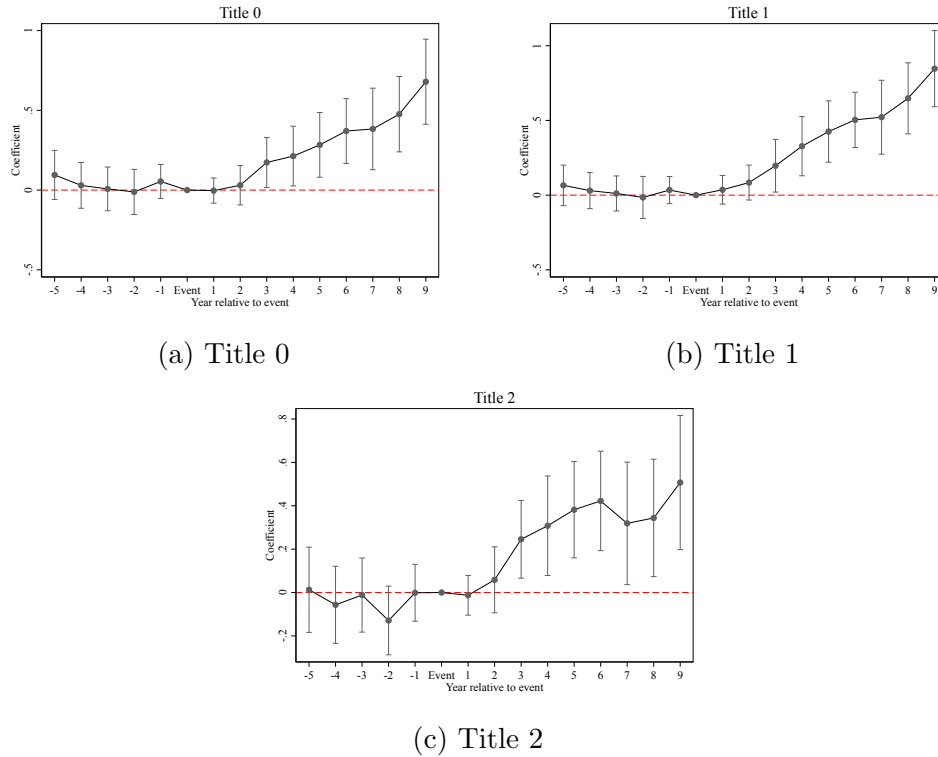
groups. Third, it is assumed that households and the housing market did not anticipate the reform before its implementation in 2011.

In general, the assumptions can be tested using the Event Study by examining the coefficients for periods prior to the reform. Additionally, we control for time fixed effects that capture annual shocks common to all neighborhoods. Therefore, part of the potential endogeneity is addressed through these variables. Furthermore, our measure of exposure to each voucher provides quasi-random variation around the property price caps that define eligibility for each voucher. For instance, a property priced at US\$37,500 is eligible for voucher Title 0, while a property priced at US\$38,000 is not. Thus, when we measure treatment exposure, some neighborhoods have more exposure than others for reasons based solely on the policy structure and not on housing market conditions. Lastly, we restrict the control group to neighborhoods with average property prices up to US\$150,000, meaning that very high-income neighborhoods are not included as part of the control group. Different thresholds for the support of the control group do not significantly change the results.

Results

The result of the estimation of Equation 3.6 is reported in Figure C9. The first important finding is the absence of pre-trends for the exposure to vouchers of type Title 0 and Title 1. However, there is evidence of linear pre-trends for Title 2. While the existence of pre-event trends for voucher Title 2 can be seen as evidence against the identifying assumptions, Freyaldenhoven et al. (2021) suggests that a pre-adoption trend in the outcome can be used to learn the slope of the trend in the error. Extrapolating this slope into the post-adoption periods then permits accounting for the confounding factors. We applied this method to make the necessary corrections, which is also reported in Figure C9 to show the change between the “raw” estimates and the adjusted ones. Figure 3.4 reports the results of the Event Study with pre-trend adjustment.

Figure 3.4: Event Study on Local Prices



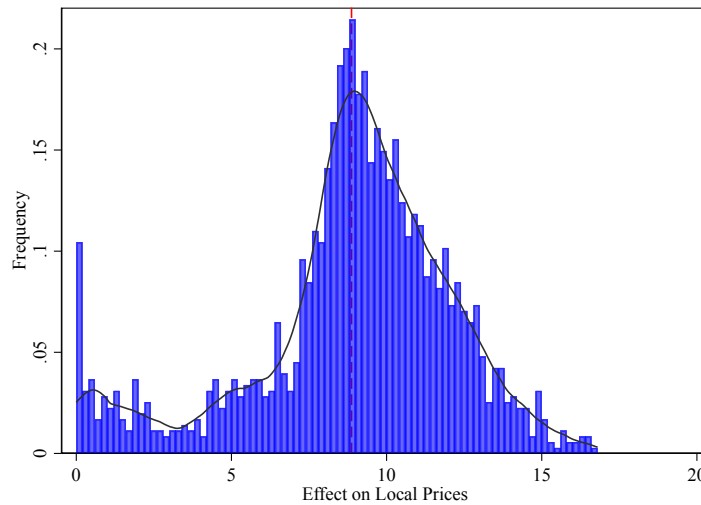
Note: This figure presents the results of an event study analyzing local property prices. The analysis is performed for three voucher types: Title 0, Title 1, and Title 2. Panel (a) shows the impact on prices for Title 0, Panel (b) for Title 1, and Panel (c) for Title 2. We make adjustments for pre-trends, following Freyaldenhoven et al. (2019).

Figure 3.4 indicates a strong and consistent effect of the exposure to the vouchers on the increase in housing prices. The Y-axis shows the percentage change in prices, while the bars represent 95 percent confidence intervals. Panel (a) indicates that in the first two years of the policy, there was no sizable impact of the eligibility for voucher Title 0 on prices. This is logical because Title 0 was implemented at the end of 2012, and individuals usually take an average of two years to purchase a property. From the third year, there is a positive and significant effect of the exposure to vouchers. Overall, we calculate an average effect of 0.36% per year for voucher Title 0. Panel (b) shows the effects of exposure to vouchers of the Title 1 type. The figure shows a larger effect in year 2 compared to Title 0, but still insignificant. After year 2, we find an increasing effect on prices (an average of 0.39%), which is consistent with the fact that more individuals use the voucher to purchase properties over time, as they decide where to buy. Lastly, for Title 2, we observe a smaller but still significant effect from year 3 after the event.

To put the results in context, the estimated effect on prices of each voucher tier enables

us to estimate the overall effect of vouchers on housing prices in Santiago. To do so, we need to calculate the cumulative effect (between 2011 and 2019) for each voucher tier and multiply it by the exposure share at each neighborhood ($\sum_{k=1}^3 \text{C.E.}_k \times \text{Exp}_i^k$). The resulting cumulative effects (C.E.) for each voucher type were 7.72% (Title 0), 16.64% (Title 1), and 7.55% (Title 2). Then, after summing the effects of all vouchers in each CZ, we end up with an average price effect of 8.87%. However, we find an important heterogeneous effect across neighborhoods. Figure 3.5 reports the distribution of price effects across Census Zones.

Figure 3.5: Distribution of Price Effects (2011-2019)



Note: This figure illustrates the distribution of the estimated price effects (in percentage) at the local level based on the share of eligible properties for each voucher titer. The effects are calculated using cumulative impacts for each voucher type across neighborhoods between 2011 and 2019 ($\sum_{k=1}^3 \text{C.E.}_k \times \text{Exp}_i^k$). The average price effect (8.87%) is highlighted in a red dashed line.

Figure 3.5 suggests a bimodal distribution of price effects, with one concentration around very low increases and another around a 9% increase. Additionally, it is noteworthy that some CZs, particularly in areas with greater exposure to Title 1, exhibit increases of up to 15%. Conversely, CZs with low exposure to all three voucher types experienced minimal local effects, with price changes close to zero.

These findings help in validating the estimated supply elasticity presented in the previous section. As shown in Table 3.5, the implied elasticities were consistent with price increases of between 6.29% and 10.13% (column 1). The average estimate in this section (8.87%) falls between those estimates. If we assume an increase in the demand of 5.5%, an increase in prices of 8.87% would be consistent with a housing supply elasticity of 0.62, which is just slightly larger than the main estimate of Table 3.2 (0.543).

Therefore, the event study supports the empirical strategy followed in the previous section and provides robustness in the calculation of the housing supply elasticity.

Heterogeneous effects

Taking advantage of this empirical framework, we provide some evidence of the heterogeneous response of housing prices to housing market characteristics. We show evidence of three different variables: CBD distance, population density, and share of new construction per CZ. Results are reported in Figure C9.

The top row of the figure breaks down the effects by distance from the CBD. It shows that property prices increased more in areas located farther than 10 km from the CBD compared to those closer to it. Panel (a), which presents the results for Title 0, indicates a positive and significant effect on prices in distant areas, while the effect in areas near the CBD is not significant. Panel (b) shows similar results for Title 1, though the differential effect across distance groups is smaller. Panel (c) reports the effects of Title 2, suggesting positive price effects but also some evidence of pre-trends. Overall, we cannot conclude that housing supply elasticity is larger downtown, as there is a correlation between CBD distance and exposure to the treatment. However, we can conclude that suburban areas do not exhibit high housing supply elasticity.

The middle row of Figure C9 presents the heterogeneous analysis by population density. We observe larger price effects in CZs with higher population density. This pattern holds across all voucher types, but is particularly pronounced for Title 0 and Title 1. Additionally, we do not observe any pre-trends in this analysis for any of the voucher types.

In the bottom row, the focus shifts to the share of new properties in each neighborhood. The price effects are most pronounced in areas with a higher share of new construction (Panels g, h, and i) with respect to the average. This is consistent with the greater responsiveness of supply in places that traditionally allow for new constructions or where developers are investing in new buildings.

These heterogeneous effects align with the previously estimated elasticities, particularly the differences between rent-to-own and new construction elasticities that suggest a low elasticity of supply in neighborhoods that are distant from the CBD. This evidence is valuable for designing housing assistance programs for low- and middle-income families. Both the overall supply response and its geographic variation across housing markets provide critical insights into the effects of subsidies for owner-occupied housing and other policies directed at vulnerable households.

3.6 Conclusions

This paper estimates housing supply elasticity, a crucial factor in understanding how housing markets respond to assistance policies. Using quasi-experimental variation from Chile's *Subsidio Habitacional* program and detailed administrative data on property transactions, we find that Santiago's short-term housing supply elasticity is 0.543. This implies that

subsidizing owner-occupied housing will likely be capitalized into higher prices, rather than directly improving housing affordability for low- and middle-income households.

To examine the policy's differential impacts across housing markets, we estimate the elasticities of new-property (NP) and rent-to-own (RtO) transactions. The results indicate that the responsiveness of new properties to price increases is more than double that of rental properties transitioning to owner-occupied housing. This is particularly relevant for low- and middle-income neighborhoods in suburban areas, where the share of new constructions is relatively low compared to downtown areas. This rigidity suggests that developers may be converting too few rental units into owner-occupied housing, limiting opportunities for families seeking to transition to homeownership. Additionally, we argue that the low elasticity observed in suburban areas may be a combination of lack of infrastructure, low profitability for developers, and a relatively high rate of redevelopment in central areas.

Using these elasticity estimates, we assess the impact of the homeownership voucher on housing prices. The findings indicate that real prices rose by between 6.29% and 10.13% due to the policy. To validate our empirical strategy and strengthen the evidence on price effects, we conduct an event study that exploits the timing of the introduction of the program. We find a consistent positive effect across all vouchers, with an overall increase of 8.87%. Notably, there is significant heterogeneity across neighborhoods, which depends on the overall exposure to the three vouchers. We notice that higher price effects are observed in Census Zones with higher population density, fewer new properties, and greater distance from the city's downtown.

Should governments subsidize owner-occupied housing? This study suggests that policymakers must account for the responsiveness of housing supply when designing assistance programs, especially in neighborhoods that are more exposed to the program. Future research could explore the benefits of implementing complementary measures that stimulate supply, such as promoting new construction in low- and middle-income neighborhoods.

Bibliography

- ABDULKADIROĞLU, A., P. A. PATHAK, J. SCHELLENBERG, AND C. R. WALTERS (2020): “Do parents value school effectiveness?” *American Economic Review*, 110, 1502–1539.
- ABED (2018): “Censo EAD.BR: relatório analítico da aprendizagem a distância no Brasil 2017,” *National Center for Education Statistics*.
- AGNESS, D. AND T. GETAHUN (2024): “Housing and Human Capital: Condominiums in Ethiopia,” .
- AINSWORTH, R., R. DEHEJIA, C. POP-ELECHES, AND M. URQUIOLA (2023): “Why Do Households Leave School Value Added on the Table? The Roles of Information and Preferences,” *American Economic Review*, 113, 1049–82.
- ALLENDE, C. (2019): “Competition under social interactions and the design of education policies,” *Working Paper*.
- ANDREWS, I., T. KITAGAWA, AND A. MCCLOSKEY (2024): “Inference on winners,” *The Quarterly Journal of Economics*, 139, 305–358.
- ANGRIST, J., P. HULL, AND C. WALTERS (2023): “Methods for measuring school effectiveness,” *Handbook of the Economics of Education*, 7, 1–60.
- ANGRIST, J. D., P. D. HULL, P. A. PATHAK, AND C. R. WALTERS (2017): “Leveraging Lotteries for School Value-Added: Testing and Estimation,” *The Quarterly Journal of Economics*, 132, 871–919.
- ARADILLAS-LÓPEZ, A. (2020): “The econometrics of static games,” *Annual Review of Economics*, 12, 135–165.
- ARMONA, L. AND S. CAO (2024): “Redesigning Federal Student Aid in Sub-Baccalaureate Education,” *Available at SSRN 4300755*.
- ASAHI, K., A. HERRERA, AND H. SILVA (2023): “The Effects of Transport Infrastructure on Housing Supply: The Role of Land-Use Regulation,” *Working Paper*.
- ATAL, J. P., J. I. CUESTA, AND M. SÆTHRE (2025): “Quality regulation and competition: Evidence from pharmaceutical markets,” Tech. rep., National Bureau of Economic Research.

- AUCEJO, E. M., A. S. PERRY, AND B. ZAFAR (2024): “Assessing the Costs of Balancing College and Work Activities: The Gig Economy Meets Online Education,” Tech. rep., National Bureau of Economic Research.
- BAJARI, P., H. HONG, J. KRAINER, AND D. NEKIPELOV (2010): “Estimating static models of strategic interactions,” *Journal of Business & Economic Statistics*, 28, 469–482.
- BARAHONA, N., C. DOBBIN, H. HO, S. OTERO, AND C. YANNELIS (2023a): “Skin in the game: colleges’ financial incentives and student outcomes,” *Working Paper*.
- BARAHONA, N., C. DOBBIN, AND S. OTERO (2023b): “Affirmative Action in Centralized College Admissions Systems,” Tech. rep., Columbia University.
- BARNHARDT, S., E. FIELD, AND R. PANDE (2017): “Moving to opportunity or isolation? network effects of a randomized housing lottery in urban india,” *American Economic Journal: Applied Economics*, 9, 1–32.
- BARROW, L., W. T. MORRIS, AND L. SARTAIN (2024): “The Expanding Landscape of Online Education: Who Engages and How They Fare,” *Journal of Labor Economics*, 42, S417–S443.
- BARTIK, T. (1991a): *Who Benefits from State and Local Economic Development Policies?*, W.E. Upjohn Institute.
- BARTIK, T. J. (1991b): *Who Benefits from State and Local Economic Development Policies?*, W.E. Upjohn Institute.
- BAU, N. (2022): “Estimating an equilibrium model of horizontal competition in education,” *Journal of Political Economy*, 130, 1717–1764.
- BAUM-SNOW, N. AND L. HAN (2024): “The Microgeography of Housing Supply,” *Journal of Political Economy*, 132, 1897–1946, publisher: The University of Chicago Press.
- BECKER, G. S., W. H. HUBBARD, AND K. M. MURPHY (2010): “Explaining the Worldwide Boom in Higher Education of Women,” *Journal of Human Capital*, 4, 203–241.
- BEHRMAN, J. R., S. W. PARKER, AND P. E. TODD (2005): “Long-Term Impacts of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico,” Ibero America Institute for Econ. Research (IAI) Discussion Papers 122, Ibero-America Institute for Economic Research.
- BELCHIOR, C., G. GONZAGA, AND G. ULYSSEA (2023): “Unpacking Neighborhood Effects: Experimental Evidence from a Large-Scale Housing Program in Brazil,” *Institute of Labor Economics (IZA)*, IZA Discussion Papers 16113.
- BERRY, S. AND P. HAILE (2016): “Identification in differentiated products markets,” *Annual review of Economics*, 8, 27–52.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63, 841–890.

- BERTOLIN, J., T. MCCOWAN, AND H. BITTENCOURT (2023): “Expansion of the Distance Modality in Brazilian Higher Education: Implications for Quality and Equity,” *High Educ Policy*, 36, 231–249.
- BETTINGER, E. P., L. FOX, S. LOEB, AND E. S. TAYLOR (2017): “Virtual Classrooms: How Online College Courses Affect Student Success,” *American Economic Review*, 107, 2855–2875.
- BIS (2024): “Residential Property Prices (RPP),” *Bank for International Settlements*, accessed: 2024-08-21.
- BLOURI, Y., S. BÜCHLER, AND O. SCHÖNI (2023): “The geography of mortgage interest deductions,” *Journal of Urban Economics*, 138, 103604.
- BODÉRE, P. (2023): “Dynamic spatial competition in early education: An equilibrium analysis of the preschool market in Pennsylvania,” *Working Paper*.
- BONEVA, T. AND C. RAUH (2018): “Parental Beliefs about Returns to Educational Investments—The Later the Better?” *Journal of the European Economic Association*, 16, 1669–1711.
- BORUSYAK, K. AND P. HULL (2023): “Nonrandom Exposure to Exogenous Shocks,” *Econometrica*, 91, 2155–2185.
- BORUSYAK, K., P. HULL, AND X. JARAVEL (2021): “Quasi-Experimental Shift-Share Research Designs,” *The Review of Economic Studies*, 89, 181–213.
- BRUECKNER, J. (2000): “Urban Sprawl: Diagnosis and Remedies,” *International Regional Science Review*, 23, 160–171.
- CALABRESE, S., D. EPPLE, AND R. ROMANO (2011): “Inefficiencies from Metropolitan Political and Fiscal Decentralization: Failures of Tiebout Competition,” *The Review of Economic Studies*, 79, 1081–1111.
- CALLAWAY, B., A. GOODMAN-BACON, AND P. H. SANT’ANNA (2024): “Difference-in-differences with a continuous treatment,” Tech. rep., National Bureau of Economic Research.
- CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014): “Robust Data-Driven Inference in the Regression-Discontinuity Design,” *Stata Journal*, 14, 909–946.
- CAMACHO, A., V. DUQUE, M. GILRAINE, AND F. SANCHEZ (2022): “The Effects of Public Housing on Children: Evidence from Colombia,” Working Paper 30090, National Bureau of Economic Research, nBER Working Paper No. 30090.
- CAMPAIGN, H.-W. (2024): “Economic Opportunity Policy Book,” Accessed: 2024-11-11.
- CARD, D. (1999): “Chapter 30 - The Causal Effect of Education on Earnings,” Elsevier, vol. 3 of *Handbook of Labor Economics*, 1801–1863.

- CARD, D., E. CHYN, AND L. GIULIANO (2024): “Can Gifted Education Help Higher-Ability Boys from Disadvantaged Backgrounds?” .
- CARLANA, M., E. LA FERRARA, AND P. PINOTTI (2022): “Goals and Gaps: Educational Careers of Immigrant Children,” *Econometrica*, 90, 1–29.
- CAROZZI, F., C. A. HILBER, AND X. YU (2024): “On the economic impacts of mortgage credit expansion policies: Evidence from help to buy,” *Journal of Urban Economics*, 139, 103611.
- CATTANEO, M., M. JANSSON, AND X. MA (2020a): “Simple Local Polynomial Density Estimators,” *Journal of the American Statistical Association*, 115, 1449–1455.
- CATTANEO, M., R. TITIUNIK, AND G. VAZQUEZ-BARE (2020b): “The Regression Discontinuity Design,” *The SAGE Handbook of Research Methods in Political Science and International Relations*, 835–857.
- CATTANEO, M. D., S. GALIANI, P. J. GERTLER, S. MARTINEZ, AND R. TITIUNIK (2009): “Housing, Health, and Happiness,” *American Economic Journal: Economic Policy*, 1, 75–105.
- CHETTY, R. AND N. HENDREN (2018): “The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates*,” *The Quarterly Journal of Economics*, 133, 1163–1228.
- CHETTY, R., N. HENDREN, AND L. F. KATZ (2016a): “The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment,” *American Economic Review*, 106, 855–902.
- (2016b): “The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment,” *American Economic Review*, 106, 855–902.
- CHETTY, R., N. HENDREN, P. KLINE, AND E. SAEZ (2014): “Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *,” *The Quarterly Journal of Economics*, 129, 1553–1623.
- CHYN, E. (2018): “Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children,” *American Economic Review*, 108, 3028–56.
- CHYN, E. AND L. F. KATZ (2021): “Neighborhoods Matter: Assessing the Evidence for Place Effects,” *Journal of Economic Perspectives*, 35, 197–222.
- CLAMPET-LUNDQUIST, S., K. EDIN, J. KLING, AND G. DUNCAN (2011): “Moving teenagers out of high-risk neighborhoods: How girls fare better than boys,” *American Journal of Sociology*, 116, 1154–1189.
- COLLINSON, R., J. E. HUMPHRIES, N. MADER, D. REED, D. TANNENBAUM, AND W. VAN DIJK (2023): “Eviction and Poverty in American Cities*,” *The Quarterly Journal of Economics*, 139, 57–120.

- CONLON, C. AND J. GORTMAKER (2020): “Best practices for differentiated products demand estimation with pyblp,” *The RAND Journal of Economics*, 51, 1108–1161.
- COSMAN, J., T. DAVIDOFF, AND J. WILLIAMS (2018): “Housing Appreciation and Marginal Land Supply in Monocentric Cities with Topography,” *Working Paper*, journal: Working Paper.
- COUTURE, V., C. GAUBERT, J. HANDBURY, AND E. HURST (2023): “Income Growth and the Distributional Effects of Urban Spatial Sorting,” *The Review of Economic Studies*, 91, 858–898.
- COX, W. AND H. PAVLETICH (2024): “Demographia International Housing Affordability: 2024 Edition,” *Center for Demographics and Policy, Chapman University*, accessed: 2024-08-21.
- CUTLER, D. M. AND E. L. GLAESER (1997): “Are Ghettos Good or Bad?” *The Quarterly Journal of Economics*, 112, 827–872.
- DAHL, G. B. AND L. LOCHNER (2012): “The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit,” *American Economic Review*, 102, 1927–56.
- DAHLSTRAND, A. (2024): “Defying Distance? The Provision of Services in the Digital Age,” Tech. rep., University of Zurich.
- DAHLSTRAND, A., N. LE NESTOUR, AND G. MICHAELS (2024): “Online versus In-Person Services: Effects on Patients and Providers,” Tech. rep., Institute of Labor Economics (IZA), Bonn.
- DAVIDOFF, T. (2016): “Supply Constraints Are Not Valid Instrumental Variables for Home Prices Because They Are Correlated with Many Demand Factors,” *Critical Finance Review*, 177–2065.
- DAVIS, M. (2019): “The Distributional Impact of Mortgage Interest Subsidies: Evidence from Variation in State Tax Policies,” *Working Paper*.
- DEMING, D. (2009): “Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start,” *American Economic Journal: Applied Economics*, 1, 111–34.
- DEMING, D. J., C. GOLDIN, AND L. F. KATZ (2012): “The For-Profit Postsecondary School Sector: Nimble Critters or Agile Predators?” *Journal of Economic Perspectives*, 26, 139–64.
- DEMING, D. J., C. GOLDIN, L. F. KATZ, AND N. YUCHTMAN (2015): “Can Online Learning Bend the Higher Education Cost Curve?” *American Economic Review*, 105, 496–501.
- DEMING, D. J., M. LOVENHEIM, AND R. W. PATTERSON (2016a): “The Competitive Effects of Online Education,” Working Paper 22749, National Bureau of Economic Research.
- DEMING, D. J., N. YUCHTMAN, A. ABULAFI, C. GOLDIN, AND L. F. KATZ (2016b): “The Value of Postsecondary Credentials in the Labor Market: An Experimental Study,” *American Economic Review*, 106, 778–806.
- DEUTSCHER, N. (2020): “Place, Peers, and the Teenage Years: Long-Run Neighborhood Effects in Australia,” *American Economic Journal: Applied Economics*, 12, 220–49.

- DIAMOND, R., T. MCQUADE, AND F. QIAN (2019): “The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco,” *American Economic Review*, 109, 3365–94.
- DINERSTEIN, M., D. MORALES, C. NEILSON, AND S. OTERO (2023): “The equilibrium effects of public provision in education markets,” *Working paper*.
- DINERSTEIN, M. AND T. D. SMITH (2021): “Quantifying the supply response of private schools to public policies,” *American Economic Review*, 111, 3376–3417.
- DOBBIN, C., N. BARAHONA, AND S. OTERO (2021): “The Equilibrium Effects of Subsidized Student Loans,” Tech. rep., Working paper, Stanford University.
- DURANTON, G. AND D. PUGA (2004): “Chapter 48 - Micro-Foundations of Urban Agglomeration Economies,” in *Cities and Geography*, ed. by J. V. Henderson and J.-F. Thisse, Elsevier, vol. 4 of *Handbook of Regional and Urban Economics*, 2063–2117.
- EL GALAD, A., D. H. BETTS, AND N. CAMPBELL (2024): “Flexible learning dimensions in higher education: aligning students’ and educators’ perspectives for more inclusive practices,” in *Frontiers in Education*, Frontiers Media SA, vol. 9, 1347432.
- ELLEN, I. G. (2018): “Housing Choice Vouchers,” Accessed: 2024-11-11.
- ERIKSEN, M. D. AND A. ROSS (2015): “Housing Vouchers and the Price of Rental Housing,” *American Economic Journal: Economic Policy*, 7, 154–76.
- FABREGAS, R. AND L. NAVARRO-SOLA (2024): “Broadcasting Education at Scale: Long-Term Labor Market Impacts of Television-Based Schools,” *Working paper*.
- FIGLIO, D., M. RUSH, AND L. YIN (2013): “Is It Live or Is It Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning,” *Journal of Labor Economics*, 31, 763–784.
- FREYALDENHOVEN, S., C. HANSEN, J. P. PÉREZ, AND J. M. SHAPIRO (2021): “Visualization, Identification, and Estimation in the Linear Panel Event-Study Design,” *NBER Working Papers*.
- GALIANI, S., P. J. GERTLER, R. UNDURRAGA, R. COOPER, S. MARTÍNEZ, AND A. ROSS (2017): “Shelter from the storm: Upgrading housing infrastructure in Latin American slums,” *Journal of Urban Economics*, 98, 187–213, urbanization in Developing Countries: Past and Present.
- GANDHI, A. AND J.-F. HOUDE (2019): “Measuring substitution patterns in differentiated-products industries,” *NBER working paper*.
- GARRETT, R., B. SIMUNICH, R. LEGON, AND E. E. FREDERICKSEN (2022): “Tracking Online Learning from Mainstream Acceptance to Universal Adoption,” Tech. rep.
- GENNETIAN, L. A., L. SANBONMATSU, L. F. KATZ, J. R. KLING, M. SCIANDRA, J. LUDWIG, G. J. DUNCAN, AND R. C. KESSLER (2012): “The Long-Term Effects of Moving to Opportunity on Youth Outcomes,” *Cityscape*, 14, 137–167.

- GLAESER, E. AND J. GYOURKO (2005): “Urban Decline and Durable Housing,” *Journal of Political Economy*, 113, 345–375.
- GLAESER, E., J. GYOURKO, AND R. SAKS (2005): “Why Have Housing Prices Gone Up?” *American Economic Review*, 95, 329–333.
- GOLDSMITH-PINKHAM, P., I. SORKIN, AND H. SWIFT (2020): “Bartik instruments: What, when, why, and how,” *American Economic Review*, 110, 2586–2624.
- GONZÁLEZ-VELOSA, C., G. RUCCI, M. SARZOSA, AND S. URZUA (2015): “Returns to Higher Education in Chile and Colombia,” Working Paper No. IDB-WP-58, IDB, IDB WORKING PAPER SERIES No. IDB-WP-58.
- GOODMAN, J., J. MELKERS, AND A. PALLAIS (2019): “Can Online Delivery Increase Access to Education?” *Journal of Labor Economics*, 37, 1–34, publisher: The University of Chicago Press.
- GOUX, D. AND E. MAURIN (2005): “The effect of overcrowded housing on children’s performance at school,” *Journal of Public Economics*, 89, 797–819.
- GRAY-LOBE, G., P. A. PATHAK, AND C. R. WALTERS (2022): “The Long-Term Effects of Universal Preschool in Boston*,” *The Quarterly Journal of Economics*, 138, 363–411.
- GURRYAN, J., E. HURST, AND M. KEARNEY (2008): “Parental Education and Parental Time with Children,” *Journal of Economic Perspectives*, 22, 23–46.
- HALTIWANGER, J. C., M. J. KUTZBACH, G. E. PALLONI, H. POLLAKOWSKI, M. STAIGER, AND D. WEINBERG (2020): “The Children of HOPE VI Demolitions: National Evidence on Labor Market Outcomes,” Working Paper 28157, National Bureau of Economic Research.
- HAUSMAN, J., G. LEONARD, AND J. D. ZONA (1994): “Competitive analysis with differentiated products,” *Annales d’Economie et de Statistique*, 159–180.
- HECKMAN, J. J. AND T. KAUTZ (2012): “Hard evidence on soft skills,” *Labour Economics*, 19, 451–464, european Association of Labour Economists 23rd annual conference, Paphos, Cyprus, 22-24th September 2011.
- HECKMAN, J. J., S. H. MOON, R. PINTO, P. A. SAVELYEV, AND A. YAVITZ (2010): “The rate of return to the HighScope Perry Preschool Program,” *Journal of Public Economics*, 94, 114–128.
- HECKMAN, J. J. AND L. K. RAUT (2016): “Intergenerational long-term effects of preschool-structural estimates from a discrete dynamic programming model,” *Journal of Econometrics*, 191, 164–175.
- HECKMAN, J. J., J. STIXRUD, AND S. URZUA (2006): “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior,” *Journal of Labor Economics*, 24, 411–482.

- HENDREN, N. AND B. SPRUNG-KEYSER (2020): “A Unified Welfare Analysis of Government Policies,” *Quarterly Journal of Economics*, 135, 1209–1318, view the estimates online at www.policyinsights.org Watch the Econimate Video.
- HILBER, C. AND T. TURNER (2014): “The mortgage interest deduction and its impact on homeownership decisions,” *The Review of Economics and Statistics*, 96, 618–637.
- HOXBY, C. M. (2018): “Online Postsecondary Education and Labor Productivity,” in *Education, Skills, and Technical Change: Implications for Future US GDP Growth*, ed. by C. R. Hulten and V. A. Ramey, Chicago: University of Chicago Press, vol. 2, 401–460, conference held in October 2015, published in December 2018.
- IBGE (2010): “Household internet access,” *Population Census*.
- IHLANFELDT, K. (2007): “The effect of land use regulation on housing and land prices,” *Journal of Urban Economics*, 61, 420–435.
- JACKSON, C. K. (2018): “What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes,” *Journal of Political Economy*, 126, 2072–2107.
- JACOB, B. A. (2004): “Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago,” *American Economic Review*, 94, 233–258.
- JACOB, B. A., M. KAPUSTIN, AND J. LUDWIG (2015): “The impact of housing assistance on child outcomes: Evidence from a randomized housing lottery,” *The Quarterly Journal of Economics*, 130, 465–506.
- JACOB, B. A. AND J. LUDWIG (2012): “The effects of housing assistance on labor supply: Evidence from a voucher lottery,” *American Economic Review*, 102, 272–304.
- KALIL, A. AND K. M. ZIOL-GUEST (2008): “Parental employment circumstances and children’s academic progress,” *Social Science Research*, 37, 500–515.
- KLING, J., J. LIEBMAN, AND L. KATZ (2007a): “Experimental Analysis of Neighborhood Effects,” *Econometrica*, 75, 83–119.
- KLING, J. R., J. B. LIEBMAN, AND L. F. KATZ (2007b): “Experimental Analysis of Neighborhood Effects,” *Econometrica*, 75, 83–119.
- KLING, J. R., J. LUDWIG, AND L. F. KATZ (2005): “Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment*,” *The Quarterly Journal of Economics*, 120, 87–130.
- KOFOED, M. S., L. GEBHART, D. GILMORE, AND R. MOSCHITTO (2024): “Zooming to Class? Experimental Evidence on College Students’ Online Learning during COVID-19,” *American Economic Review: Insights*, 6, 324–40.

- KUMAR, T. (2021): “The housing quality, income, and human capital effects of subsidized homes in urban India,” *Journal of Development Economics*, 153, 102738.
- LALIBERTÉ, J.-W. (2021): “Long-Term Contextual Effects in Education: Schools and Neighborhoods,” *American Economic Journal: Economic Policy*, 13, 336–77.
- LAVY, V., M. D. PASERMAN, AND A. SCHLOSSER (2012): “Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom,” *The Economic Journal*, 122, 208–237.
- LEE, D. S. AND T. LEMIEUX (2010): “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature*, 48, 281–355.
- LUDWIG, J., G. J. DUNCAN, L. A. GENNETIAN, L. F. KATZ, R. C. KESSLER, J. R. KLING, AND L. SANBONMATSU (2013a): “Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity,” *American Economic Review*, 103, 226–31.
- (2013b): “Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity,” *American Economic Review*, 103, 226–31.
- MARTINEZ, J. AND E. LOPEZ (2018): “Vivienda masiva en barrios centrales: ¿quién es responsable de la densificación excesiva?” *ARQ (Santiago)*, 144 – 153.
- MCCRARY, J. (2008): “Manipulation of the running variable in the regression discontinuity design: A density test,” *Journal of Econometrics*, 142, 698–714, the regression discontinuity design: Theory and applications.
- MDS (2017): “Encuesta CASEN año 2017,” *Ministerio de Desarrollo Social y Familia, Observatorio Social*.
- MILLS, G., D. GUBITS, L. ORR, D. LONG, J. FEINS, B. KAUL, AND D. MCINNIS (2006): “Effects of Housing Vouchers on Welfare Families,” Tech. rep., Abt Associates, Prepared for the U.S. Department of Housing and Urban Development, Cambridge, MA.
- MINEDUC: AGENCIA DE CALIDAD DE LA EDUCACIÓN (2024): “Base de Datos de la Agencia de Calidad de la Educación [2003-2024],” Santiago, Chile.
- MINVU (2024): “Administrative data on applicants to the Chilean homeownership voucher from 2011 to 2020.” .
- MOFFITT, R. (1992): “Incentive Effects of the U.S. Welfare System: A Review,” *Journal of Economic Literature*, 30, 1–61.
- MOUNTJOY, J. (2022): “Community Colleges and Upward Mobility,” *American Economic Review*, 112, 2580–2630.
- NCES (2022): “Digest of Education Statistics 2022, Table 311.15,” Tech. rep., U.S. Department of Education.

- NEILSON, C. ET AL. (2013): “Targeted vouchers, competition among schools, and the academic achievement of poor students,” *Job Market Paper*, 48.
- OREOPOULOS, P. AND K. G. SALVANES (2011): “Priceless: The Nonpecuniary Benefits of Schooling,” *Journal of Economic Perspectives*, 25, 159–84.
- PETRIN, A. (2002): “Quantifying the benefits of new products: The case of the minivan,” *Journal of political Economy*, 110, 705–729.
- PINTO, R. (2022): “Beyond Intention to Treat: Using the Incentives in Moving to Opportunity to Identify Neighborhood Effects,” Working Paper w29167, National Bureau of Economic Research.
- PNAD (2019): “Household internet access,” *Pesquisa Nacional por Amostra de Domicílios*.
- POLLAKOWSKI, H. O., D. H. WEINBERG, F. ANDERSSON, J. C. HALTIWANGER, G. PALLONI, AND M. J. KUTZBACH (2022): “Childhood Housing and Adult Outcomes: A Between-Siblings Analysis of Housing Vouchers and Public Housing,” *American Economic Journal: Economic Policy*, 14, 235–72.
- POTERBA, J. AND T. SINAI (2008): “Tax Expenditures for Owner-Occupied Housing: Deductions for Property Taxes and Mortgage Interest and the Exclusion of Imputed Rental Income,” *American Economic Review*, 98, 84–89.
- RAPPOPORT, D. E. (2016): “Do Mortgage Subsidies Help or Hurt Borrowers?” *Finance and Economics Discussion Series 2016*, 081.
- ROJAS-AMPUERO, F. AND F. CARRERA (2022): “Sent away: The long-term effects of slum clearance on children and families,” *JMP*.
- ROTHSTEIN, J. (2010): “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement*,” *The Quarterly Journal of Economics*, 125, 175–214.
- ROUSE, C. E. (1995): “Democratization or Diversion? The Effect of Community Colleges on Educational Attainment,” *Journal of Business & Economic Statistics*, 13, 217–224.
- (1998): “Do two-year colleges increase overall educational attainment? Evidence from the states,” *Journal of Policy Analysis and Management*, 17, 595–620.
- SACERDOTE, B. (2001): “Peer Effects with Random Assignment: Results for Dartmouth Roommates*,” *The Quarterly Journal of Economics*, 116, 681–704.
- SAIZ, A. (2010): “The Geographic Determinants of Housing Supply,” *The Quarterly Journal of Economics*, 125, 1253–1296.
- SANCHEZ, C. (2023): “Equilibrium Consequences of Vouchers Under Simultaneous Extensive and Intensive Margins Competition,” Tech. rep., Working Papers.

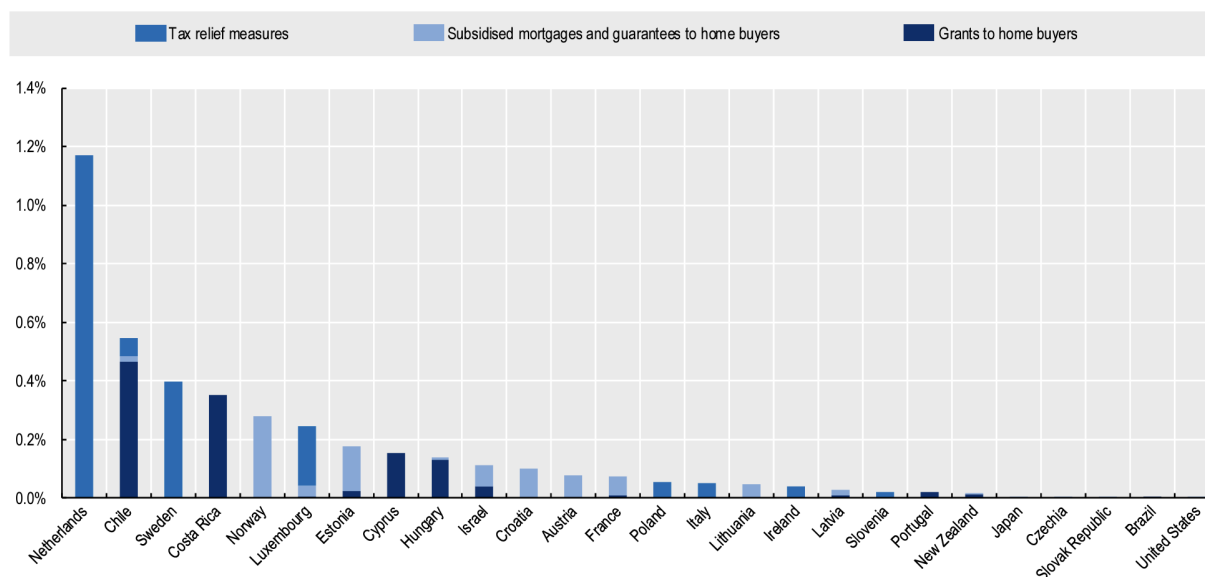
- SCHWARTZ, A. E., L. STIEFEL, AND S. A. CORDES (2020): “Housing Vouchers and Children’s Educational Outcomes: Evidence from New York City,” *Journal of Policy Analysis and Management*, 39, 131–158.
- SEIM, K. (2006): “An empirical model of firm entry with endogenous product-type choices,” *The RAND Journal of Economics*, 37, 619–640.
- SELMAN, J. (2022): “Rental Voucher Programs in Middle Income Countries: Quasi- experimental Evidence from Chile,” *Working Paper*.
- SHEEHAN, M. (2024): “Everyday verticality: Migrant experiences of high-rise living in Santiago, Chile,” *Urban Studies*, 61, 726–742.
- SINAI, T. AND J. GYOURKO (2004): “The (Un)changing Geographical Distribution of Housing Tax Benefits: 1980-2000,” *Tax Policy and the Economy*, 18, 175–208, publisher: University of Chicago Press.
- SOLTAS, E. (2024): “Tax Incentives and the Supply of Low-Income Housing,” *Working Paper*.
- SWEETING, A. (2009): “The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria,” *The RAND Journal of Economics*, 40, 710–742.
- THISSE, J.-F. AND M. FUJITA (2002): “Economics of Agglomeration,” *Cambridge: Cambridge University Press*.
- TOCTOC (2024): “Property transactions at the Metropolitan Region of Santiago from 1998 to 2020,” .
- UN (2023): “The Sustainable Development Goals Report 2023,” *United Nations DESA*.
- UN-HABITAT (2022): “World Cities Report 2022: Envisaging the future of cities,” .
- VALENZUELA, P. (2023): “Santiago’s Zoning Policy,” *Mimeo*.
- VAN DIJK, W. (2019): “The socio-economic consequences of housing assistance,” *University of Chicago Kenneth C. Griffin Department of Economics job market paper*, 0–46 i–xi, 36.
- WORLD BANK (2023): “Exciting, Boring, and Nonexistent Skylines Vertical Building Gaps in Global Perspective,” *Urban, Disaster Risk Management, Resilience and Land Global Practice*.
- ZELTZER, D., L. EINAV, J. RASHBA, AND R. D. BALICER (2023): “The Impact of Increased Access to Telemedicine,” *Journal of the European Economic Association*, 22, 712–750.
- ZHANG, Y., E. HADDAD, B. TORRES, AND C. CHEN (2011): “The Reciprocal Relationships Among Parents’ Expectations, Adolescents’ Expectations, and Adolescents’ Achievement: A Two-Wave Longitudinal Analysis of the NELS Data,” *Journal of Youth and Adolescence*, 40, 479–489.

Appendix A

Appendix of The Effect of a Homebuyer Subsidy on Children

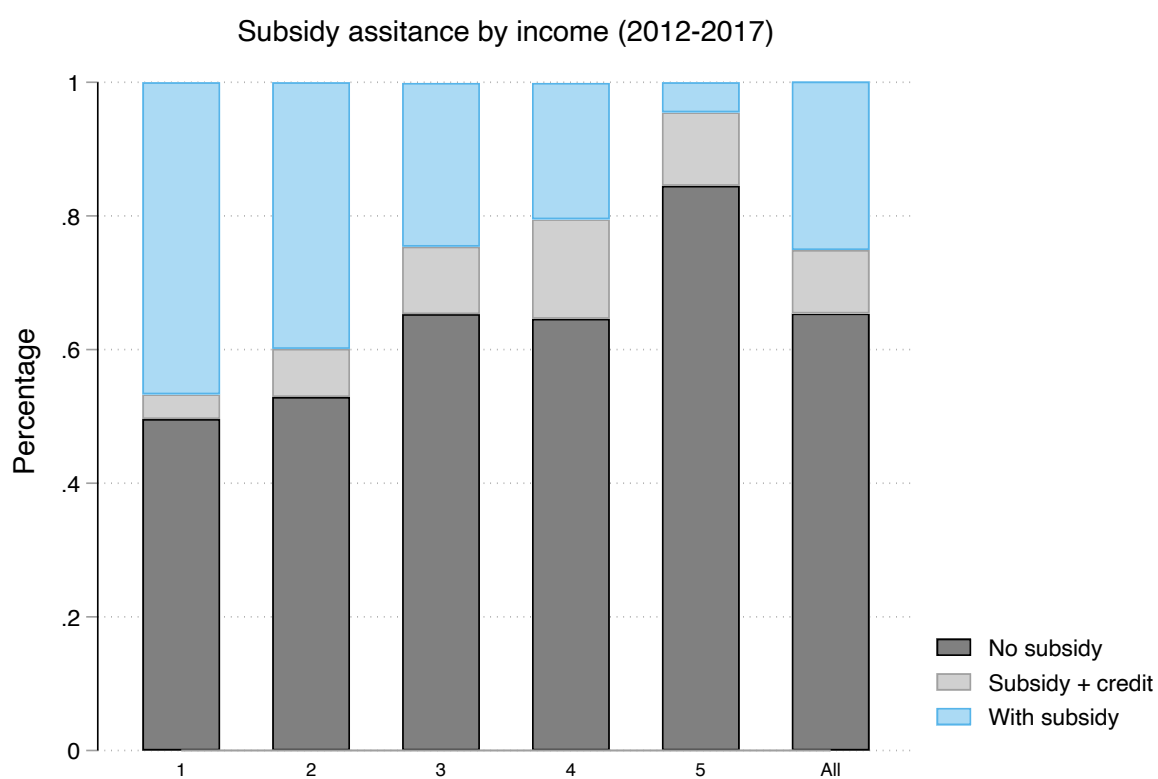
A.1 Additional Figures

Figure A1: Public spending, financial support and tax relief for homebuyers



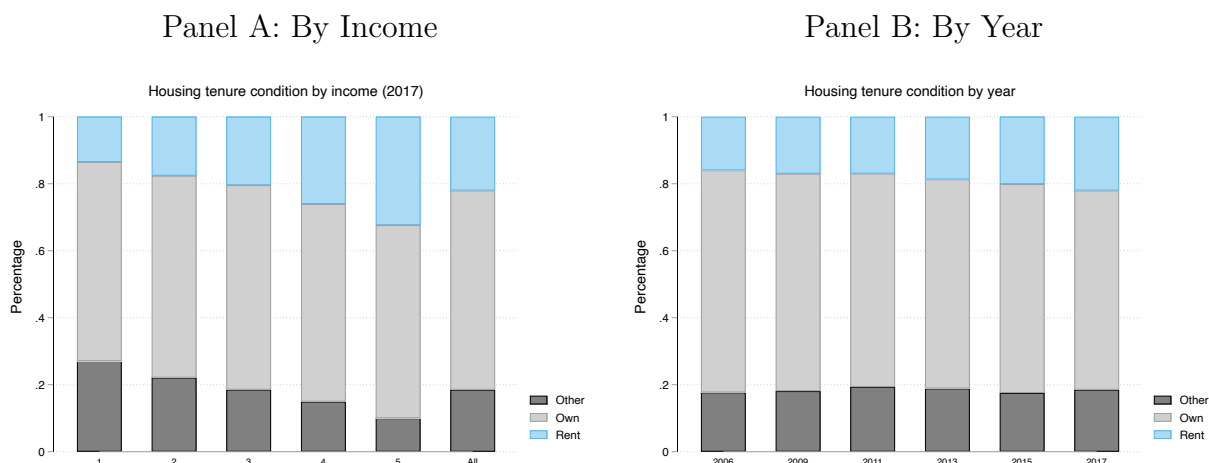
Notes: Figure displays the amount of public spending, financial support, and tax relief for homebuyers. Amounts are shown as % of GDP, 2022 or the latest available year. Chile is the country that spends the most on grants to homebuyers. *Source:* OECD Questionnaire on Affordable and Social Housing (QuASH).

Figure A2: Subsidy Assistance by Income level



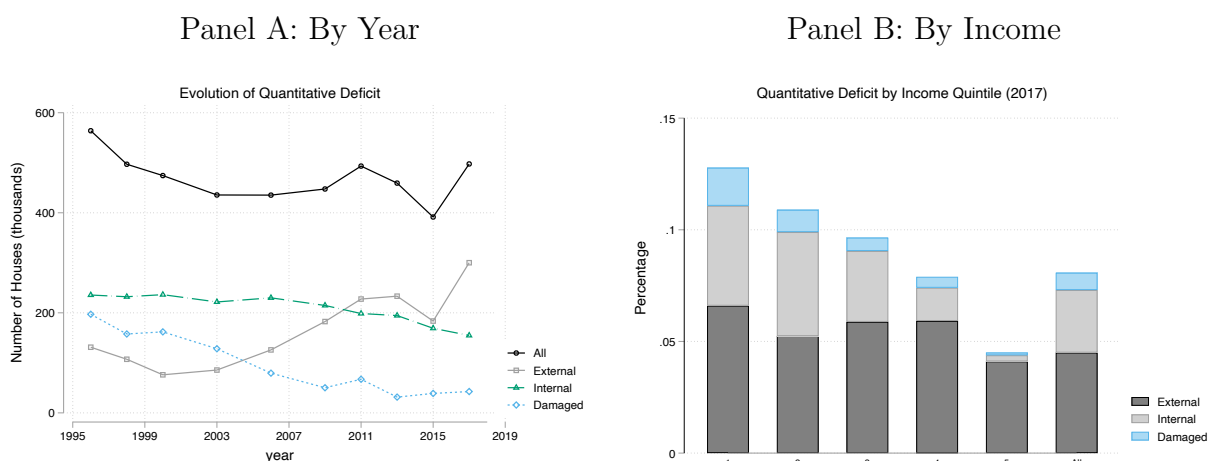
Notes: This figure shows the distribution of the percentage of houses sold using 1) Only the subsidy; 2) Combination of subsidy and credit; and 3) No subsidy, by income quintile. It considers all of the housing units sold between 2012 and 2017. More than half of the units sold to the lowest quintile were bought only using the subsidy.

Figure A3: Housing Tenure



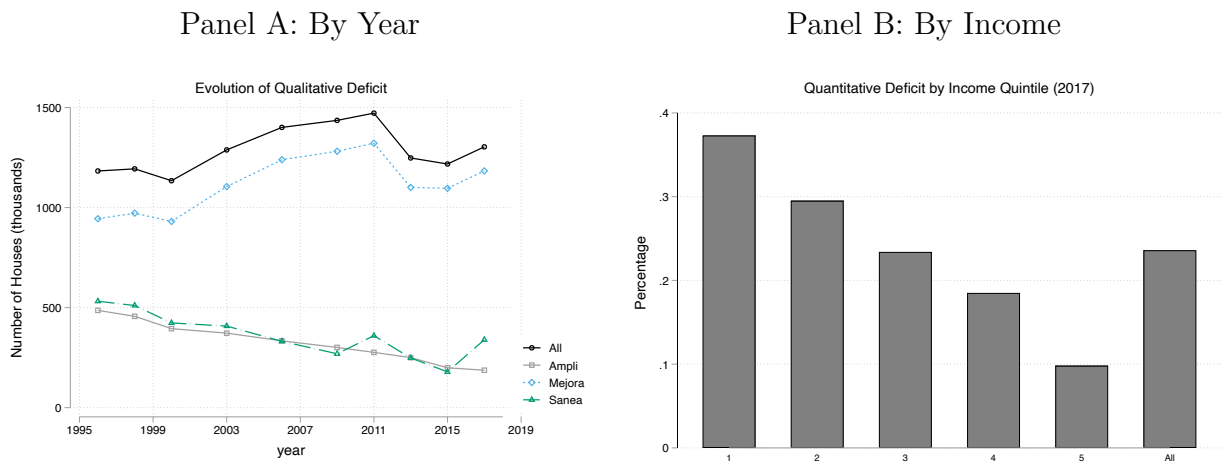
Notes: This figure shows the distribution of the housing tenure condition by income quintile and year. In Panel A, we show the distribution by income quintile in the year 2017. In Panel B, we show the aggregated distribution by year, from 2006 to 2017.

Figure A4: Quantitative Housing Deficit



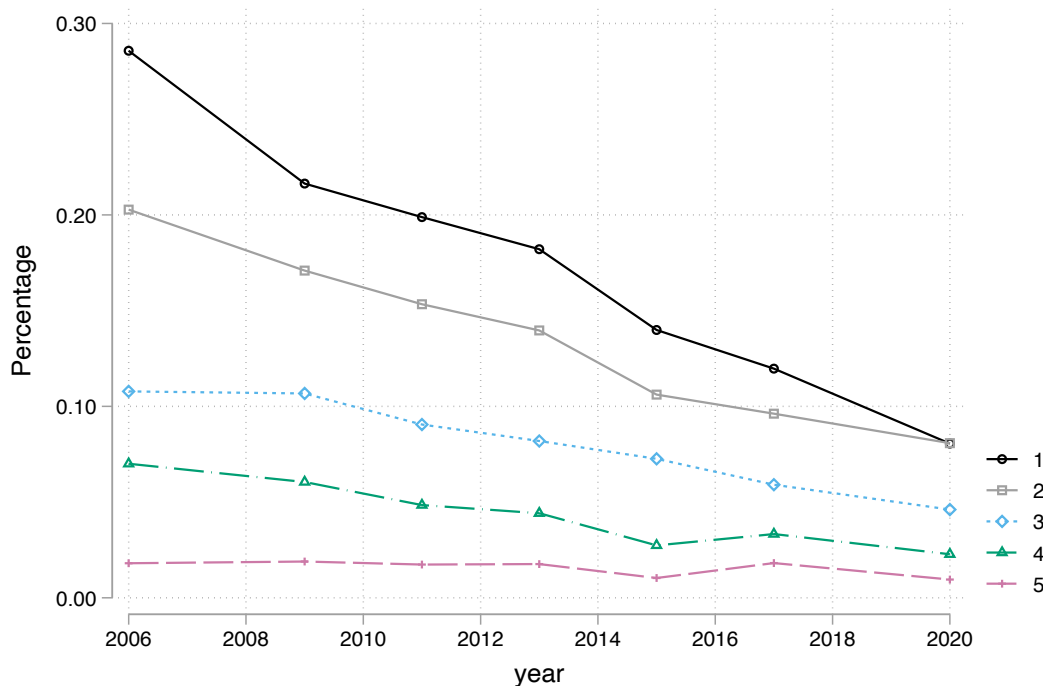
Notes: This figure shows the distribution of the quantitative housing deficit by income quintile and year. *External* corresponds to more than one household living in the same unit; *Internal* to more than one family living in the same unit with more than 2.5 persons per bedroom; and *Damaged* corresponds to housing units with irrecoverable damage. Panel A shows the evolution of this deficit by year, while Panel B displays the deficit by income quintile in 2017.

Figure A5: Qualitative Housing Deficit



Notes: This figure displays the distribution of the qualitative housing deficit by income quintile and year. *Ampli* corresponds to overcrowded houses (more than 2.5 persons per bd); *Mejora* to more than one family living in the same unit with more than 2.5 persons per bedroom; and *Z* corresponds to housing units with irrecoverable damage. The total number includes all of the housing units that need at least one of the 3 categories.

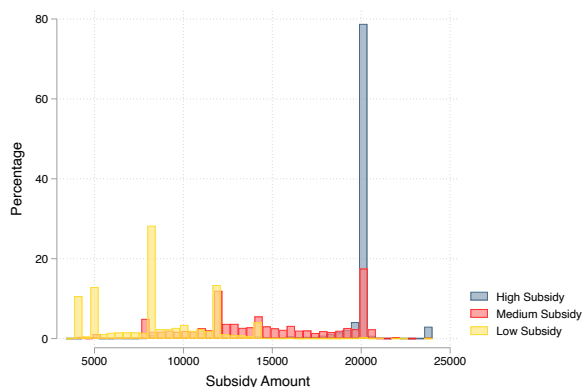
Figure A6: Evolution of Overcrowding by Income Quintile



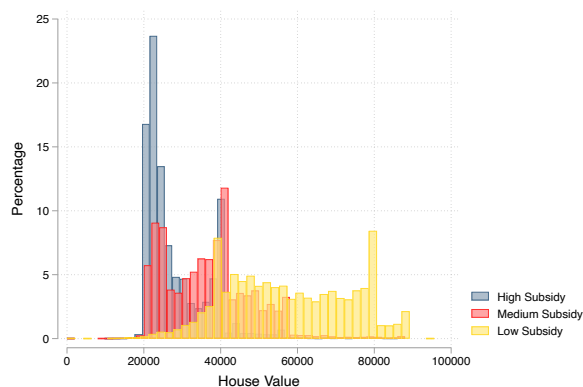
Notes: This figure shows the evolution of the distribution of overcrowded households by income quintile over the years, measured as housing units with more than 2.5 people per bedroom. We show that, even in 2020, many families live in overcrowded housing conditions, especially the ones in the lower quintiles of the income distribution.

Figure A7: Distribution of Vouchers and Housing Prices for Recipients

Panel A: Voucher Amounts

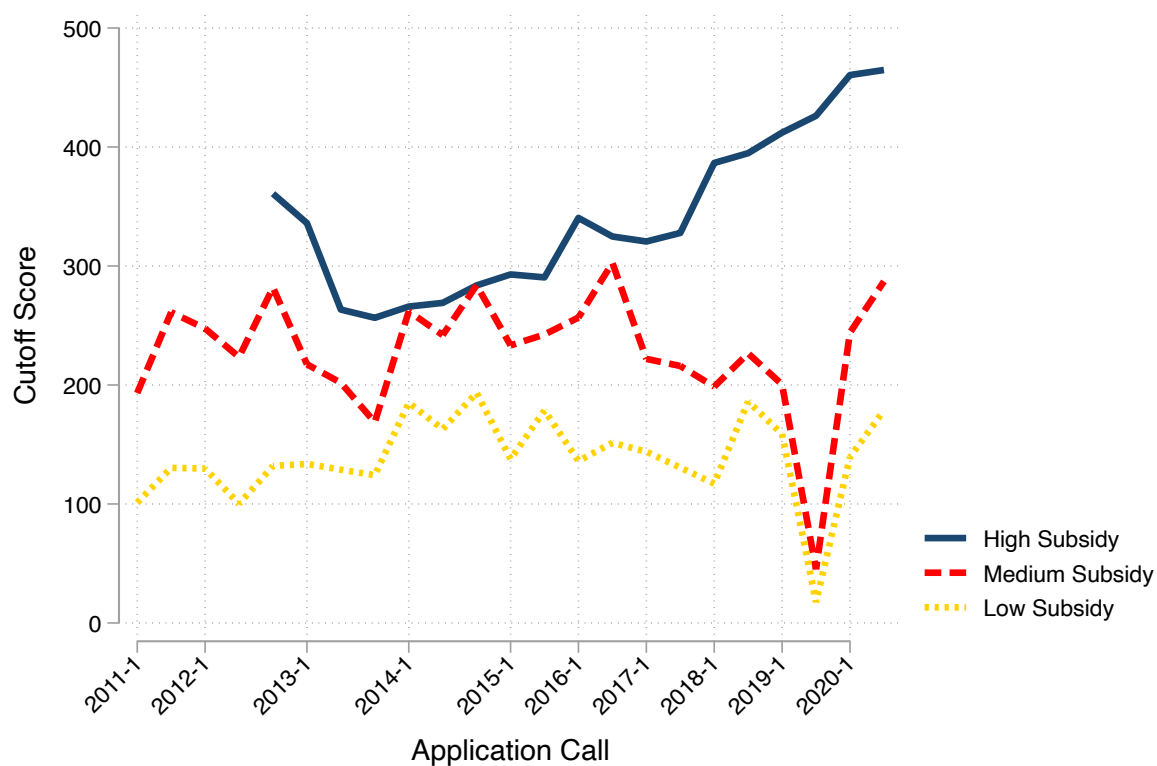


Panel B: Housing Prices



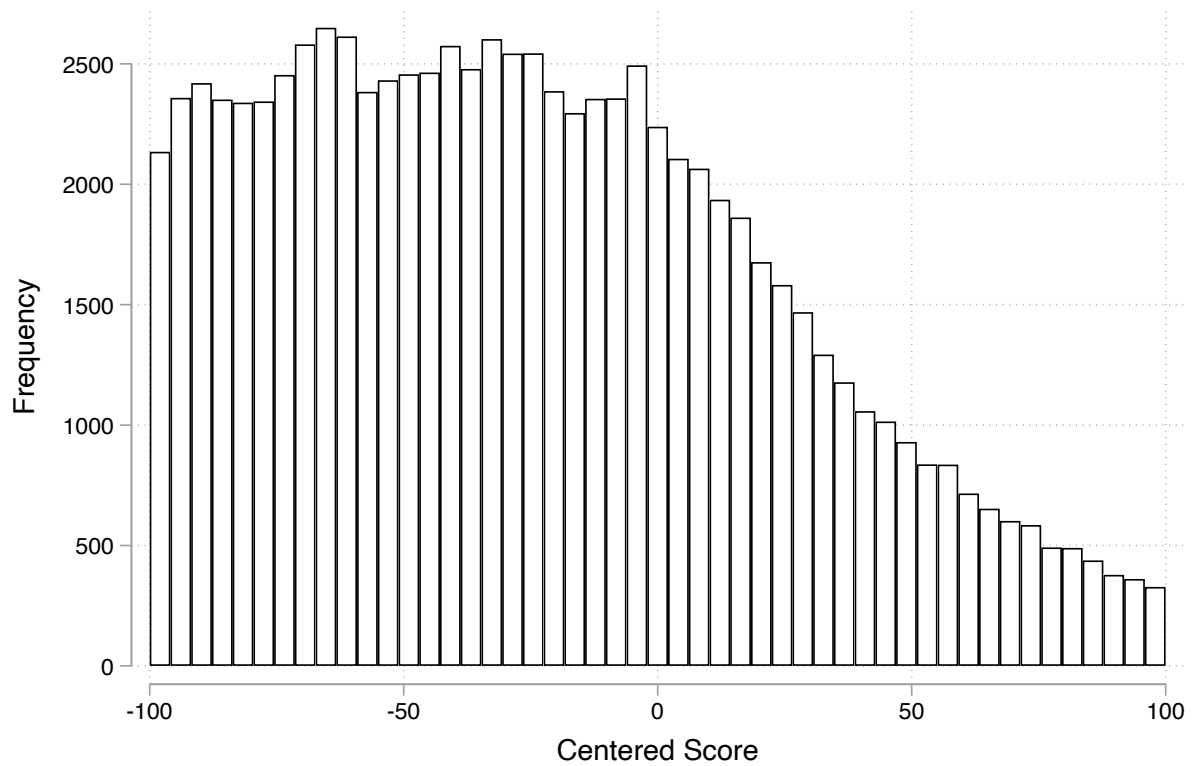
Notes: This figure shows the distribution of voucher amounts and housing prices for voucher recipients. Panel A presents the distribution of voucher amounts by subsidy tier, while Panel B shows the distribution of housing prices.

Figure A8: Cutoff Evolution by Call and Tier



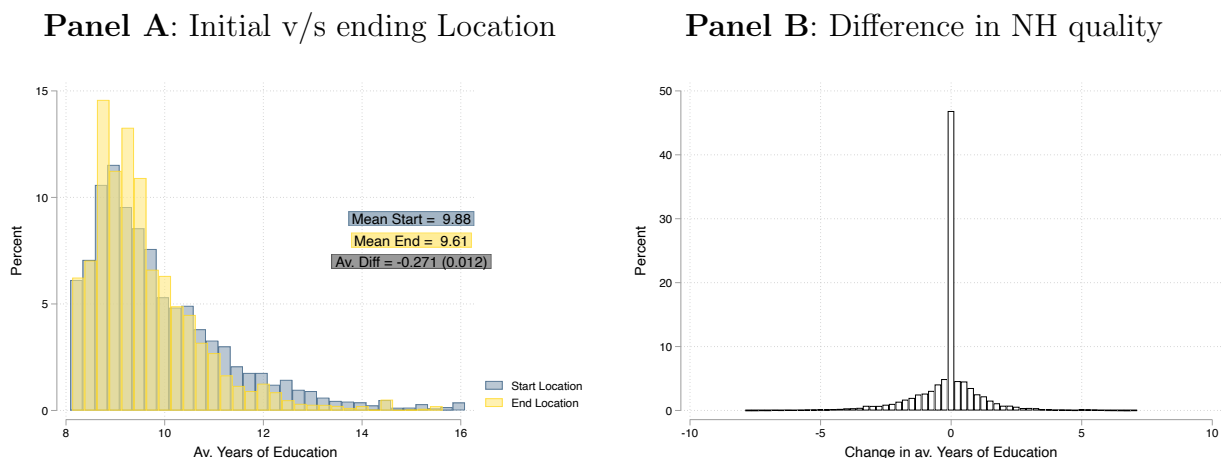
Notes: This figure displays the cutoff scores by application call and tier for Santiago. The variance in cutoff scores by application call is mostly driven by changes in the pool of applicants, as the budget for different regions and tiers is relatively stable over time.

Figure A9: Score Distribution



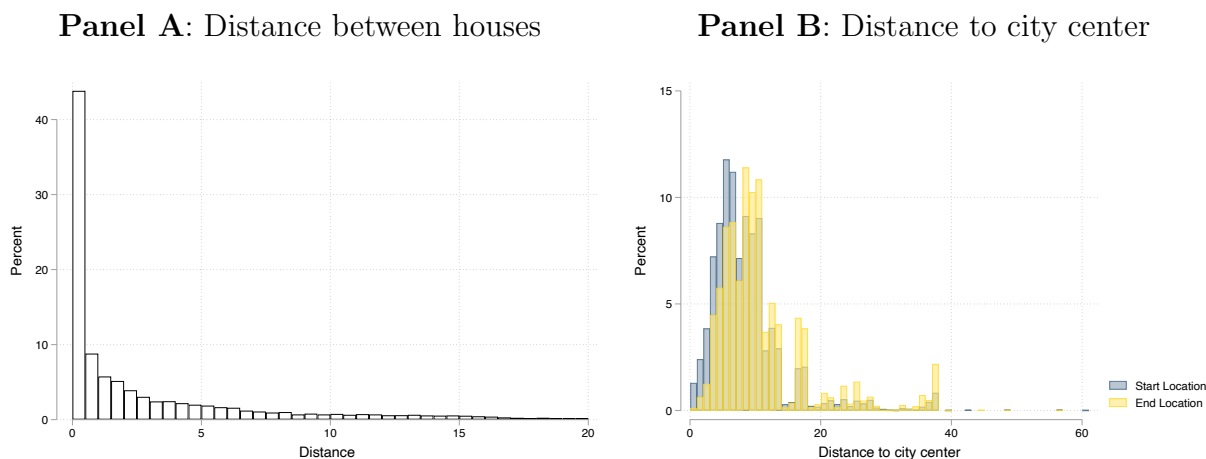
Notes This figure shows the distribution of scores relative to the cutoff, pooling across application calls. There is a large mass on both sides of the cutoff, and there is no evidence of bunching at zero. We take this as evidence that there is no manipulation of the scores and that we will have enough power to detect effects with the regression discontinuity

Figure A10: Distribution of average neighborhood quality



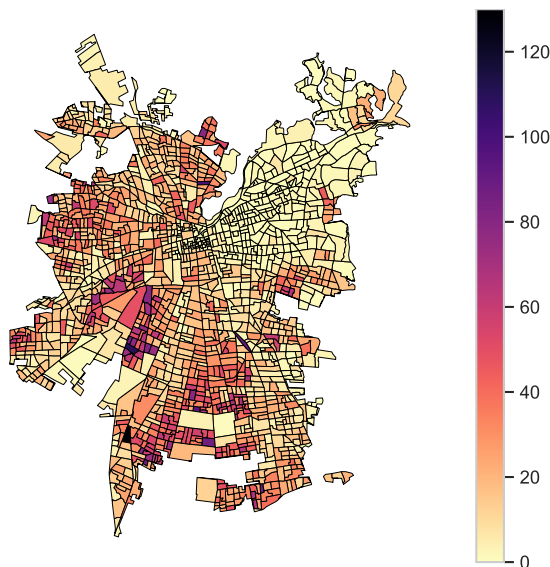
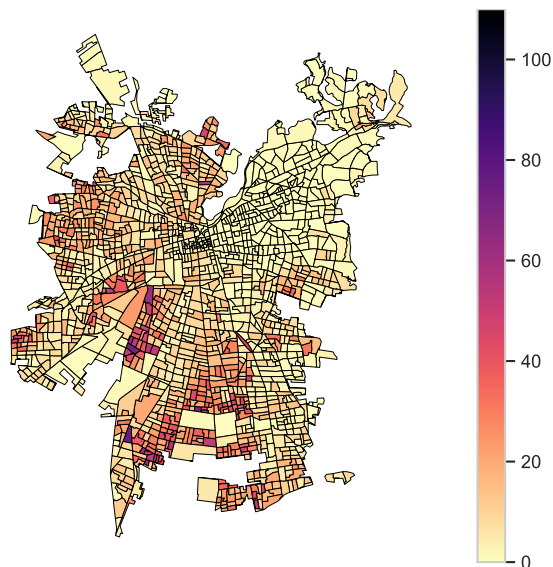
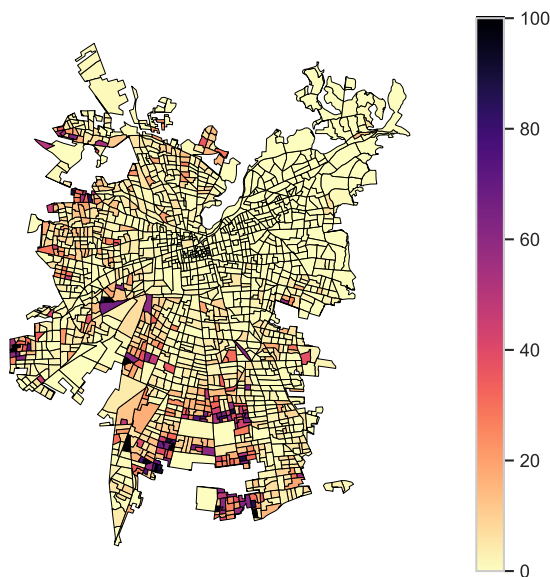
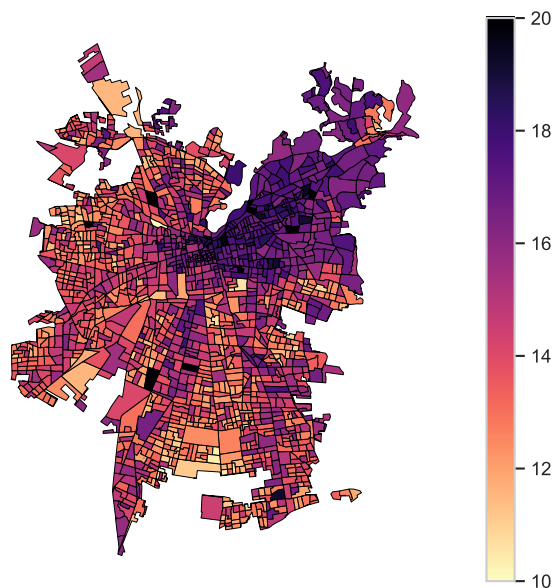
Notes: This figure shows the distribution of neighborhood quality measured as average years of schooling in the census tract. Panel A presents the distribution of average years of education associated with the initial and ending locations of voucher users. Panel B shows the distribution of the change in years of education from the initial and ending locations. We show that almost half of voucher users stay in a census tract with the same average years of education. For the ones moving to different neighborhoods, we find that they tend to move to slightly less educated neighborhoods.

Figure A11: Distance between houses and to the city center



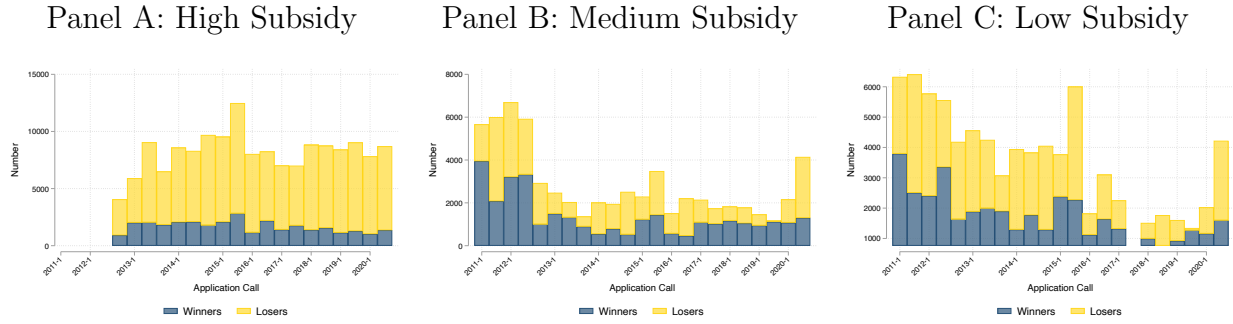
Notes: This figure presents the distribution of the distance between the initial and ending locations of the houses for voucher users and the distribution of distances to the city center. Panel A presents the distribution of the distance between former and new housing units. Panel B shows the distribution of the distance from the initial and ending locations to the city center. We show that more than half of voucher users do not move more than one mile away, and most of the remaining move relatively close. However, they move to places slightly further from the city center.

Figure A12: Maps of the distribution of winners and users

Panel A: Initial location of winners**Panel B:** Initial location of users**Panel C:** Ending location of users**Panel D:** Av. Years of schooling

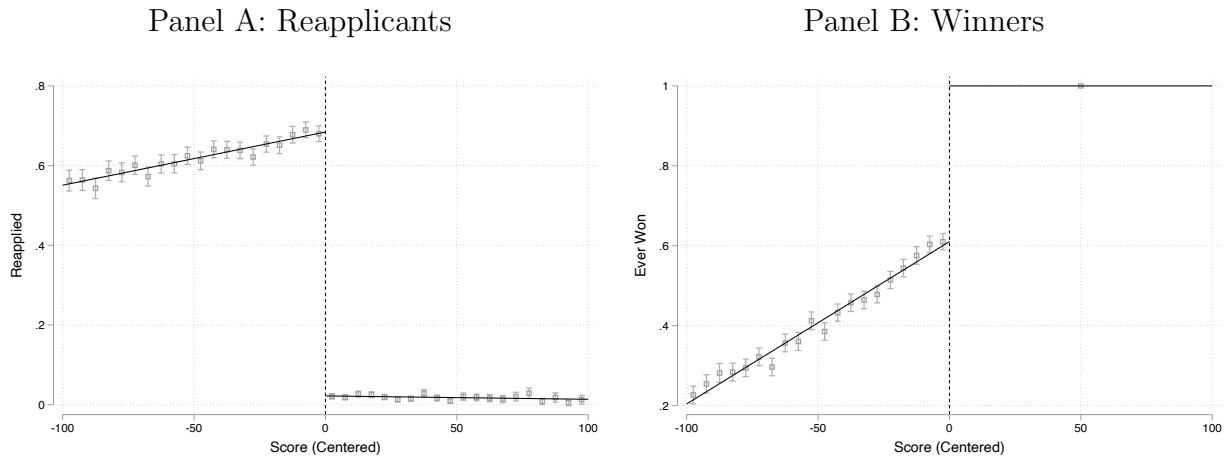
Notes: This figure displays the distribution of the number of households in each census tract. Panel A shows the distribution of the location at the time of application for voucher-awarded households, and Panel B shows the distribution of the location for voucher users. Panel C displays the distribution of the final location of voucher users, and panel D presents the average years of schooling by census tract, using census data from 2017.

Figure A13: Evolution of winners



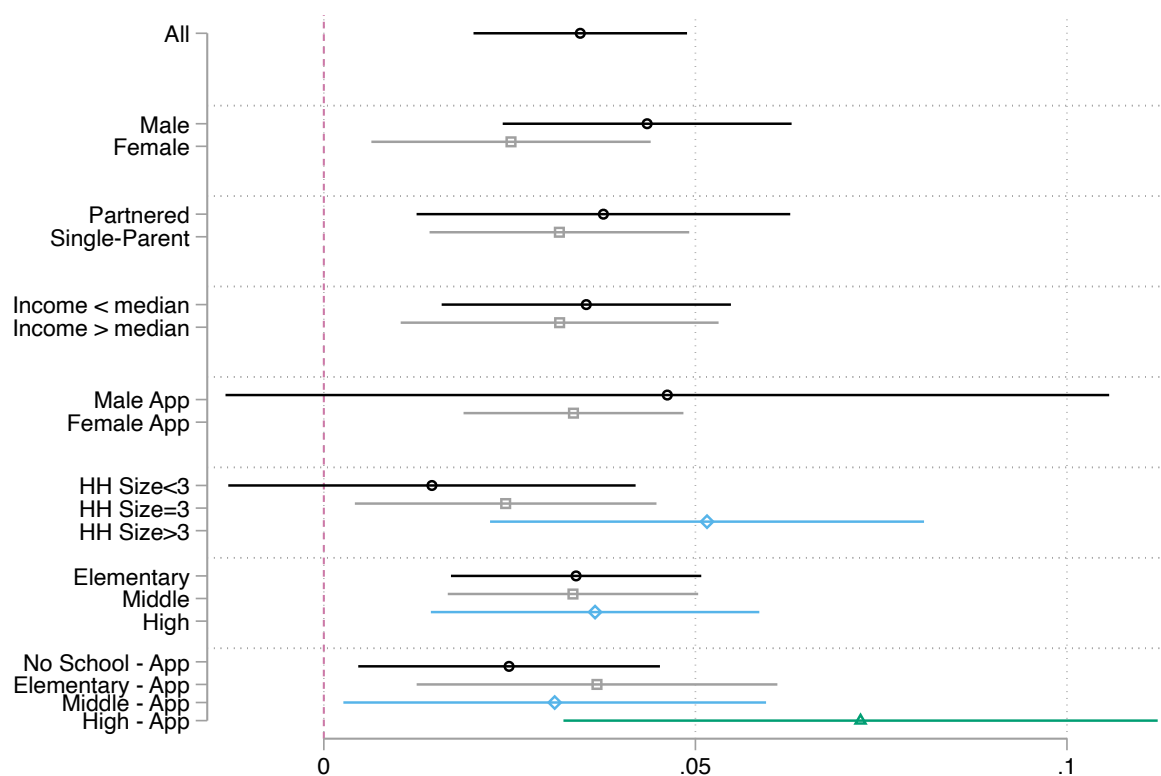
Notes: This figure displays the evolution of applicants in Santiago by application call and tier, separated by awarded and non-awarded applicants.

Figure A14: Reapplicants and Winners around the cutoff



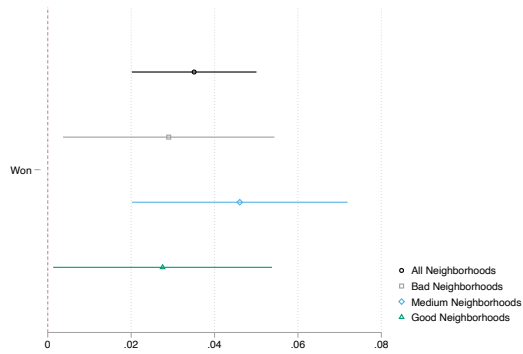
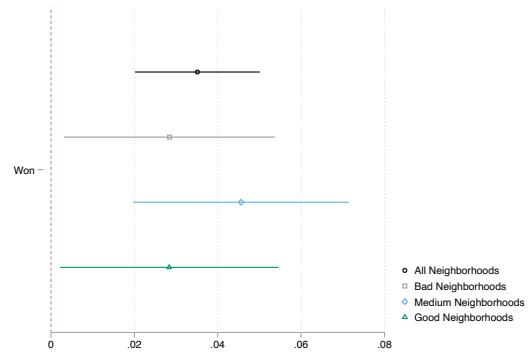
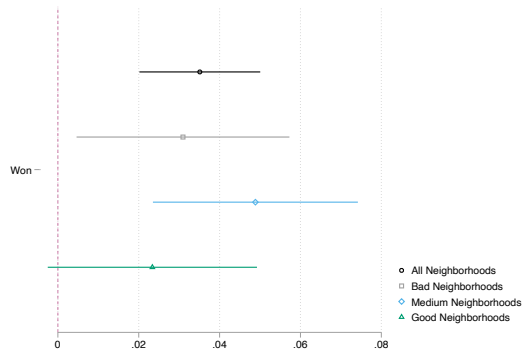
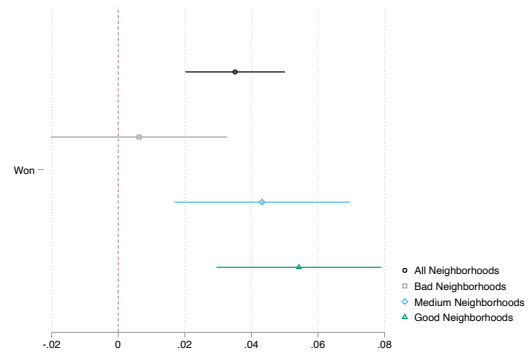
Notes: This figure shows the result of the regression discontinuity, using [Equation 1.1](#) on the probability of reapplying and winning the subsidy. Panel A uses an indicator of reapplying in the future, and Panel B an indicator of ever winning the subsidy as an outcome. We show that almost 70% of non-awarded applicants end up reapplying in the future, while only 5% of awarded applicants do so. Moreover, over 60% of non-awarded applicants end up winning the voucher in the future.

Figure A15: Heterogeneity by Demographics: Grades



Notes: This figure displays the heterogeneous effects depending on demographic characteristics for the ITT estimates on grades. Each horizontal bar reports the coefficient of the jump at the cutoff and the confidence interval of the estimate at the 10% significance level, using Equation 1.1 and fitting a linear polynomial at each side of the cutoff. All estimates include year and gender by age fixed effects and consider scores the IMSE-optimal bandwidth. Standard errors are clustered at the applicant level.

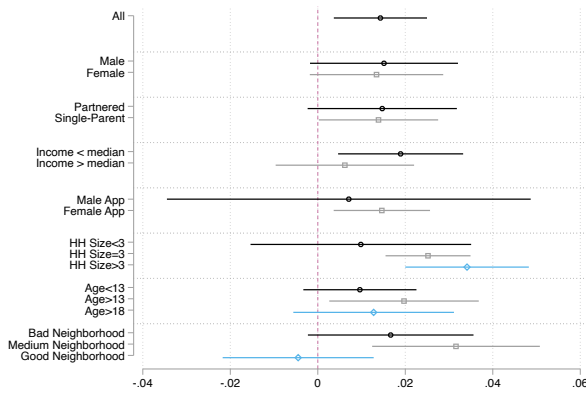
Figure A16: Heterogeneity on Neighborhood: alternative measures

Panel A: Av. Years of Schooling**Panel B: Socio-Territorial Index****Panel C: Overcrowding****Panel D: Material Quality**

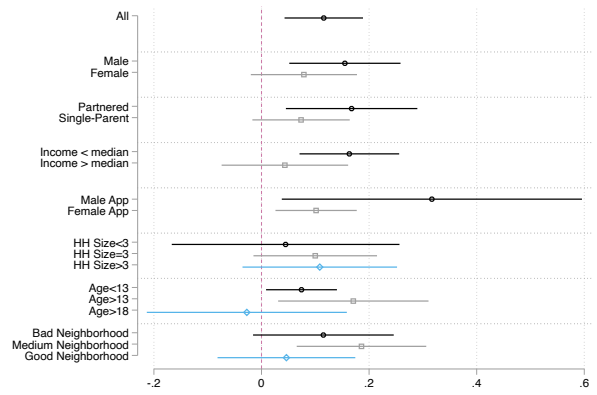
Notes: This figure displays the heterogeneous effects of the result of the estimate of the ITT estimate for grades depending on neighborhood quality, pooling years after the application. Measures of neighborhood quality include average years of schooling (A), socio-territorial index (B), overcrowding index (C), and material quality index (D). Standard errors are clustered at the applicant level, and regression considers scores using the IMSE-optimal bandwidth. Controls include year and gender-by-age fixed effects. ***, **, * indicates significance at 1, 5, and 10%.

Figure A17: Heterogeneities in high school attainment and years of college

Panel A: High School Completion

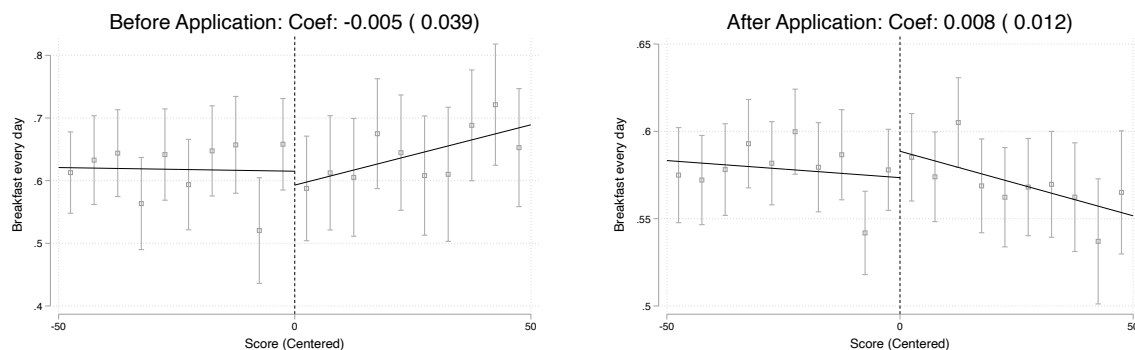
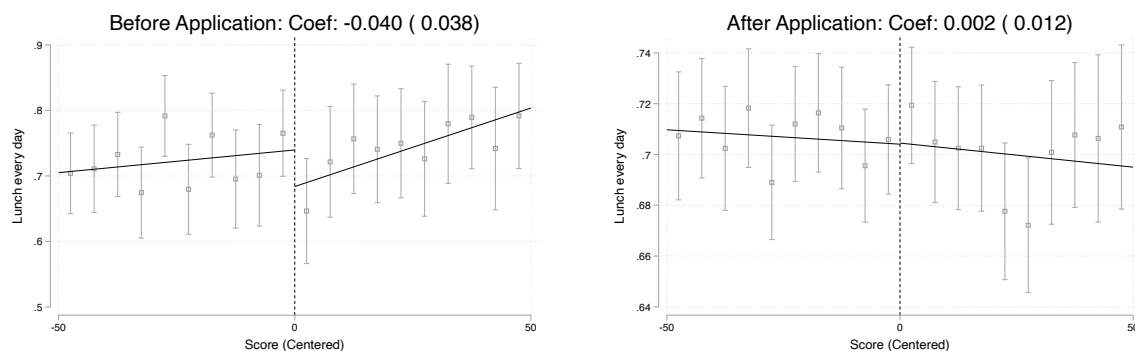
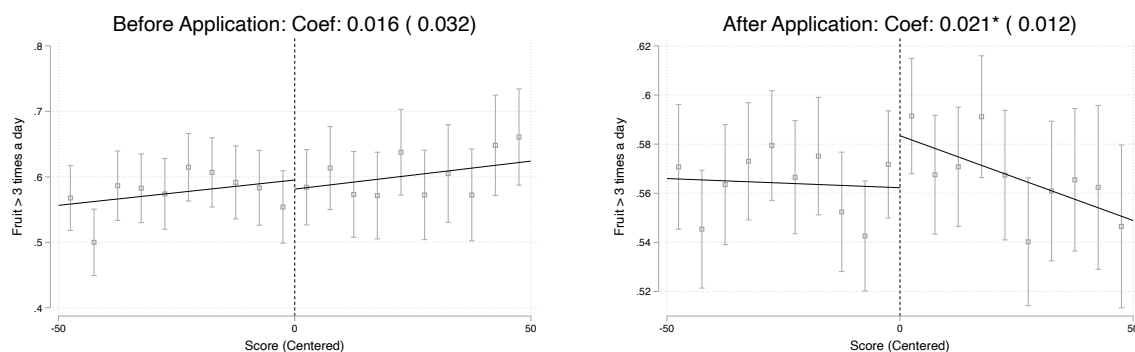


Panel B: Years of College



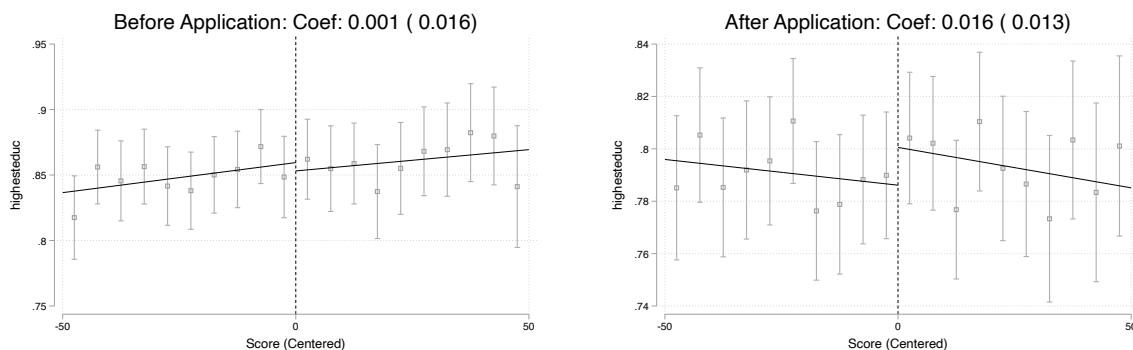
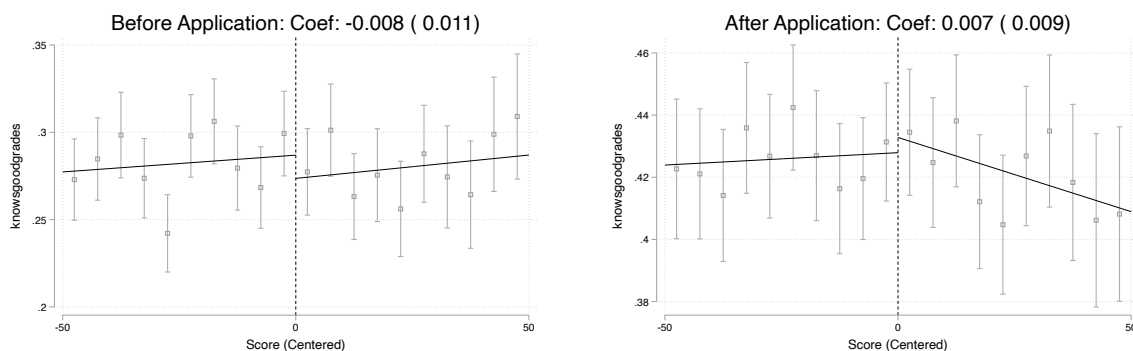
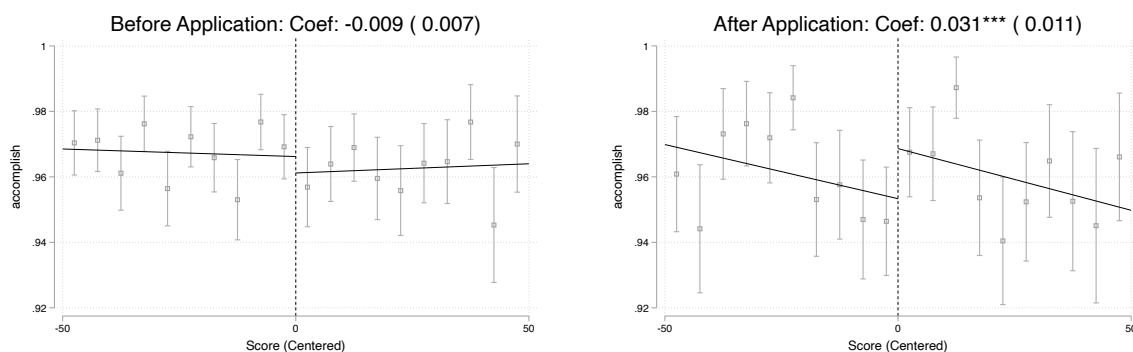
Notes: This figure displays the heterogeneous effects depending on demographic characteristics for the ITT estimates on high school completion and years of college. Each horizontal bar reports the coefficient of the jump at the cutoff and the confidence interval of the estimate at the 10% significance level, using [Equation 1.1](#). All estimates include year and gender by age fixed effects and consider scores the IMSE-optimal bandwidth. Standard errors are clustered at the applicant level.

Figure A18: RD of children's behavior regarding food consumption

Panel A: Breakfast**Panel B: Lunch****Panel C: Fruit Consumption**

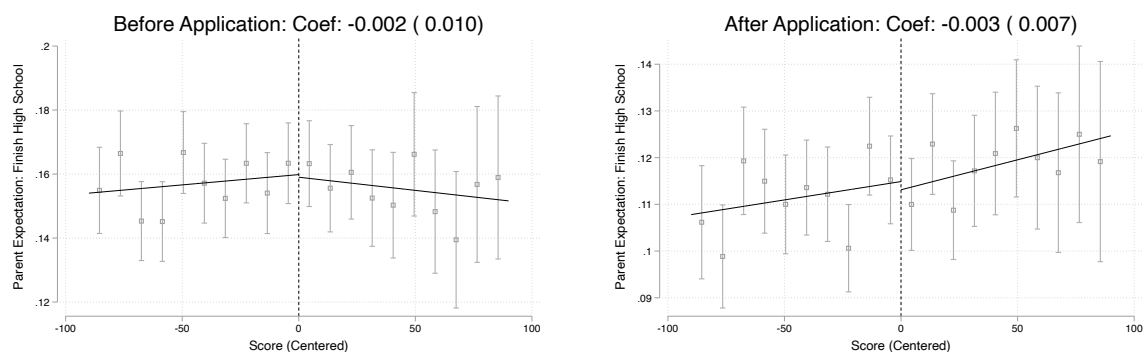
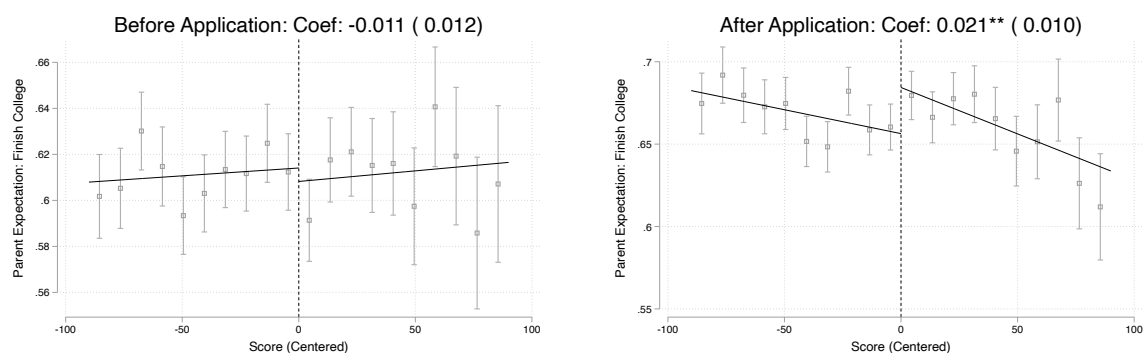
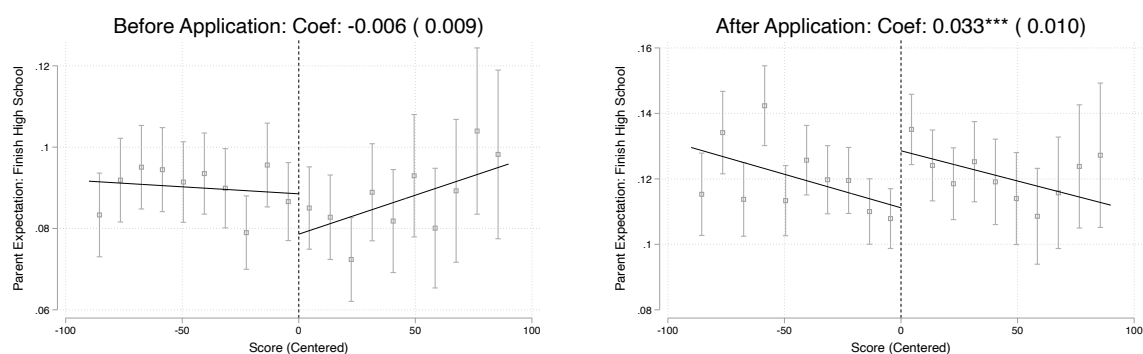
Notes: This figure shows the result of the ITT estimate of the regression discontinuity regarding children's food consumption before and after the application. In Panels A, B, and C, the outcome variable indicates breakfast, lunch, and fruit consumption, respectively. For breakfast and lunch, the outcome variable indicates having had lunch on all school days. For fruit consumption, it indicates consuming two or more times a day. Standard errors are clustered at the applicant level, and scores are considered using the IMSE-optimal bandwidth. Controls include year and gender-by-age fixed effects. ***, **, * indicates significance at 1, 5, and 10%.

Figure A19: RD of children's expectations

Panel A: Will attend post-secondary education**Panel B: Knows they can get good grades****Panel C: Will accomplish what they want as adults**

Notes: This figure shows the result of the ITT estimate of the regression discontinuity regarding children's expectations about the future. In Panel A, the outcome variable indicates whether the student thinks they will complete post-secondary education. Panels B and C indicate whether the student believes they can get good grades and will accomplish what they want as adults, respectively. Standard errors are clustered at the applicant level, and scores are considered using the IMSE-optimal bandwidth. Controls include year and gender-by-age fixed effects. ***, **, * indicates significance at 1, 5, and 10%.

Figure A20: RD of parents' expectations about children's education

Panel A: High School completion**Panel B: College completion****Panel C: Graduate studies completion**

Notes: This figure shows the result of the ITT estimate of the regression discontinuity regarding the parents' expectations regarding the maximum education level reached by their child before and after the application. In Panel A, the outcome variable indicates that parents think their children will finish high school. In Panels B and C, the outcome variable indicates that parents think their child will complete a college degree and a graduate degree, respectively. Standard errors are clustered at the applicant level, and scores are considered using the IMSE-optimal bandwidth. Controls include year and gender-by-age fixed effects. ***, **, * indicates significance at 1, 5, and 10%.

A.2 Additional Tables

Table A1: Score Components

Name of the Variable	Description	Score	Average Score
Household Size	Number of Household Members besides the applicant	40	50
Single-Parent	Indicator applicant is single-parent	35	13.3
Elderly	Number of elderly in the household	30	2.2
Savings	Amount Saved in addition to the minimum	1	40.2
Past Applications	Number of applications in the past	20	35
Vulnerability	Vulnerability Index	1	51
Disability	Number of household members with disabilities	30	1.3
Valech	Part of the Valech Report	100	0.2
Military	Indicator applicant participated in the military	20	0.3
Children < 5	Number of children younger 5 years old	30	9.1
Children > 5	Number of children older than 5 years old	20	10.6
Total Score	Sum of the individual scores	-	224

Notes: This table describes the variables used to construct the score, along with their weight on the overall score constructed by the Ministry of Housing. Column (1) displays the name of the variable, column (2) describes it, column (3) shows how much impact it has on the overall score, and column (4) shows the average score considering all applicants.

Table A2: Neighborhood characteristics

Dep. Variable	Initial Location (1)	End Location (2)	Difference (3)	Applicants (4)
Panel A: Census Data				
Education Index	9.879 (0.011)	9.608 (0.009)	-0.271*** (0.012)	16,059
Socio - Territorial Index	0.499 (0.001)	0.487 (0.000)	-0.012*** (0.001)	16,059
Overcrowding Index	1.629 (0.002)	1.603 (0.001)	-0.026*** (0.002)	16,059
Material Quality Index	8.829 (0.001)	8.866 (0.001)	0.036*** (0.001)	16,059
Housing Deficit	0.101 (0.000)	0.083 (0.000)	-0.018*** (0.000)	16,059
Population Density	191.954 (0.833)	216.754 (0.908)	24.800*** (0.987)	16,059
People / Houses	3.363 (0.004)	3.375 (0.004)	0.012*** (0.004)	16,059
Panel B: Unidad Vecinal				
Green Area Surface pp	3.318 (0.026)	3.374 (0.024)	0.056* (0.029)	13,918
Education Facilities pp	0.579 (0.004)	0.507 (0.003)	-0.072*** (0.004)	13,903
Supermarkets pp	0.042 (0.001)	0.039 (0.001)	-0.003*** (0.001)	14,120
Bus Stops pp	1.543 (0.008)	1.256 (0.008)	-0.287*** (0.009)	14,037
Drugstores pp	0.149 (0.002)	0.111 (0.002)	-0.037*** (0.002)	14,232
Health Facilities pp	0.086 (0.001)	0.071 (0.001)	-0.015*** (0.002)	14,142
Panel C: SII data				
Total Units	1720.642 (10.389)	1782.734 (11.733)	62.092*** (14.546)	16,027
Price m2 const.	21.580 (0.096)	17.138 (0.058)	-4.443*** (0.100)	16,027
Size of Const.	41.958 (0.074)	40.059 (0.062)	-1.899*** (0.079)	16,027
House Valuation	683.770 (3.165)	562.433 (2.100)	-121.337*** (3.401)	16,027

Notes: This table presents the average neighborhood characteristic for awarded applicants who buy a house using the subsidy on a variety of outcomes. Columns (1) and (2) show the neighborhood characteristics of the initial and end locations, respectively, while column (3) shows the difference, and column (4) displays the number of users. Panel A shows census outcomes at the census tract level. Panel B shows other outcomes gathered from *Espacio Público* at the *unidad vecinal* level, which is more aggregated than the census tract. Panel C displays characteristics regarding prices using data from the Servicio de Impuestos Internos, also at the census tract level. We find that users of the high subsidy tend to move to slightly worse neighborhoods in terms of average years of schooling and socio-territorial index, and have access to fewer amenities. They also move to cheaper places, measured as the price per square meter.

Table A3: Optimal Bandwidths by Outcome

Dep. Variable	High	Medium	Low
Grades	55.14	59.16	35.32
Grades (sd)	56.32	71.64	43.56
Percentile	59.25	74.85	53.83
Ind. Percentile > 0.5	57.54	69.96	35.35
Av. Score	58.73	55.35	45.27
Score Verbal	60.26	55.27	47.50
Score Math	71.15	51.80	47.69
Ind. Progression	65.30	59.32	54.25
Ind. at School	77.34	64.69	39.88
Ind. Dropout	76.76	63.12	34.11
Absenteeism	75.81	70.47	41.81
Chronic Absenteeism	82.95	52.30	42.71
Ind. Repeated	81.94	53.66	54.61
Ind. Priority Student	50.38	37.38	44.43
Ind. Public School	93.04	38.86	57.37
Ind. Private	58.74	68.54	43.56
School Quality	75.50	78.99	48.51
Class Size	74.00	68.63	48.10
Ind. Change Comuna	74.98	67.14	49.94
Ind. Change School	72.07	72.44	57.08

Notes: This table shows the value of the IMSE-optimal bandwidth at application for the main variables, separated by subsidy tier. It includes year and age by gender fixed effects and clusters the standard errors at the applicant level.

Table A4: Differences in demographics for voucher Users

	All	Winners	Users	Non-Users	Difference
Dep. Variable	(1)	(2)	(3)	(4)	(5)
Applicant's Age	37.373 (7.803)	36.971 (7.281)	35.827 (6.912)	38.040 (7.453)	-2.213*** (0.086)
Applicant's Gender	0.924 (0.265)	0.922 (0.269)	0.925 (0.263)	0.918 (0.274)	0.007** (0.003)
Married = 1 (Applicant)	0.307 (0.461)	0.319 (0.466)	0.318 (0.466)	0.321 (0.467)	-0.003 (0.006)
Single-Parent	0.620 (0.485)	0.601 (0.490)	0.607 (0.488)	0.596 (0.491)	0.011* (0.006)
Family Members	3.130 (1.048)	3.835 (1.094)	3.825 (1.094)	3.843 (1.094)	-0.018 (0.013)
Children at School	1.443 (0.710)	1.705 (0.837)	1.686 (0.810)	1.723 (0.860)	-0.037*** (0.010)
Former Applications	1.191 (1.972)	1.221 (1.976)	1.031 (1.780)	1.398 (2.128)	-0.367*** (0.023)
Self-reported Income	11.831 (5.201)	11.923 (5.277)	11.394 (4.943)	12.417 (5.525)	-1.023*** (0.063)
Av. Years Schooling NH	9.995 (1.476)	9.976 (1.470)	9.804 (1.386)	10.136 (1.526)	-0.332*** (0.017)
Normalized score	-47.432 (106.789)	66.225 (67.321)	58.382 (61.227)	73.552 (71.784)	-15.171*** (0.798)
Bandwidth	All	All	All	All	All
Observations	90,362	28,126	13,584	14,542	28,126

Notes: This table displays the differences in demographic characteristics for users and non-users among voucher winners. Columns (1), (2), (3), and (4) display the mean and standard deviation for the whole sample, all voucher winners, and split by users and non-users, respectively. Column (5) presents the difference in these demographic characteristics for users vs non-users. We consider all voucher winners and users, without restricting the sample within a bandwidth. ***, **, * indicates significance at 1, 5, and 10%.

Table A5: Achievement outcomes: Bandwidth selection

	Mean	BW = 30	BW = 40	BW = 50	BW = 60	BW = 70	BW = 80	BW = 90	BW = 100
Dep. Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Grades	5.685	0.035*** (0.012)	0.039*** (0.011)	0.036*** (0.010)	0.035*** (0.009)	0.033*** (0.008)	0.031*** (0.008)	0.028*** (0.007)	0.027*** (0.007)
Grades (sd)	-0.151	0.061*** (0.019)	0.066*** (0.017)	0.061*** (0.015)	0.058*** (0.014)	0.055*** (0.013)	0.052*** (0.012)	0.048*** (0.011)	0.045*** (0.011)
Percentile	0.485	0.019*** (0.006)	0.020*** (0.005)	0.018*** (0.005)	0.017*** (0.004)	0.016*** (0.004)	0.015*** (0.004)	0.015*** (0.003)	0.014*** (0.003)
Ind. Percentile > 0.5	0.473	0.032*** (0.009)	0.031*** (0.008)	0.028*** (0.007)	0.026*** (0.006)	0.024*** (0.006)	0.023*** (0.006)	0.022*** (0.005)	0.021*** (0.005)
Av. Score	-0.241	0.062** (0.030)	0.065** (0.026)	0.059** (0.023)	0.057*** (0.021)	0.054*** (0.019)	0.050*** (0.018)	0.047*** (0.017)	0.045*** (0.016)
Score Math	-0.246	0.053* (0.030)	0.056** (0.026)	0.051** (0.023)	0.051** (0.021)	0.050*** (0.019)	0.047*** (0.018)	0.045*** (0.017)	0.044*** (0.016)
Score Verbal	-0.205	0.050* (0.030)	0.051** (0.026)	0.047** (0.023)	0.045** (0.021)	0.044** (0.020)	0.041** (0.018)	0.038** (0.017)	0.036** (0.017)
Ind. Progression	0.830	0.014* (0.008)	0.015** (0.007)	0.014** (0.006)	0.015*** (0.006)	0.015*** (0.005)	0.014*** (0.005)	0.013*** (0.005)	0.012*** (0.004)
Ind. at School	0.956	0.003 (0.003)	0.004 (0.003)	0.004* (0.002)	0.004* (0.002)	0.004** (0.002)	0.004** (0.002)	0.004** (0.002)	0.004** (0.002)
Ind. Dropout	0.010	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)
Ind. Repeated	0.044	-0.003 (0.002)	-0.003* (0.002)	-0.003** (0.002)	-0.003** (0.001)	-0.003** (0.001)	-0.003** (0.001)	-0.002** (0.001)	-0.002** (0.001)
Absenteeism	10.020	-0.145 (0.180)	-0.178 (0.156)	-0.163 (0.139)	-0.143 (0.126)	-0.120 (0.116)	-0.096 (0.108)	-0.090 (0.102)	-0.095 (0.097)
Chronic Absenteeism	0.362	-0.008 (0.008)	-0.010 (0.007)	-0.011* (0.006)	-0.011** (0.006)	-0.010** (0.005)	-0.009* (0.005)	-0.008* (0.004)	-0.008* (0.004)
Observations	-	227,707	291,395	348,282	399,831	449,220	491,659	528,324	560,293
Children	-	37,676	48,361	58,060	66,793	75,298	82,628	89,177	95,020
Bandwidth	-	30	40	50	60	70	80	90	100
Controls	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the differences in our achievement outcomes after application, presenting the ITT estimate of the regression discontinuity using [Equation 1.1](#). Column (1) shows the control mean, and columns (2)-(9) display the ITT coefficients using different bandwidths from 30 to 100. Results are very stable across bandwidths for all achievement outcomes, suggesting our results are not driven by the bandwidth selection. All estimates include year and gender by age fixed effects and clustered standard errors at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Table A6: Achievement outcomes by Age at application

	No School		Elementary		Middle		High	
	ITT	ATE	ITT	ATE	ITT	ATE	ITT	ATE
Dep. Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grades (sd)	0.060*** (0.020)	0.245*** (0.083)	0.056** (0.025)	0.290** (0.135)	0.056** (0.027)	0.291* (0.152)	0.078** (0.038)	0.534** (0.264)
Percentile	0.021*** (0.006)	0.084*** (0.024)	0.012* (0.007)	0.064* (0.039)	0.019** (0.008)	0.095** (0.044)	0.012 (0.012)	0.083 (0.081)
Av. Score	0.036 (0.042)	0.150 (0.195)	0.078** (0.033)	0.461** (0.203)	0.043 (0.036)	0.271 (0.226)	0.044 (0.054)	0.254 (0.317)
Ind. Progression	0.005 (0.006)	0.021 (0.025)	0.013 (0.010)	0.067 (0.054)	0.029** (0.014)	0.163** (0.076)	0.001 (0.020)	0.007 (0.138)
Absenteeism	-0.001 (0.002)	-0.004 (0.006)	-0.001 (0.002)	-0.006 (0.009)	-0.001 (0.002)	-0.006 (0.011)	0.001 (0.004)	0.006 (0.025)
Mean Grades (sd)	-0.099	-0.099	-0.172	-0.172	-0.184	-0.184	-0.179	-0.179
Mean Percentile	0.491	0.491	0.481	0.481	0.482	0.482	0.482	0.482
Mean Av. Score	-0.168	-0.168	-0.219	-0.219	-0.234	-0.234	-0.305	-0.305
Mean Ind. Progression	0.933	0.933	0.874	0.874	0.837	0.837	0.799	0.799
Mean Absenteeism	0.108	0.108	0.101	0.101	0.095	0.095	0.099	0.099
Bandwidth	Opt	Opt	Opt	Opt	Opt	Opt	Opt	Opt
Observations	153,676	153,676	127,405	127,405	72,516	72,516	21,216	21,216

Notes: This table shows the differences in our main achievement outcomes, separated by age at application. Columns (1), (3), (5) and (7) display the ITT estimates of winning the subsidy, while (2), (4), (6) and (8) show the LATE estimates of using the voucher to buy a house. All estimates include year and gender by age fixed effects and clustered standard errors at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Table A7: Achievement outcomes by Household size

	HH Size<3		HH Size=3		HH Size>3	
	ITT	ATE	ITT	ATE	ITT	ATE
Dep. Variable	(1)	(2)	(3)	(4)	(5)	(6)
Grades (sd)	0.034 (0.028)	0.123 (0.124)	0.043** (0.020)	0.226** (0.108)	0.082*** (0.029)	0.324*** (0.120)
Percentile	0.017** (0.008)	0.063* (0.038)	0.012** (0.006)	0.067** (0.031)	0.021** (0.008)	0.082** (0.034)
Av. Score	0.006 (0.051)	0.000 (0.246)	0.059** (0.029)	0.377** (0.193)	0.056 (0.040)	0.258 (0.184)
Ind. Progression	0.006 (0.010)	0.029 (0.045)	0.016** (0.007)	0.081** (0.038)	0.015 (0.011)	0.057 (0.045)
Absenteeism	0.330 (0.204)	1.635* (0.915)	-0.300* (0.162)	-1.531* (0.829)	0.107 (0.227)	0.511 (0.909)
Mean Grades (sd)	-0.051	-0.051	-0.168	-0.168	-0.207	-0.207
Mean Percentile	0.506	0.506	0.480	0.480	0.476	0.476
Mean Av. Score	-0.152	-0.152	-0.240	-0.240	-0.283	-0.283
Mean Ind. Progression	0.865	0.865	0.832	0.832	0.792	0.792
Mean Absenteeism	9.671	9.671	10.076	10.076	10.416	10.416
Bandwidth	Opt	Opt	Opt	Opt	Opt	Opt
Observations	80,418	80,418	182,496	182,496	112,291	112,291

Notes: This table shows the differences in our main achievement outcomes, split by household size at application. Columns (1), (3), (5) and (7) display the ITT estimates of winning the subsidy, while (2), (4), (6) and (8) show the LATE estimates of using the voucher to buy a house. All estimates include year and gender by age fixed effects and clustered standard errors at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Table A8: Achievement outcomes by Neighborhood quality

	Bad Neighborhood		Medium Neighborhood		Good Neighborhood	
	ITT	ATE	ITT	ATE	ITT	ATE
Dep. Variable	(1)	(2)	(3)	(4)	(5)	(6)
Grades (sd)	0.050** (0.024)	0.268** (0.125)	0.083*** (0.025)	0.358*** (0.114)	0.039 (0.025)	0.150 (0.123)
Percentile	0.012* (0.007)	0.060* (0.035)	0.025*** (0.007)	0.109*** (0.033)	0.015* (0.008)	0.061 (0.037)
Av. Score	0.040 (0.035)	0.294 (0.245)	0.114*** (0.036)	0.545*** (0.184)	0.012 (0.038)	0.039 (0.210)
Ind. Progression	0.011 (0.009)	0.062 (0.046)	0.027*** (0.009)	0.115*** (0.042)	0.006 (0.009)	0.026 (0.044)
Absenteeism	-0.128 (0.194)	-0.656 (0.932)	-0.155 (0.195)	-0.623 (0.871)	-0.012 (0.189)	0.131 (0.903)
Mean Grades (sd)	-0.206	-0.206	-0.168	-0.168	-0.099	-0.099
Mean Percentile	0.484	0.484	0.482	0.482	0.486	0.486
Mean Av. Score	-0.283	-0.283	-0.257	-0.257	-0.177	-0.177
Mean Ind. Progression	0.829	0.829	0.824	0.824	0.824	0.824
Mean Absenteeism	10.379	10.379	10.175	10.175	9.751	9.751
Bandwidth	Opt	Opt	Opt	Opt	Opt	Opt
Observations	126,812	126,812	128,350	128,350	120,043	120,043

Notes: This table shows the differences in our main achievement outcomes, split by neighborhood quality at application using average years of schooling. Columns (1), (3), (5) and (7) display the ITT estimates of winning the subsidy, while (2), (4), (6) and (8) show the LATE estimates of using the voucher to buy a house. All estimates include year and gender by age fixed effects and clustered standard errors at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Table A9: High School and Post-secondary outcomes by age

	Control Mean	ITT	ATE	Observations	Bandwidth
Panel A: Children < 13 at application					
Ind. graduated HS	0.881	0.010 (0.008)	0.040 (0.037)	17,038	71.600
Grades HS	5.688	0.013 (0.019)	0.041 (0.092)	11,892	58.917
Ind. PSU	0.827	0.018 (0.013)	0.080 (0.066)	13,176	59.357
Av. Score	-0.287	0.021 (0.029)	0.111 (0.145)	8,941	55.438
Ind. college	0.486	0.029** (0.013)	0.147** (0.064)	13,027	58.387
Years of college	1.040	0.086** (0.043)	0.392* (0.214)	13,528	61.843
Ind. University	0.445	0.005 (0.023)	0.025 (0.120)	6,964	66.233
Years Accredited	5.705	0.018 (0.063)	0.112 (0.322)	6,146	60.243
Panel B: Children > 13 at application					
Ind. graduated HS	0.823	0.020** (0.010)	0.080* (0.048)	15,147	73.359
Grades HS	5.515	0.072*** (0.021)	0.343*** (0.111)	8,535	56.280
Ind. PSU	0.829	-0.010 (0.013)	-0.044 (0.057)	13,932	86.335
Av. Score	-0.359	0.098*** (0.034)	0.448*** (0.172)	7,436	56.556
Ind. college	0.715	0.007 (0.012)	0.034 (0.057)	12,691	75.338
Years of college	2.532	0.178** (0.082)	0.785** (0.381)	12,787	76.136
Ind. University	0.356	0.055*** (0.021)	0.251** (0.103)	8,364	67.507
Years Accredited	5.525	-0.161** (0.073)	-0.758** (0.352)	6,618	55.645

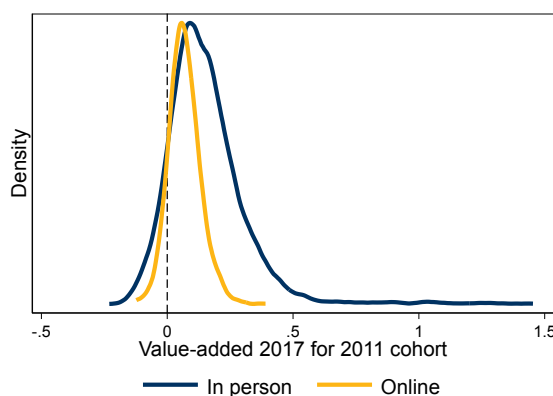
Notes: This table shows the differences in high school completion and our main post-secondary education outcomes, split by age at application. Panel A considers children who were below 13 years old when applying, while Panel B considers children aged 13 or older. Columns (1), (4), and (5) display the mean of the variable for non-awarded households, the number of applicants, and the bandwidth selected to perform the RD, respectively. Column (2) shows the ITT estimate of the regression discontinuity using Equation 1.1 by regressing each demographic characteristic on the score, corresponding to the jump at the cutoff. Column (3) presents the LATE estimates, rescaling the ITT estimates by the first stage in Equation 3.6 and presenting the effects of using the voucher to buy a house. All estimates include year and gender by age fixed effects and clustered standard errors at the application level. ***, **, * indicates significance at 1, 5, and 10%.

Appendix B

Appendix of The Effects of Widespread Online Education on Market Structure and Enrollment

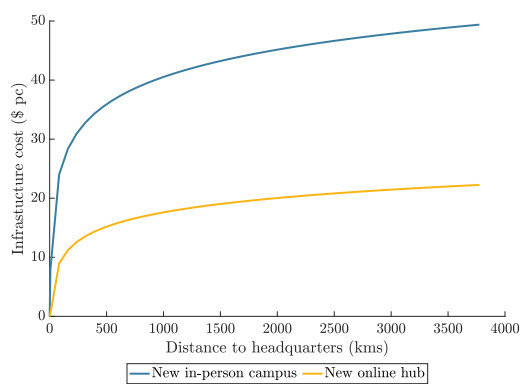
B.1 Additional Figures and Tables

Figure B1: Value-added distribution for in-person and online degree programs



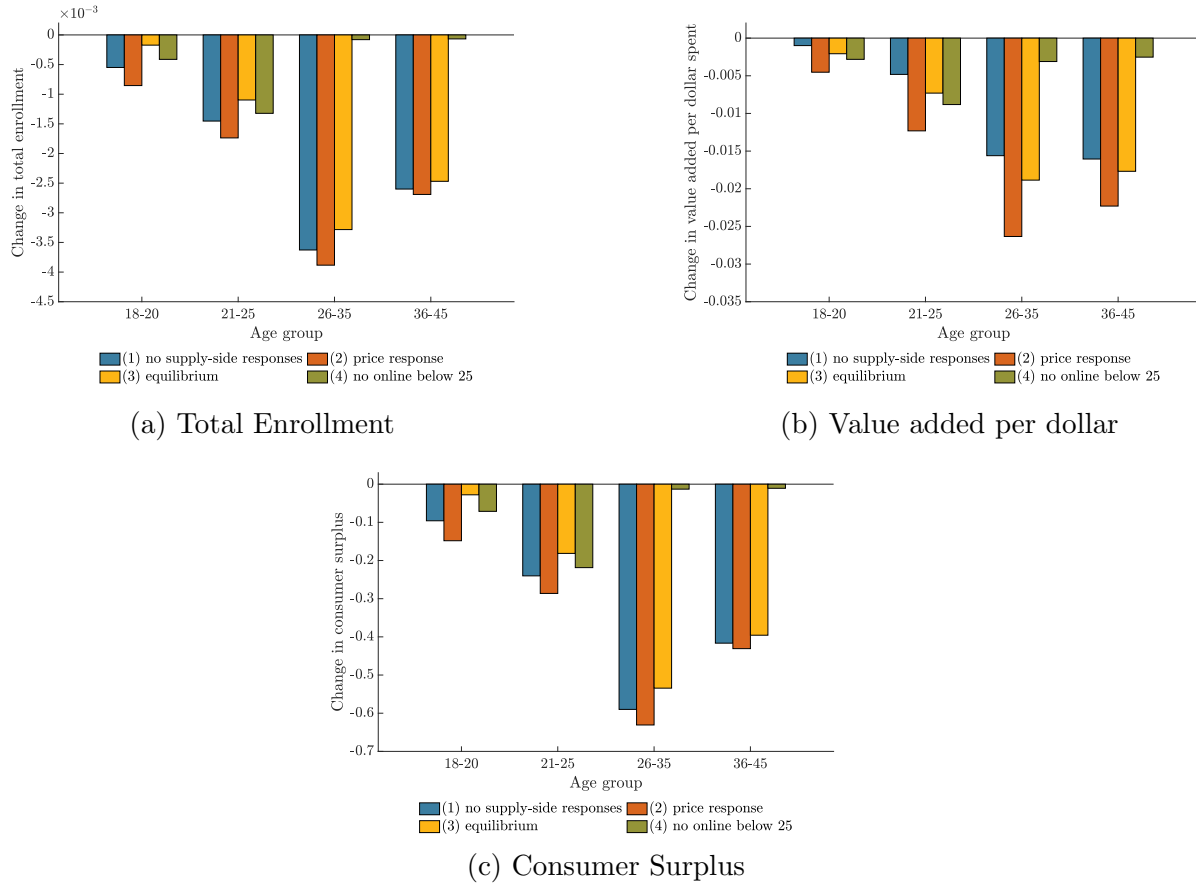
Notes: This figure shows the distribution of value added for existing degree programs. The blue lines represent the distribution of in-person degree programs, and the yellow line represents the distribution of online degree programs.

Figure B2: Cost of building infrastructure as a function of distance



Notes: This figure shows the infrastructure cost of building a new campus or hub as a function of the distance between the firm's headquarter and any given region, given by $\chi_0^k + g(d_{fr})' \chi^k$ for $k \in \{\text{campus, hub}\}$.

Figure B3: Changes in various outcomes by age group under each counterfactual



Notes: this figure presents changes in various outcomes for different age groups under different counterfactuals. We show changes in value added relative to the *Baseline* counterfactual for all counterfactuals presented in Table 2.5 and the hypothetical counterfactual from Section 2.5. In Panel (a), we show the changes in total enrollment. In Panel (b), we show changes in value added for every \$1000 spent in higher education. In Panel (c), we show changes in total consumer surplus measured as the area below the demand curve.

Table B1: Areas of Study and Degrees

Area of Study	Type	Degree programs
Arts, Humanities,	Bachelor	Library, Political Science, Social Communication, Design, Journalism, International Relations
Social Sciences	Technical	Interior Design, Fashion Design, Product Design, Graphic Design, Photography, Digital Marketing, Audiovisual Production, Multimedia Production
Business	Bachelor	Administration, Accounting, Economics, Public Administration, Social Communication, Advertisement and Marketing, Public Relations, Executive Secretariat
	Technical	Foreign Trade, Entrepreneurship, Commercial Management, Quality Management, Information Technology Management, Human Resources Management, Security Management, Public Health Management, Legal Management, Information System Management, Financial Management, Hospital Management, Public Management, Logistics, Marketing, Real Estate, Management Processes, Secretariat
Education	Bachelor	Visual Arts, Biological Sciences, Physical Education, Philosophy, Geography, History, Mathematics, Psychopedagogy, Chemistry
	Technical	Visual Arts, Biological Sciences, Social Sciences, Physical Education, Geography, History, Letras, Mathematics, Pedagogy, Chemistry, Sociology
Engineering	Bachelor	Agronomy, Architecture and Urbanism, Civil Engineering, Computer Engineering, Software Engineering, Electrical Engineering, Mechanical Engineering, Chemical Engineering
	Technical	Agribusiness, Agribusiness Management, Industrial Mechatronics, Mechanical Manufacturing
Health Sciences	Bachelor	Nursing, Pharmacy, Physiotherapy, Veterinary Medicine, Speech Therapy, Nutrition Occupational Therapy
	Technical	Radiology
Services	Bachelor	Aeronautical Sciences, Social Service, Tourism
	Technical	Beauty and Personal Image, Aesthetics and Cosmetics, Gastronomy, Environmental Management, Tourism Management, Public Security, Juridic Services
Others	Bachelor	Environmental Engineering, Production Engineering, Theology
	Technical	Industrial Automation, Social Educator, Management of Industrial Production
Law	Bachelor	Law
Medicine	Bachelor	Medicine, Odontology
Psychology	Bachelor	Psychology
Math, Computer, Natural Sciences	Bachelor	Biomedicine, Computer Science, Information Systems
	Technical	Analysis and System Development, <i>Banco de Dados, Jogos Digitais, Redes de Computadores</i> , Information Security, <i>Sistemas para Internet</i>

Notes: This table summarizes the aggregation of different degree programs into broader areas of study. We categorize programs based on the International Standard Classification of Education (ISCED) codes.

Table B2: Relationship between distance and probability of opening an online degree for selected years

	entered _{<i>f</i><i>r</i>} (1)
$\log(1 + d_{fr})$	-0.088 (0.008)
H_{fr}	-0.146 (0.076)
Obs.	4070
Regions	110
Firms	37
Mean Dep. Var	0.22

Notes: This table shows the estimated parameters from estimating Equation (2.5), given by $\text{Entered}_{fr} = g(d_{fr})'\gamma + \delta_f + \delta_r + \eta_{fr}$, where Entered_{fr} equals 1 if institution f had entered region r by 2010 and 0 otherwise; d_{fr} represents the distance between the headquarters of institution f and region r . We define the vector $g(d_{fr}) = [\log(1 + d_{fr}), H_{fr}]'$, where $H_{fr} \in \{0, 1\}$ indicates whether the headquarters of institution f are located in region r (i.e., $d_{fr} = 0$), to account for cases where the distance is zero. We estimate the regression using all 37 firms that offered at least one online degree in 2010.

Table B3: Demand parameters

Panel A: BLP parameters									
Price coef.		Price RC		Online RC		In-person RC		Nest param.	
$\bar{\alpha}$	-2.214 (1.111)	σ_{α}	-0.002 (40.041)	σ^o	-0.112 (60.217)	σ^p	0.087 (73.554)	ρ	0.852 (0.248)
Dem. shifter		Dem. shifter		Age 21-25		Age 26-35		Age 36-45	
ψ^0	0.099 (0.148)	ψ^o	-0.04 (0.785)	μ_2^p	-0.906 (15.233)	μ_3^p	-1.465 (14.835)	μ_4^p	-2.095 (25.36)
Age 21-25		Age 26-35		Age 36-45					
μ_2^o	-0.63 (5.577)	μ_3^o	-0.856 (14.237)	μ_4^o	-1.263 (1.725)				
Panel B: Mixed model parameters									
Degree r.e.		Region-area r.e.		Year-area r.e.		Year-online r.e.		Demand sock	
σ_j	0.237 (0.004)	σ_{ra}	0.663 (0.014)	σ_{ta}	0.855 (0.063)	σ_{to}	0.467 (0.08)	σ_{ξ}	0.316 (0.001)
Hours f.e.		Stem f.e.		Degree age f.e.		Av. score f.e.		Av. wages f.e.	
δ_1	0.119 (0.007)	δ_2	0.019 (0.023)	δ_3	0.044 (0.007)	δ_4	0.006 (0.0002)	δ_5	0.049 (0.022)
Const. f.e.									
δ_0	-9.249 (0.208)								

Notes: This table reports the estimated parameters from the demand model presented in Section 2.4. Panel A presents the parameters estimated in the first step following [Berry et al. \(1995\)](#). Panel B presents the parameters estimated in the second step from the mixed-effects Bayesian hierarchical model that we use to recover each of the mean utility components of δ_{jrt} .

Table B4: Diversion ratios by age group

	$t = 2010$		$t = 2019$	
	In-person	Online	In-person	Online
Median diversion ratios for cohort 18-20:				
To in-person:	0.84	0.80	0.81	0.54
To online:	0.00	0.06	0.04	0.33
To outside good:	0.14	0.14	0.14	0.13
Median diversion ratios for cohort 21-25:				
To in-person:	0.80	0.61	0.69	0.25
To online:	0.01	0.23	0.14	0.61
To outside good:	0.15	0.15	0.14	0.14
Median diversion ratios for cohort 26-35:				
To in-person:	0.73	0.30	0.41	0.06
To online:	0.06	0.50	0.40	0.78
To outside good:	0.15	0.15	0.14	0.15
Median diversion ratios for cohort 36-45:				
To in-person:	0.61	0.12	0.22	0.02
To online:	0.18	0.68	0.61	0.82
To outside good:	0.15	0.16	0.15	0.15

Notes: This table reports diversion ratios across products and markets for 2010 and 2019 for in-person and online degrees for each age group. We calculate diversion ratios as the share of students from each age group that decide to stop attending degree j upon an increase in tuition price that would switch to either an in-person degree, an online degree, or the outside option. Formally, we calculate diversion ratios as $D_{jb\mathcal{K}} = \left(\left| \frac{\partial s_{jb}}{\partial p_j} \right| \right)^{-1} \left(\sum_{k \in \mathcal{K}/j} \frac{\partial s_{kb}}{\partial p_j} \right)$, where $\mathcal{K}$ the set of all degrees that are either in-person, online, or the outside option, and s_{jb} is the share of students among those in group age b that decide to attend degree j . For example, the diversion ratio from in-person degrees to online degrees in 2010 for age group 36-45 is 0.2, which is given by the median of $D_{jb\mathcal{K}}$ across all degrees j that are in-person and for $\mathcal{K}$ the set of all online degrees, with $b = \{36-45\}$.

Table B5: Marginal cost parameters

Panel A: Main parameters									
Demand instrument		log-distance		same-region					
γ_{z_1}	0.152 (0.399)	γ_{z_2}	0.151 (1.04)	γ_{z_3}	0.748 (1.961)				
Panel B: Mixed model parameters									
Degree r.e.		Region-area r.e.		Year-area r.e.		Year-online r.e.		Cost sock	
σ_j	0.539 (0.008)	σ_{ra}	0.267 (0.009)	σ_{ta}	0.101 (0.033)	σ_{to}	0.989 (0.16)	σ_ω	0.56 (0.001)
Hours f.e.		Stem f.e.		Degree age f.e.		Av. score f.e.		Av. wages f.e.	
δ_1	0.187 (0.012)	δ_2	-0.12 (0.038)	δ_3	-0.014 (0.015)	δ_4	0.007 (0.0004)	δ_5	0.257 (0.045)
Const. f.e.									
δ_0	-7.756 (0.388)								

Notes: This table reports the estimated parameters from the marginal cost model presented in Section 2.4. Panel A presents the parameters estimated in the first step via OLS. Panel B presents the parameters estimated in the second step from the mixed-effects Bayesian hierarchical model that we use to recover each of the marginal cost components of γ_{jrt} .

Table B6: Fixed cost parameters

Panel A: Infrastructure costs											
New campus						New hub					
χ_0^c	-0.795 (1.093)	χ_H^c	0.795 (1.093)	χ_d^c	0.953 (0.165)	χ_0^h	-0.94 (0.917)	χ_H^h	0.94 (0.916)	χ_d^h	0.5 (0.125)
Panel B: Degree costs											
κ_0	1.057 (0.15)	$\kappa_{\iota 1}^{old}$	0 (0)	$\kappa_{o 1}^{old}$	0.17 (0.334)	$\kappa_{\iota 2}^{old}$	0 (0)	$\kappa_{o 2}^{old}$	0 (0)	$\kappa_{\iota 3}^{old}$	0 (0)
$\kappa_{o 3}^{old}$	0 (0)	$\kappa_{\iota 4}^{old}$	0 (0)	$\kappa_{o 4}^{old}$	0 (0)	$\kappa_{\iota 5}^{old}$	0 (0)	$\kappa_{o 5}^{old}$	0 (0)	$\kappa_{\iota 6}^{old}$	0 (0)
$\kappa_{\iota 7}^{old}$	0 (0)	$\kappa_{o 7}^{old}$	0 (0)	$\kappa_{\iota 8}^{old}$	0 (0)	$\kappa_{\iota 9}^{old}$	0 (0)	$\kappa_{o 9}^{old}$	0 (0)	$\kappa_{\iota 10}^{old}$	0 (0)
$\kappa_{\iota 11}^{old}$	0 (0)	$\kappa_{o 11}^{old}$	0 (0)	$\kappa_{\iota 1}^{new}$	0.852 (0.201)	$\kappa_{o 1}^{new}$	0 (0)	$\kappa_{\iota 2}^{new}$	1.039 (0.295)	$\kappa_{o 2}^{new}$	0 (0)
$\kappa_{\iota 3}^{new}$	0.874 (0.213)	$\kappa_{o 3}^{new}$	0 (0)	$\kappa_{\iota 4}^{new}$	0 (0)	$\kappa_{o 4}^{new}$	0 (0)	$\kappa_{\iota 5}^{new}$	0 (0.043)	$\kappa_{o 5}^{new}$	0 (0.06)
$\kappa_{\iota 6}^{new}$	0.345 (0.207)	$\kappa_{\iota 7}^{new}$	0.515 (0.137)	$\kappa_{o 7}^{new}$	0 (0)	$\kappa_{\iota 8}^{new}$	0 (0)	$\kappa_{\iota 9}^{new}$	0.049 (0.093)	$\kappa_{o 9}^{new}$	0 (0.014)
$\kappa_{\iota 10}^{new}$	0 (0)	$\kappa_{\iota 11}^{new}$	0 (0.064)	$\kappa_{o 11}^{new}$	0 (0)	σ_ε	0.007 (0.001)				

Notes: This table reports the estimated fixed-cost parameters from the entry model presented in Section 2.4. Panel A presents the infrastructure costs parameters from Equation (2.16). Panel B presents the degrees fixed costs parameters from Equation (2.15). Standard errors are estimated via Bootstrap with 100 repetitions and clustering at the market level.

B.2 Data Construction

In this Appendix, we provide details on the construction process of some key variables.

Tuition fees

To construct program-level prices, we combine four different data sources. The first two sources come from Brazil’s government fellowship and loan programs, PROUNI and FIES. We utilize administrative records from the National Education Fund (FNDE), which track the payments made by the government to students in these programs, allowing us to estimate the tuition fees at participating institutions. The third source is a nationally representative survey conducted by Hoper, a consultancy specializing in higher education. The fourth source is administrative data from QueroBolsa, Brazil’s largest degree search platform.

Using these datasets, we construct prices as follows: First, we use the information from all sources to recover an average posted price by program, campus or hub, and year. This is done by regressing log-prices on program-region-year and source fixed effects. Controlling for source fixed effects helps account for persistent differences across data sources and allows us to recover a program-region-year price. In cases where information for a certain year is missing, we run a regression for the predicted price on program-region and year fixed effects, imputing the missing values based on the sum of the coefficients. Finally, in situations where information on both year and region is unavailable, we regress the predicted price on program and year fixed effects, filling the missing values using the sum of these coefficients.

Value added

We track all university entrance exam (ENEM) participants and assign them to degree programs based on their initial college enrollment, or to an outside option if they do not enroll. We then follow these individuals in the labor market seven years after the exam, recording their highest salary from labor market administrative records (RAIS).

Our value-added estimation relies on a standard selections-on-observables model (Rothstein, 2010; Angrist et al., 2017), which compares the labor market earnings of students across various programs while controlling for test scores and an extensive range of student characteristics. All values are normalized relative to the outside option of not attending college. We allow the outside option to vary by region to account for local labor market conditions.

Specifically, we estimate the following model for each region r :

$$\log(Y_i) = \sum_j \text{VA}_{jr} \cdot D_{ij} + X_i' \beta_r + \varepsilon_i \quad (\text{B.1})$$

where Y_i represents the wage income of student i , $D_{ij} \in \{0, 1\}$ indicates whether student i is enrolled in degree program j , X_i includes students’ characteristics such as gender, age, ENEM

score, and a constant, and ε_i captures other determinants of income that are uncorrelated with school enrollment. The term VA_{jr} denotes the region-specific value added of each program. We normalize the value added of the outside option to zero in each region and estimate the model via OLS.

We follow Angrist et al. (2023) and use empirical Bayes shrinkage methods to improve the precision of estimates. The estimates $\widehat{VA}_{jr}$ from estimating Equation (B.1) via OLS are unbiased but noisy measures of the underlying program-specific value added. We investigate the distribution of VA_{jr} using the following hierarchical model:

$$\begin{aligned}\widehat{VA}_{jr}|VA_{jr} &\sim \mathcal{N}(VA_{jr}, s_{jr}^2) \\ VA_{jr} &\sim \mathcal{N}(\mu_k, \sigma_k^2)\end{aligned}$$

where ω_{jr}^2 is the sampling variance of the estimator $\widehat{VA}_{jr}$, while μ_k and σ_k^2 are the hyperparameters that govern the distribution of the latent parameters across programs that we allow to be different for in-person and online degrees (i.e., $k \in \{\text{in-person}, \text{online}\}$). Method of moments estimates of these hyperparameters are given by:

$$\begin{aligned}\hat{\mu}_k &= \frac{1}{J_k} \sum_{j \in k} \widehat{VA}_{jr} \\ \hat{\sigma}_k^2 &= \frac{1}{J_k} \sum_{j \in k} \left[\left(\widehat{VA}_{jr} - \hat{\mu}_{VA} \right)^2 - s_{jr}^2 \right]\end{aligned}$$

where J_k is the number of programs offered in format k .

The final step in the empirical Bayes estimation is to construct posteriors for the value added of each program. Given the model above, the posterior means are given by

$$VA_{jr}^* = \left(\frac{\sigma_{k(j)}^2}{\sigma_{k(j)}^2 + s_j^2} \right) \widehat{VA}_{jr} + \left(\frac{s_j^2}{\sigma_{k(j)}^2 + s_j^2} \right) \hat{\mu}_{k(j)}. \quad (\text{B.2})$$

By shrinking the noisy estimate of VA_{jr} toward the prior mean, the posterior mean reduces variance, with more shrinkage for programs with noisier estimates. We use the posterior means in the analysis from Section 2.2.

To estimate the model, we use the universe of ENEM 2010 applicants from the 2011 admissions cycle and use student labor market earnings in 2017, seven years after their enrollment decisions. Because we use only one student cohort, our program value-added measure remains constant over time. We are able to estimate value added for 22% of the program-region pairs that exist in the data, which covers 73% of enrollment across all years. Among program-region pairs that existed in 2011, we cover 86% of program-region pairs and 98% of total enrollment.

Degree-level aggregation and inputation for counterfactual analysis

The counterfactual analysis from Section 2.5 requires value-added estimates for all degrees that can potentially enter a market. To get reliable estimates for degrees that we do not observe in the data, we estimate a simple model using the program-level value-added measures presented above. In our model, each degree-region pair is associated with a specific value added given by

$$VA_{jr} = x_j^{(1)} \nu_o + x_j^{(2)} \nu_x + \nu_j + \nu_{ra} + \varsigma_{jr} \quad (\text{B.3})$$

where VA_{jr} is the (weighted) average of the posterior mean of the estimated value-added of all degree programs that compose degree j , $x_j^{(1)} = [\iota_j, o_j]$ and $x_j^{(2)}$ are the degree-level characteristics from Equations (2.8) and (2.9) from the main text, ν_j are degree specific components of value added, ν_{ra} are field of study-region specific components, and ς_{jr} is a degree-region specific idiosyncratic shock. As before, the outside option is normalized to be zero in each region.

Following Section 2.4, we estimate the value added model using mixed-effects Bayesian hierarchical model with $\nu_j \sim \mathcal{N}(0, \vartheta_j^2)$, $\nu_{ra} \sim \mathcal{N}(0, \vartheta_{ra}^2)$, and $\varsigma_{jr} \sim \mathcal{N}(0, \vartheta_{\varsigma}^2)$, and (ν_o, ν_x) non-random parameters. We estimate the model via maximum likelihood to recover posterior means of each component of the value-added model that we use to impute value-added for all degree-markets, regardless of whether they are offered in the data or not.

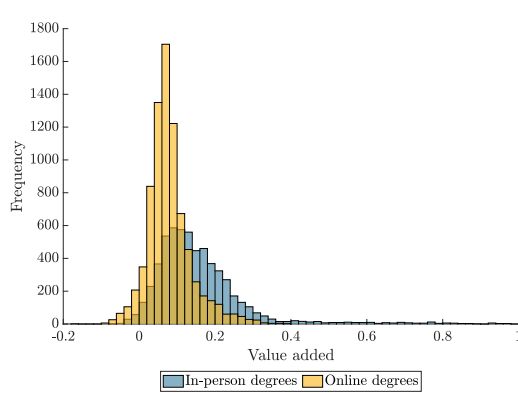
We present the estimated parameters in Table B7. In Figure B4, Panel (a), we show the distribution of value added for online and in-person degrees. We find that in-person and online degrees have an average value added of 0.157 and 0.079, respectively. Finally, in Figure B4, Panel (b), we show that students' preferences over degrees have a positive but weak correlation with value added. The correlation is particularly low for online degrees, which tend to be newer, and students might be less informed about their quality.

Table B7: Value added parameters

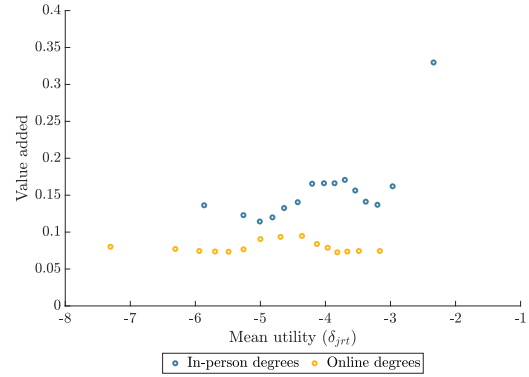
Panel A: Mixed model parameters									
Degree r.e.		Region-area r.e.		VA shock		Const. f.e.		Online f.e.	
σ_j	0.057 (0.002)	σ_{ra}	0.068 (0.003)	σ_{ta}	0.068 (0.001)	δ_0	-1.237 (0.056)	δ_o	-0.041 (0.007)
Hours f.e.		Stem f.e.		Degree age f.e.		Av. score f.e.		Av. wages f.e.	
δ_1	0.066 (0.003)	δ_2	-0.026 (0.007)	δ_3	-0.015 (0.004)	δ_4	0.001 (0.0001)	δ_5	0.103 (0.008)

Notes: This table reports the estimated fixed-cost parameters from the model presented in Section B.2.

Figure B4: Estimated parameters from value added model



(a) Distribution of value added



(b) Relationship between preferences and value added

Notes: The figure summarizes the estimation results of the value-added model. Panel (a) shows the distribution of value added across all in-person and online degrees that are offered in the data. Panel (b) shows the relationship between the degrees' mean utility components of δ_{jrt} and the estimated value added. In both panels, in-person degrees are depicted in blue and online degrees in yellow.

B.3 Empirical Model Details

Internet penetration and degree entry

In Section 2.4, we use internet penetration as an excluded variable to estimate the entry fixed cost. This appendix presents the reduced-form relationship between internet penetration and changes in enrollment and degree availability for in-person and online education. Specifically, we estimate the following equation:

$$\Delta y_{ra} = \phi \Delta I_r + \varepsilon_{ra} \quad (\text{B.4})$$

where ΔI_r denotes the change in the number of internet accesses per person in region r between 2010 and 2019, and Δy_{ra} represents the corresponding change in the following outcomes: (i) the total number of online degrees, (ii) the total number of in-person degrees, (iii) the share of online students relative to market size, and (iv) the share of in-person students relative to market size. We estimate the model using OLS, with results summarized in Table B8. Our findings indicate that increased internet penetration predicts a rise in the number of online degrees and a small, statistically insignificant decline in the number of in-person degrees. Additionally, internet penetration is associated with increased online enrollment and a decrease in in-person enrollment.

Table B8: Effects of internet expansion on degree availability and enrollment

	Δ in online degrees (1)	Δ in in-person degrees (2)	Δ in online students (3)	Δ in in-person students (4)
Panel A: OLS regression				
Δ internet penetration	3.383 (0.800)	-0.221 (0.323)	1.705 (0.607)	-1.094 (0.246)
Panel B: Average value of the dependant variable in levels in 2010 and 2019				
2010	2.82	10.98	0.72	3.10
2019	9.48	12.19	2.79	2.94
Obs.	1,210	1,210	1,210	1,210

Notes:

Entry probabilities prediction model

We estimate the entry fixed costs from Section 2.4 using the two-step estimator developed in Sweeting (2009). In the first step, we estimate a multinomial choice logit model using all own and competitors' drivers of variable profits, and the determinants of fixed costs from Equation (2.14), to predict the probability that each bundle is offered in the data. To do that,

we start by computing proxies of a bundle's value by estimating three different measures. The first measure proxies for the expected utility that a consumer perceives when facing the opportunity of choosing a given degree j that belongs to bundle $\mathcal{J}_{f_{rt}}$, and is given by

$$\text{log-sum}(\vec{\delta}_{\mathcal{J}_{f_{rt}}}) = \log \left(1 + \sum_{j \in \mathcal{J}_{f_{rt}}} \exp(x_j^{(2)}\delta + \delta_j + \delta_{ra} + \delta_{ta} + \delta_{to_j}) \right)$$

where the terms inside the exponential are the posterior means of the mean utility components of $\delta_{j_{rt}}$ from Equation (2.9).

The second measure proxies for the expected utility that a consumer perceives when facing the opportunity of choosing a given degree j that belongs to bundle $\mathcal{J}_{f_{rt}}$ but taking into account the marginal cost of each associated degree, and is given by

$$\text{log-sum}(\vec{\delta}_{\mathcal{J}_{f_{rt}}}, \vec{c}_{\mathcal{J}_{f_{rt}}}) = \log \left(1 + \sum_{j \in \mathcal{J}_{f_{rt}}} \exp(x_j^{(2)}\delta + \delta_j + \delta_{ra} + \delta_{ta} + \delta_{to_j} - \bar{\alpha}c_{j_{rt}}|_{\omega_{j_{rt}}=0}) \right)$$

where $c_{j_{rt}}|_{\omega_{j_{rt}}=0}$ corresponds to the marginal cost from Equation (2.12) evaluated at $\omega_{j_{rt}} = 0$, and $\bar{\alpha}$ is the average price sensitivity across all students.

The third measure proxies for the total market share that firm f would capture offering bundle $\mathcal{J}_{f_{rt}}$ and is given by

$$\hat{s}(\vec{\delta}_{\mathcal{J}_{f_{rt}}}, \vec{c}_{\mathcal{J}_{f_{rt}}}) = \frac{\sum_{j \in \mathcal{J}_{f_{rt}}} \exp(x_j^{(2)}\delta + \delta_j + \delta_{ra} + \delta_{ta} + \delta_{to_j} - \bar{\alpha}c_{j_{rt}}|_{\omega_{j_{rt}}=0})}{1 + \sum_{\mathcal{J}} \sum_{k \in \mathcal{J}} \exp(x_k^{(2)}\delta + \delta_k + \delta_{ra} + \delta_{ta} + \delta_{to_k} - \bar{\alpha}c_{k_{rt}}|_{\omega_{k_{rt}}=0})}$$

where the numerator sums over all degrees belonging to bundle $\mathcal{J}_{f_{rt}}$, and the denominator adds over all degrees available in market rt .

We use $\Theta_{\mathcal{J}_{f_{rt}}} = [\text{log-sum}(\vec{\delta}_{\mathcal{J}_{f_{rt}}}), \text{log-sum}(\vec{\delta}_{\mathcal{J}_{f_{rt}}}, \vec{c}_{\mathcal{J}_{f_{rt}}}), \hat{s}(\vec{\delta}_{\mathcal{J}_{f_{rt}}}, \vec{c}_{\mathcal{J}_{f_{rt}}})]$, the linear combination of $\Theta_{\mathcal{J}_{f_{rt}}}$, to approximate the variable profits that firm f would receive after entering with bundle $\mathcal{J}_{f_{rt}}$, and a linear combination of the fixed cost determinants from Equation (2.14) to approximate for the fixed costs associated with offering bundle $\mathcal{J}_{f_{rt}}$. Specifically, we approximate the probability that bundle $\mathcal{J}_{f_{rt}}$ is offered by

$$\phi_{\mathcal{J}_{f_{rt}}} = \frac{\exp(\Theta'_{\mathcal{J}_{f_{rt}}}\beta + \text{Degrees}(\mathcal{J}_{f_{rt}}|\mathcal{J}_{f_{rt}0}) + \text{Infrastructure}(\mathcal{J}_{f_{rt}}|\mathcal{J}_{f_{rt}0}))}{\sum_{\mathcal{J} \in \mathcal{B}_f} \exp(\Theta'_{\mathcal{J}}\beta + \text{Degrees}(\mathcal{J}|\mathcal{J}_{f_{rt}0}) + \text{Infrastructure}(\mathcal{J}|\mathcal{J}_{f_{rt}0}))}. \quad (\text{B.5})$$

where

$$\Theta'_{\mathcal{J}}\beta = \beta_1 \text{log-sum}(\vec{\delta}_{\mathcal{J}}) + \beta_2 \text{log-sum}(\vec{\delta}_{\mathcal{J}}, \vec{c}_{\mathcal{J}}) + \beta_3 \hat{s}(\vec{\delta}_{\mathcal{J}}, \vec{c}_{\mathcal{J}}).$$

We estimate Equation (B.5) via maximum likelihood and use the estimated probabilities, $\hat{\phi}_{\mathcal{J}_{f_{rt}}}$, to estimate the expected variable profits, $\mathbb{E}[\pi_{f_{rt}}(\mathcal{J})]$, as described in the main text.

Counterfactual Simulation Details

To implement our counterfactual analysis in Section 2.5, we simulate equilibrium outcomes using our estimated model. We begin with a dataset that includes all firms and the respective bundles $\mathcal{J}_{frt} \in \mathcal{B}_f$ they can offer in each market, together with the estimated profits of offering each bundle, $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]$, as estimated from the data, and the estimated entry probabilities, $\phi_{\mathcal{J}_{frt}}$. For each market, we proceed as follows:

1. Take the vector of expected profits from iteration $t - 1$ and call it $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_{t-1}$. For the first iteration, use the expected profits estimated from the data.
 - a) Take draws of the private-information fixed cost shocks, $\varepsilon_{fr\mathcal{J}_{frt}}$, for all potential bundles, $\mathcal{J}_{frt}$.
 - b) Solve the firms' problem from Equation (2.17) using the fixed cost draws and $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_{t-1}$ to determine which degrees are offered in the market.
 - c) Take draws from demand and marginal cost shocks, ξ_{jrt} and ω_{jrt} , solve for optimal pricing and compute variable profits using Equation (2.11) for all firms.
 - d) Repeat steps (a)-(c) 10,000 times and use the simulated variable profits to estimate $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]$. Call the estimate $\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J}_{frt})]$.
2. Update the vector of expected profits to $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_t = \sqrt{\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_{t-1} \hat{\mathbb{E}}[\pi_{frt}(\mathcal{J}_{frt})]}$ and calculate the mean square difference between iteration t and $t - 1$ as $MSD_t = \sum (\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_t - \mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_{t-1})^2$, where the sum is over all markets, firms, and bundles.
3. Iterate until $MSD_t < \overline{MSD}$ for $\overline{MSD}$ small.

Outcomes of Interest

In this appendix, we present detailed formulas to calculate the main outcomes of interest presented in Section 2.5 of the main text. For each counterfactual k , we calculate

$$\begin{aligned}
\text{Total value-added}^k &= \frac{\sum_r M_r \sum_j s_{jr}^k VA_{jr}}{\sum_r M_r \sum_j s_{jr}^{BL}} \\
\text{Total expenditure}^k &= \frac{\sum_r M_r \sum_j s_{jr}^k p_{jr}^k}{\sum_r M_r \sum_j s_{jr}^{BL}} \\
\text{Av. price of in-person degrees}^k &= \frac{\sum_r M_r \frac{\sum_{j:\iota_j=1} p_{jr}^k}{\sum_{j:\iota_j=1} 1}}{\sum_r M_r} \\
\text{Total profits per capita}^k &= \frac{\sum_r M_r \sum_j s_{jr}^k (p_{jr}^k - c_{jr})}{\sum_r M_r \sum_j s_{jr}^{BL}} \\
\text{Total consumer welfare}^k &= \frac{\sum_r M_r CS_r^k}{\sum_r M_r},
\end{aligned}$$

where VA_{jr} is the value added of degree j in region r , p_{jr}^k is the price of degree j in region r under counterfactual k , c_{jr} is the marginal cost of degree j in region r , M_r is the market size for region r , s_{jr}^k is the market share of degree j in region r under counterfactual k , and s_{jr}^{BL} is the market share of degree j in region r under the baseline counterfactual. Note that the denominator is constant across counterfactuals and we use it to normalize the values per student base.

Appendix C

Appendix of The Role of Housing Supply in Shaping Affordability: A Quasi-experimental Approach

C.1 Additional Figures and Tables

Table C1: OLS results for housing supply

	(1) lnPrice	(2) lnPrice	(3) lnPrice	(4) lnPrice	(5) lnPrice	(6) lnPrice
ln(Stock)	0.255*** (0.00383)	0.275*** (0.00615)	0.180*** (0.0121)			
ln(Transactions)				0.103*** (0.0157)		
ln(NP Transactions)					0.0802*** (0.00580)	
ln(RtO Transactions)						0.0531** (0.0185)
Implied elasticity ($\frac{1}{\beta}$)	3.921	3.636	5.556	9.701	12.469	18.832
Observations	15,857	15,857	15,857	15,858	15,858	15,858
R-squared	0.189	0.207	0.661	0.631	0.641	0.626
Fixed Effects	No	No	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Each cell corresponds to the estimate from running an OLS regression of housing stock or transactions on housing prices (both in logarithms). The regressions include controls at the neighborhood level, such as the share of different property types, the average construction year of properties, and total population. Year and neighborhood fixed effects are also included. “NP transactions” refers to new builds, while “RtO transactions” refers to the conversion of rental units to owner-occupied housing. Standard errors are clustered at the neighborhood level. Additionally, implied housing supply elasticities are calculated.

Table C2: OLS results for other measures of prices

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Effects on Price per sq ft (in log)						
ln(Stock)	0.229*** (0.00284)	0.148*** (0.00444)	0.0736*** (0.00702)			
ln(Transactions)				0.0477*** (0.00801)		
ln(NP Transactions)					0.0362*** (0.00344)	
ln(RtO Transactions)						0.0260** (0.00962)
Implied Elasticity ($\frac{1}{\beta}$)	4.366	6.757	13.583	20.964	27.624	38.461
Observations	15,857	15,857	15,857	15,858	15,858	15,858
R-squared	0.276	0.333	0.670	0.660	0.663	0.658
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes
	(7)	(8)	(9)	(10)	(11)	(12)
Panel B: Effects on SII's Property Appraisal Value (in log)						
ln(Stock)	0.406*** (0.00422)	0.348*** (0.00742)	0.223*** (0.0165)			
ln(Transactions)				0.234*** (0.0207)		
ln(NP Transactions)					0.203*** (0.00774)	
ln(RtO Transactions)						0.00102 (0.0226)
Implied Elasticity ($\frac{1}{\beta}$)	2.463	2.874	4.484	4.273	4.926	N.R.
Observations	15,862	15,862	15,862	15,863	15,863	15,863
R-squared	0.330	0.364	0.663	0.654	0.702	0.630
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes

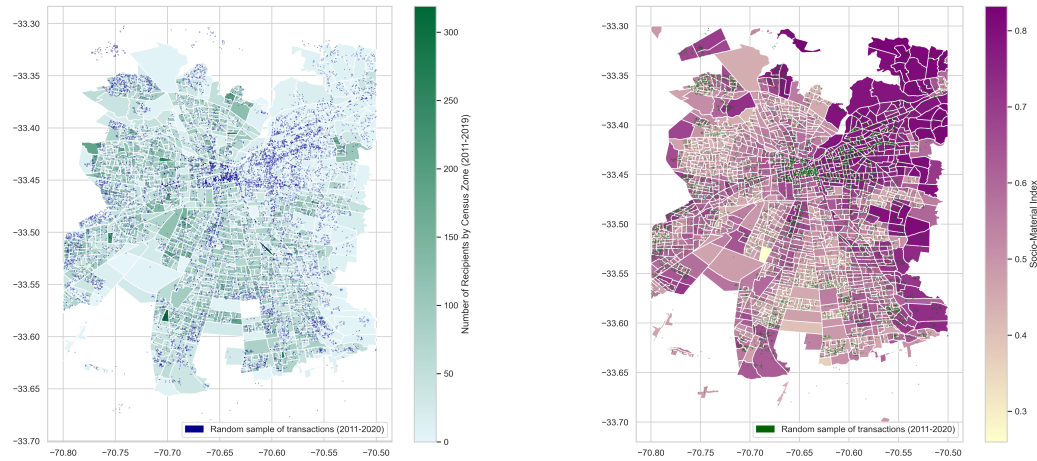
Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Each cell corresponds to the estimate from running an OLS regression of housing stock or transactions on a measure of housing prices (all are log-transformed). Panel A shows results on price per square foot, while Panel B shows results on the appraisal value estimated by the Chilean Revenue Service (SII). The regressions include controls at the neighborhood level, such as the share of different property types, the average construction year of properties, and total population. Year and neighborhood fixed effects are also included. “NP transactions” refer to new builds, while “RtO transactions” refer to the conversion of rental units to owner-occupied housing. Standard errors are clustered at the neighborhood level. Additionally, implied housing supply elasticities are calculated. For non-significant coefficients, the elasticity is not reported.

Table C3: Reduced forms and first stages

	(1) ln(Price)	(2) ln(Price sqft)	(3) ln(AV)	(4) ln(Stock)	(5) ln(Trans.)
Shift-share instrument	0.0534*** (0.00894)	0.0404*** (0.00437)	0.0634*** (0.0108)	0.0271* (0.0118)	0.0833*** (0.00946)
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	15,831	15,831	15,834	15,834	15,834
R-squared	0.637	0.667	0.639	0.666	0.569

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Each cell corresponds to the estimate of an OLS regression of a variable on the shift-share instrument. Columns (1), (2), and (3) present the reduced form regressions on the logarithm of price, price per square foot, and appraisal value, respectively. Columns (4) and (5) show the first stage results for housing stock and transactions (in logarithms). The regressions include controls at the neighborhood level, such as the share of different property types, the average construction year of properties, and total population. Year and neighborhood fixed effects are also included. Standard errors are clustered at the neighborhood level.

Figure C1: Transactions, voucher recipients and neighborhood quality by CZ

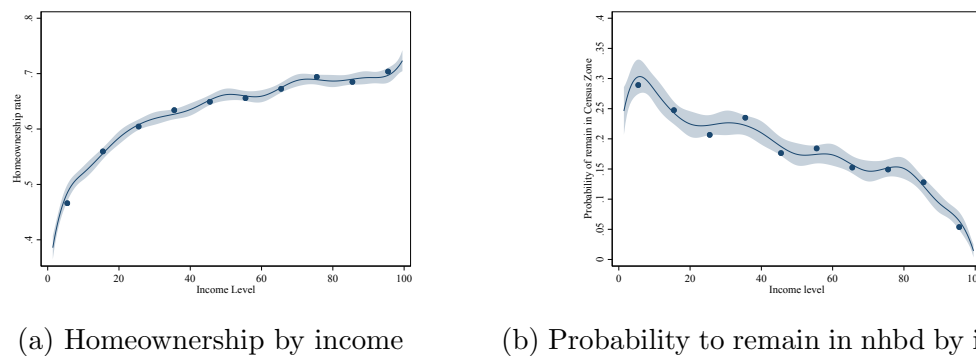


(a) Voucher Recipients & Transactions

(b) Socio-Territorial Index & Transactions

Note: This figure shows the spatial relationship between transactions, number of vouchers, and housing quality at the neighborhood level. Panel (a) shows the number of voucher recipients at the Census Zone (CZ) level from 2011 to 2019 and a random sample of transactions. The number of recipients includes the sum of vouchers classified under Title 0, Title 1, and Title 2. Panel (b) displays the Socio-Territorial Index (ISMT), which measures housing quality and socioeconomic conditions, alongside the corresponding transactions aggregated at the CZ level for the same period.

Figure C2: Homeownership and moving probability by income level



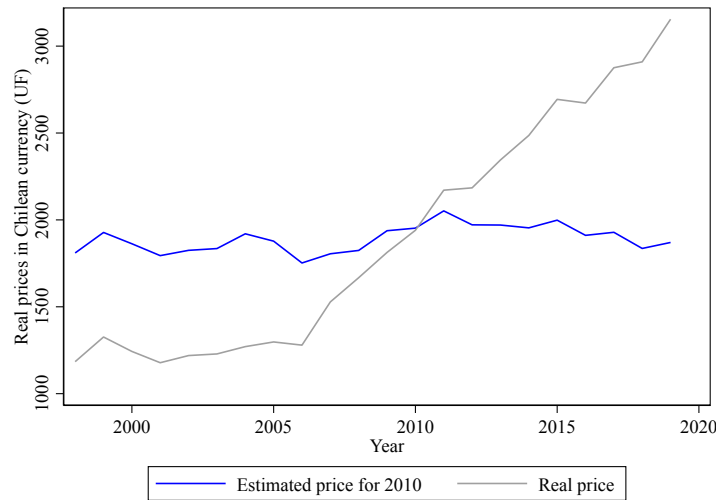
Note: The bin-scatter plots illustrate the homeownership rate and the probability of remaining in the Census Zone, both as functions of income level. In panel (a), we use data from the CASEN 2017 survey, which reports the tenure choices of individuals. In panel (b), we use voucher data, aggregated at the municipality level. Each bin represents deciles based on the average income of individuals in each municipality. The plots also include 95% confidence intervals. To make the plots, we use the method of Cattaneo et al. (2023) under the command *binsreg*, with a polynomial of degree 3 and 3 smoothness constraints.

Figure C3: Neighborhood choice by voucher recipients



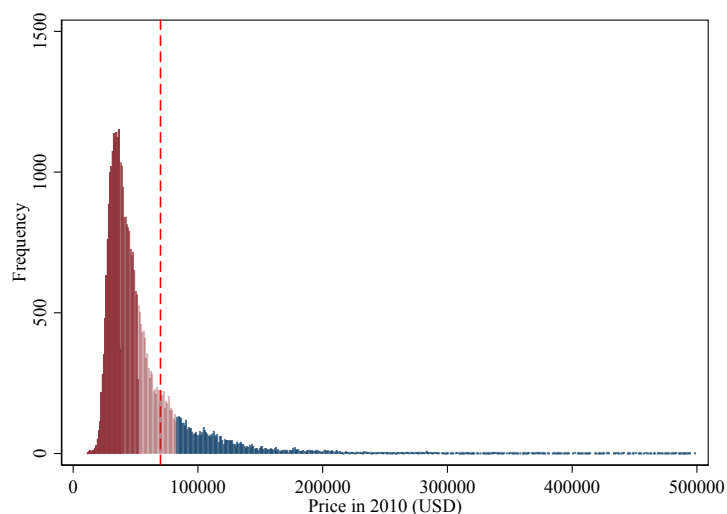
Note: This figure illustrates the residential location of voucher recipients at the Census Zone (CZ) level. The dashed line highlights the borders of the City's downtown, which we define as the municipality of Santiago. Panel (a) displays the neighborhoods of Title 0 voucher recipients, while Panels (b) and (c) show the same for recipients of Title 1 and Title 2 vouchers, respectively. Title 0 and Title 1 recipients tend to reside in the northern, western, and southern suburban areas, whereas Title 2 recipients are more widely distributed throughout the city, with a higher likelihood of residing in or around the downtown area.

Figure C4: Evolution of prices



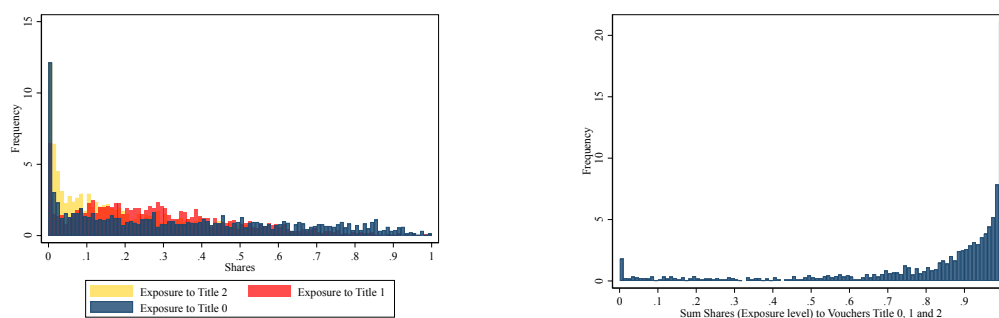
Note: This figure shows the evolution of average property prices over the years (in real terms) alongside the predicted prices for properties traded in all years, estimated at 2010 levels. The predicted prices are based on a regression of property prices on home characteristics such as size, age, type, and construction year. We include neighborhood, monthly, and yearly fixed effects. The prediction is made using the estimated coefficients from 2010. As expected, estimated 2010-adjusted values are relatively constant for properties traded in different years. To build the indicator of exposure to the policy, we only use data from before 2011.

Figure C5: Distribution of prices



Note: This figure shows the distribution of average Census Zone property prices in the year 2010, which serves as the basis for defining exposure to each voucher treatment. Properties eligible for Title 0 vouchers are highlighted in strong red, Title 1 eligible properties are in intermediate red, and Title 2 eligible properties are shown in light red. Areas that are not exposed to the voucher policy are depicted in blue. In the red dashed line, the figure shows the average price in Santiago de Chile in 2010, which is nearly USD70,000.

Figure C6: Histogram of Census Zone's exposure to each voucher type

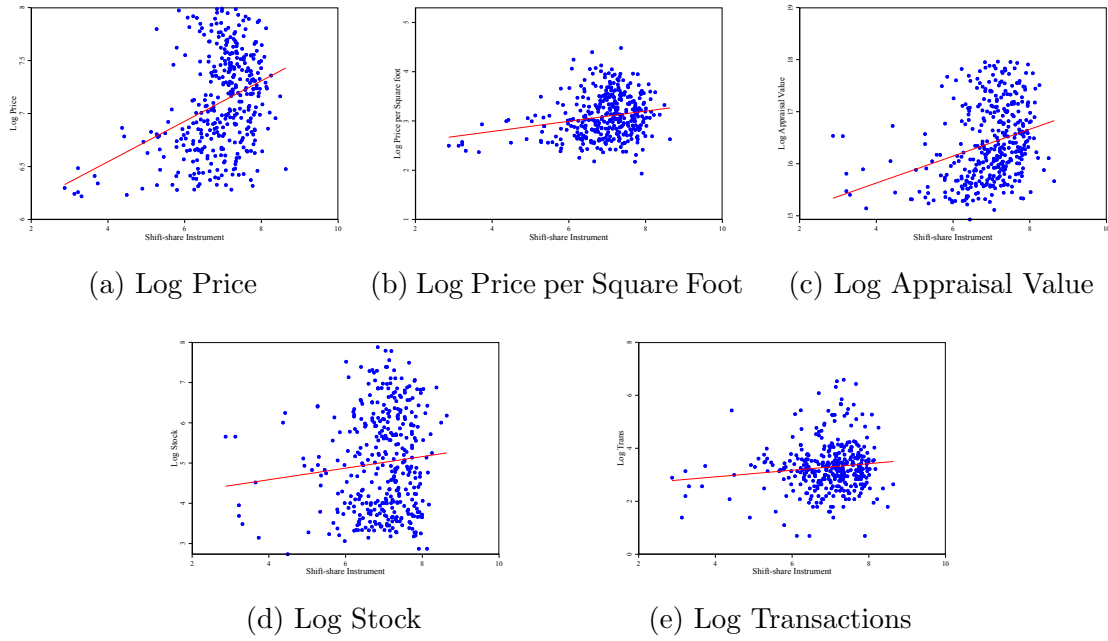


(a) Exposure to each voucher type

(b) Exposure to all vouchers

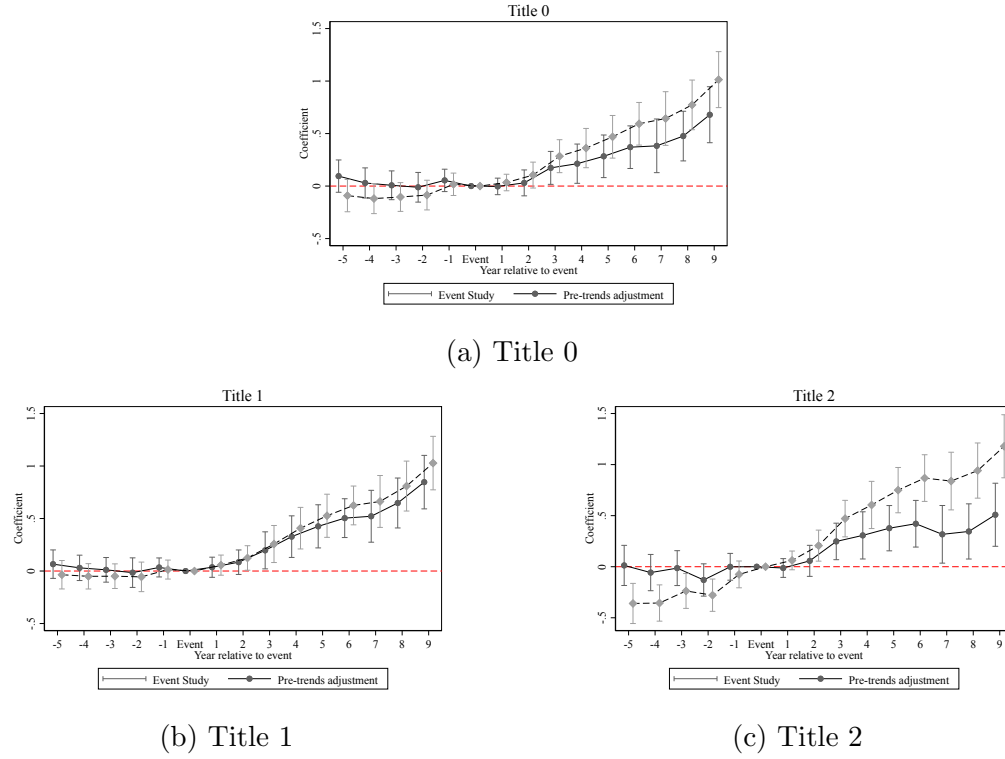
Note: This figure presents histograms depicting the distribution of Census Zones' exposure to the program. To create the exposure indicator, we count all the properties in each Census Zone that in 2010 would have had a price eligible to be purchased by each voucher. Panel (a) illustrates the exposure to each voucher type separately. Panel (b) shows the aggregate exposure of Census Zones to all vouchers, highlighting the overall level of exposure across neighborhoods. See Section 3.4 for more explanation about the methodology to define exposure levels.

Figure C7: Reduced forms and first stages scatter-plots



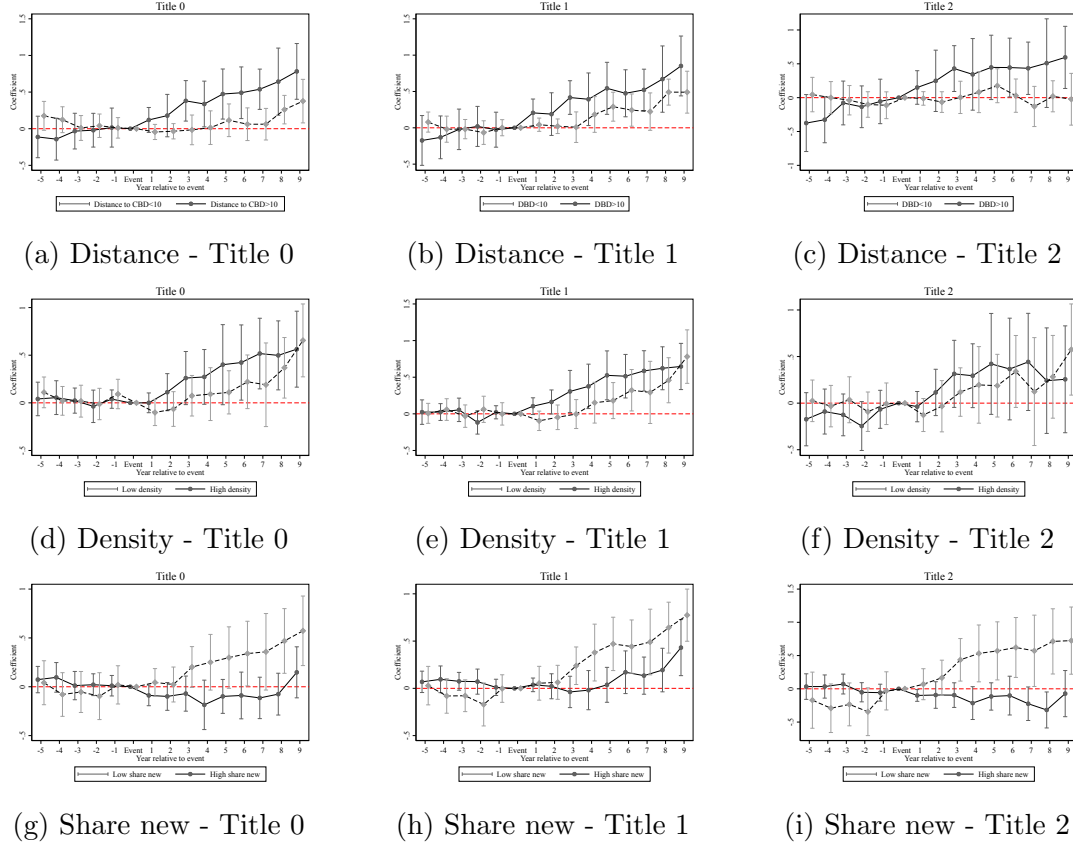
Note: This figure presents the reduced form and first stage results from the analysis. We present a random sample of 5% of the observations. Panels (a) to (c) display the reduced form relationships between the instrument and (a) Log Price, (b) Log Price per Square Foot, and (c) Log Appraisal Value. Panels (d) and (e) present the first stage results, where the instrument is used to predict (d) Log Stock and (e) Log Transactions. The significance of these relationships is detailed in Table C3.

Figure C8: Event Study on Local Prices with and without Pre-trends Adjustment



Note: This figure presents the results of an event study analyzing changes in property prices. Panel (a) shows the impact on prices for Title 0, Panel (b) for Title 1, and Panel (c) for Title 2. We make adjustments for pre-trends, following [Freyaldenhoven et al. \(2021\)](#). Bars represent 95 percent confidence intervals.

Figure C9: Heterogeneous effects in an Event Study framework



Note: The figure presents the results of an event study analyzing the heterogeneous effects of the voucher program on local property prices, broken down by three voucher types: Title 0, Title 1, and Title 2. The study examines three dimensions of heterogeneity: distance, density, and share of new properties. The top row shows the effect of the voucher program on property prices based on distance from the city center or a relevant focal point, with Panels (a), (b), and (c) corresponding to Title 0, Title 1, and Title 2, respectively, highlighting how prices respond differently depending on proximity to central areas. The middle row explores the effects based on neighborhood density, where Panels (d), (e), and (f) display the price impacts in lower- and higher-density areas for each voucher type, revealing the sensitivity of property prices to population density. The bottom row focuses on the price effects in neighborhoods with a higher share of new properties, with Panels (g), (h), and (i) showing the impact of each voucher type, illustrating how the presence of new constructions moderates the policy's impact. Bars represent 95 percent confidence intervals.