THE UNIVERSITY OF CHICAGO

ESSAYS IN URBAN POLICY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE IRVING B. HARRIS
GRADUATE SCHOOL OF PUBLIC POLICY STUDIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
LUIS ANDRES HERSKOVIC

CHICAGO, ILLINOIS
JUNE 2017

To my grandfather Enzo, my parents Ana and Pedro, and my wife María Paz, who have supported me unconditionally.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful for the help and support of many people at the University of Chicago throughout the years. In particular, I am thankful for the continued support and encouragement of my advisors, Ofer Malamud and Anjali Adukia, whose generosity, insight, and guidance I cannot begin to repay. I am privileged to have worked with them, and hope to carry with me what I have learned from them and share it with others throughout my career.

I am thankful for the generosity of Dan Black, who has shared his knowledge and ideas with every student that has attended the Harris School. His continued commitment to PhD students has been invaluable. I also thank Cynthia Cook-Conley, whose daily kindness and encouragement I've valued for these five years.

I thank my wife, María Paz, for leaving her beloved country and career to join me in this journey. Her support, enthusiasm and company made this experience better in every way.

I thank my parents, Ana and Pedro, for their unconditional support and encouragement. They have supported me without asking questions in every single step I have taken, and inspired me to try to make a contribution by following this path. I am thankful for this every day.

I will always thank Enzo, my grandfather, with whom I share so many interests. He has helped me and inspired me in many ways, and continues to do so.

Finally, I thank the friends I made in Chicago, who made the tough parts of pursuing a PhD easier. I will always treasure my years at the University of Chicago.

ABSTRACT

In the first chapter of this dissertation I present evidence examining the relationship between access to public transportation and school choice. One of the goals of school choice is to allow parents to send their children to higher-performing schools. Several studies have shown that distance to school is one of the main determinants of school choice, but challenges to address endogeneity issues remain. This chapter examines school choice in a context where vouchers have been implemented on a large scale and combines that analysis with a natural experiment to address the endogeneity concern posed above. In particular, I use information from Santiago, Chile, and take advantage of the construction of a new subway line that crosses a large area of the city previously unconnected to the subway network. I provide convincing evidence to show that the introduction of the subway line was arguably exogenous for families living close to the new subway stations. The high clustering of schools in certain areas of the city makes distance especially relevant for students who live in neighborhoods with little connectivity to transport networks. The use of rich administrative data allows me to calculate the distance from students' homes to each school in the city with high accuracy. Using an independent cross-sample difference-in-difference estimation, I find that (i) students near the new subway stations travel significantly farther to school than students who live in nearby areas with no subway stations, and (ii) that students near the subway are willing to travel farther to attend schools that perform better in standardized tests. This set of results is particularly informative for the ongoing school-choice debate, reconciling the advocates' and skeptics' views in the sense that school choice may lead to higher-performing schools once access restrictions are eased.

In the second chapter, I present causal evidence of the relationship between cell phone use and traffic accidents. Since drivers have incentives to underreport their cell phone use, and this behavior is not captured by official statistics, the magnitude of the problem

is not known with certainty. In this chapter, I take advantage of a government program implemented in Chile between 2010 and 2012, which connected rural areas to the mobile phone network for the first time. I employ two alternative differences-in-differences estimation strategies to produce a causal estimate of the impact of cell phone use on traffic accidents. As a first strategy, I take advantage of the phasing in of the program to compare localities treated in the first phase of the program with localities treated at a later stage. As a second strategy, I use GIS to generate and select random polygons in areas that already had cell phone coverage before the program started, and use these random areas as a control group for areas treated by the program. Both strategies lead to similar results. I find that the availability of cell phone coverage in a locality causes an increase of almost 30% in the number of accidents, which is consistent with the widespread and frequent cell phone use by drivers reported in several studies. This gives an indication that policy makers should approach the problem of cell phone use by drivers with the same or more urgency than they have approached other issues such as drunk driving.

In the third and final chapter, I examine the relationship between subway access and employment in areas surrounding subway stations. Although subways represent considerable investments and their construction often creates substantial disruptions to urban life, little is known about the effects of subway lines on firms located in the neighborhoods surrounding the stations. In this chapter, I look into the effect on employment of the opening of a new subway line in Santiago that connected several densely populated neighborhoods to the larger subway network. For this purpose, I take advantage of highly detailed yearly administrative data on all firms operating within the city between the years 2005 and 2009. Using an empirical strategy based on the geographic distance between firms and the new subway stations, I find no significant effect of the subway on employment in firms close to the stations. The one exception is a significant increase in employment related to real estate activities.

CHAPTER 1

THE EFFECT OF SUBWAY ACCESS ON SCHOOL CHOICE

1.1 Introduction

School choice was introduced to the Chilean school system in 1981 with the intention of generating competition among schools that would lead to an increase in the overall quality of the system. The idea behind this reform was that, if schools competed to attract students and parents valued quality, schools would improve their performance (Friedman, 1962, 1997). However, this policy does not seem to have produced the desired results (McEwan et al., 2008).

In many areas of the city with low-performing schools, students do not leave their districts to attend better schools. As I show in this paper, in Santiago during the year 2003, more than 85% of students living in the south side of the city attended schools in their own district despite the fact that schools in that area were performing worse than schools in the rest of the city. If students are not actively choosing better schools, the system does not improve as a result of implementing a school-choice policy. Did school choice fail to generate a system-wide improvement because parents don't value school quality as much as expected?

Some authors, such as Elacqua and Fabrega (2004) have argued that parents are uninformed consumers when it comes to school choice, claiming that they use only a few sources of information, don't include enough schools in their choice set, make decisions based on practical reasons, and don't have precise information about the schools that they do choose. Others, like Gallego and Hernando (2008) have disputed this claim, showing that quality seems to be one of the main factors affecting parents' choice of a school. Their results show that, in Santiago, "differences in quality seem to be the most important factor driving the decision to attend school in the home or other municipalities"

(Gallego and Hernando, 2008, pg. 211). These opposing views have persisted throughout much of the debate on the merits of the Chilean school-choice system.

It is possible, however, to reconcile the idea that parents do care about quality with the fact that in many cases they choose schools that are closer to their home instead of better-performing schools that are farther away. In this paper, I claim that one of the factors that determines school choice is access to transportation. Families sometimes will choose lower-performing schools not because they don't care but because the commute to better-performing schools is too long.

To test this idea, I take advantage of a natural experiment generated by the opening of a new subway line at the end of the year 2005 in Santiago, Chile. The inauguration of this subway line and the use of GIS tools, alongside a rich administrative dataset with detailed geographic information, allows me to estimate the effect of improved access to transportation infrastructure on school choice while addressing endogeneity issues.

This estimation strategy is based on the fact that the subway gave easier access to more and better schools to families living near the new subway stations but not to families in the same neighborhoods living farther away from the stations. Because of this, if families living near the new stations valued school quality, we would expect them to send their children to better-performing schools than families living farther away from the new subway line.

It can be argued that having access to the subway reduced transportation costs for students. Even if the cost of a subway ticket is the same as a bus ticket in Santiago (roughly $1), there was a considerable reduction in travel time when using the subway. This allowed students to potentially travel longer distances in the same time as their previous commute once the subway was inaugurated.

Using an independent cross-sample difference-in-difference estimation, I find that students who gained access to the subway with the construction of the new line traveled distances that are about 20% greater to get to school than students who lived farther away from the subway, and that the students with close access were willing to travel longer distances to attend schools that scored higher in standardized tests than students with worse access to the subway network. Thus, when travel times are reduced by better transport connectivity, families actually choose better-performing schools for their children, even if these schools are farther away.

This is in line with the evidence presented by Hastings et al. (2007), who show that parents for whom academic achievement is important are willing to leave their neighborhood to have access to better schools. Families face a trade-off between choosing a school with a shorter commute and choosing a school of higher quality. As the results of this paper show, when the restriction imposed by long commute times is eased, parents choose better-performing schools for their children.

This paper contributes to the literature on school choice, establishing that ease of access to transportation plays a prominent role in the schooling decisions made by families. Providing better access to schools helps parents make quality-oriented decisions. This shows that parents do care about school quality and that school choice combined with policies that improve access may result in system-wide improvements.

As school choice expands in the United States and other countries, it is especially relevant to understand which factors may have a direct effect on the potential success of this policy. Implementing a policy based on school choice with the goal of helping low-income students will have little or no effect if these same students do not have access to high-quality public transportation.

It is worth considering that even if all families care about getting higher-quality education, some families face greater restrictions in terms of the time it takes for their children to travel to school. Students living in districts without good schools and without good transport connectivity may end up being forced to choose worse schools than what their preferences for quality would indicate. This limits the impact that school choice may have as a policy on individual students and on the whole system as schools do not face as much competition as they could be facing.

The paper is structured as follows. Section II provides additional background on school choice, on the spatial distribution of schools, on the number of students who choose schools in other areas of the city, and on the subway network in Santiago, Chile. Section III describes the unique dataset and distance calculations used in this paper. Section IV introduces the strategy used to estimate the independent cross-sample difference-in-difference specification. Section V presents the results of the estimation of the effect of subway access on school choice. Section VI presents robustness checks. Section VII concludes the paper.

1.2 Background

The following subsections briefly discuss school choice within the Chilean school system, the spatial distribution of schools in Santiago, and descriptive statistics for these schools. They also present an overview of the subway network and its expansion.

1.2.1 School Choice in Chile

Starting in 1981, Chile began to implement an educational reform based on giving parents freedom in the choice of their children's school with the goal of improving the quality of the school system. This is clearly summarized by Hsieh and Urquiola (2002): "The notion that free choice is welfare enhancing is one of the foundations of modern, market-oriented societies. This view is prominent in the school-choice debate, where there is a widespread perception that public schools are inefficient local monopolies, and that the quality of education would improve dramatically if only parents were allowed to freely choose between schools" (Hsieh and Urquiola, 2002, pg. 1).

An interesting aspect—one that makes Chile a natural ground to study the effects of school choice—is that this reform consisted of a widespread voucher program that has lasted for more than 30 years, providing a unique case in the context of voucher programs. For a more detailed treatment of how the school system reacted to this reform over time, see Hsieh and Urquiola (2002) and McEwan and Carnoy (2000).

As part of the reform, the Chilean school system was organized around three types of schools: private, public, and voucher schools. Students could attend schools in any part of the city and received vouchers for public or voucher schools, with some voucher schools requiring an additional copayment by parents. That system still stands today but is the focus of major reforms.

Despite the long-running nature of this reform, the evidence on whether school-performance improved in Chile after implementing school choice in the 1980s is mixed. International evaluations do not show clear improvements. As Hsieh and Urquiola (2002) point out when examining international test data from TIMMS: "Using these exams we can assess whether years of unrestricted school choice have improved Chile's performance relative to the other countries... The evidence presented shows that its relative ranking, if anything, has worsened" (Hsieh and Urquiola, 2002, pg. 2).

One of the main arguments of the proponents of school choice—based on Milton Friedman's proposals—is that competition should induce schools to either improve their performance or disappear from the educational market, being replaced by better schools (Friedman, 1997). Again, in Chile it is not evident that this has happened. Corvalán and Román (2012) show that over 1,000 schools in Chile that have persistently under-performed in standardized tests have remained in the school system over time even when most of them are located in competitive environments, that is, with better-performing schools in nearby areas.

A frequent explanation to this apparent shortcoming of school choice has been that parents don't assign enough importance to quality when they choose a school (Elacqua and Fabrega, 2004). If parents care more about other factors, such as distance, religion or infrastructure, schools won't have incentives to improve their academic performance to attract more students. This is supported by survey evidence gathered by Elacqua et al. (2006), who conducted face-to-face interviews with parents in Santiago to study the determinants of school choice. They conclude that "despite the fact that parents in Santiago say they are seeking strong academic programs in their children's schools, they actually shop for schools that are widely different on academic quality but similar on socioeconomic dimensions. In short, as parents choose schools in Chile, class—not the classroom—may matter more" (Elacqua et al., 2006, pg. 578).

This follows the claims made by Hsieh and Urquiola (2002) who suggest that when families try to choose the best schools for their children, they may actually be trying to choose the best peer group. Additionally, they suggest that schools may respond to the incentives generated by school-choice policies but possibly do so by trying to attract better students in addition to improving their quality.

Other authors have argued that school quality does play a preeminent role in school choice. For example, Gallego and Hernando (2009) model school choice using information for 70,000 fourth-graders in Chile. They conclude that households consider several attributes when choosing schools but that the two most important ones are test scores and distance from home to school.

Regarding school choice and the relationship between distance and school quality, Hastings et al. (2007) have argued that there is considerable heterogeneity in preferences for schools, pointing out that "while many parents are very inelastic with respect to school test scores, there is a significant density of parents who highly value school scores and have low preferences for their neighborhood school. These parents are willing to consider schools outside of their neighborhood, and they place a high weight on average test scores when picking a school" (Hastings et al., 2007, pg. 19).

Chumacero et al. (2008) looked into the importance of distance from home to school when parents choose a school in Santiago. Even though they did not have the exact location of students' households, they estimated a probit model and found that there was a trade-off between distance and quality. They also found that parents were more likely to choose a school closer to their home if their child was a girl.

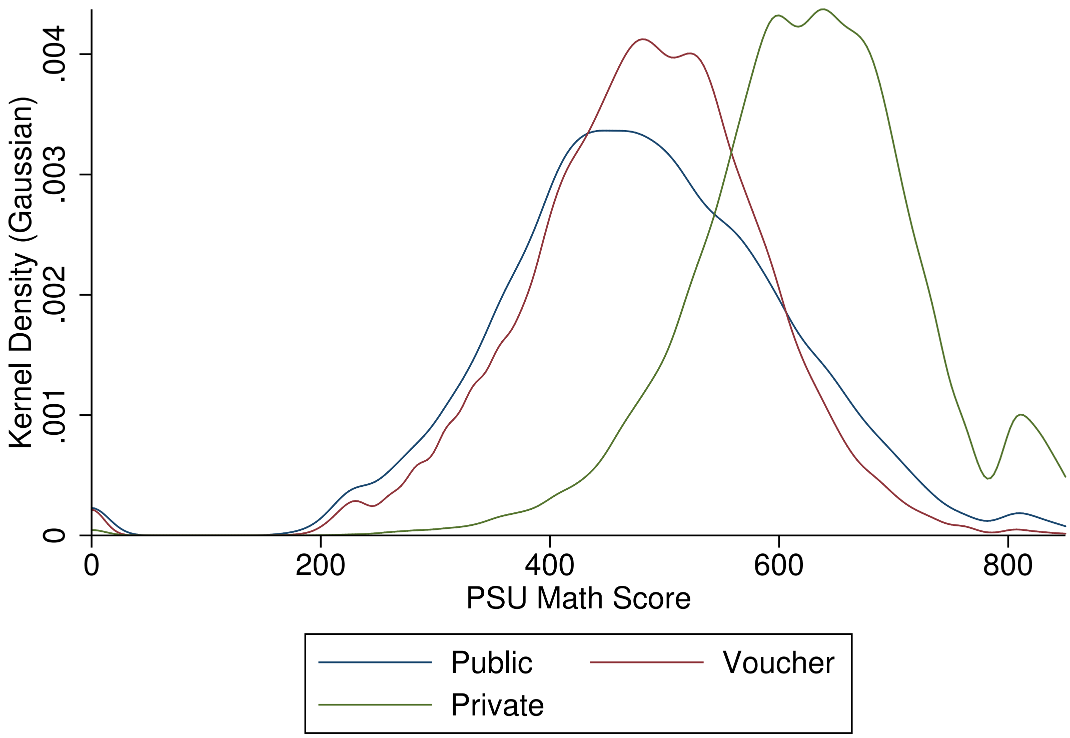
1.2.2 Spatial distribution of schools in Santiago

The previous section highlights the relevance that the distance between homes and schools may have when analyzing school choice. Families living in certain areas of the city may not have a realistic possibility of sending their children to better-performing schools, even if they want to. This is even more relevant in a city like Santiago in which schools are highly clustered by type and performance on standardized tests.

To illustrate this, I generate maps using inverse-distance-weighted interpolation (IDW) that reflect that, despite there being over 2,000 schools in the city, they are not spatially distributed in a way that makes them easily accessible to all families. Private schools, which outperform voucher and public schools in standardized tests, are in many cases more than 10 km away for a large number of students. The same applies to families that may want to send their children to high-performing schools. Using standardized math test scores as a measure of quality, it also becomes clear that high-performing schools are spatially autocorrelated. Figure 1.1 shows the distribution of university entrance exam scores for each type of school in 2007.

The maps in Figure 1.2 and Figure 1.3 show, respectively, that private and high-performing schools are highly clustered in the northeast area of the city. In Figure 1.2, greener areas reflect a higher proportion and clustering of private schools, yellow areas

Figure 1.1: Kernel Density of PSU Math Scores by School Type in 2007

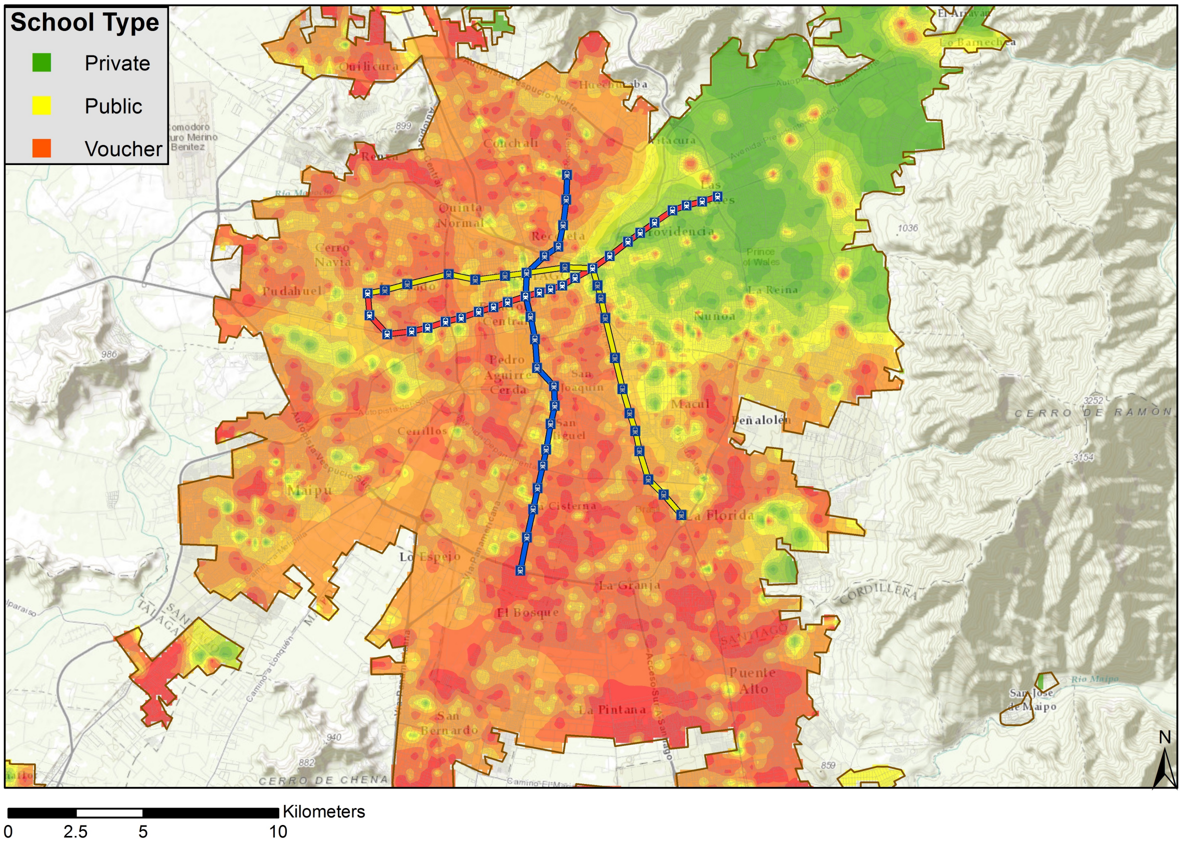


Note: Private schools outperform both voucher and public schools in all sections of the PSU. This graph shows the density of test scores for each type of school in the year 2007.

indicate a larger number of public schools, and red areas indicate a higher concentration of voucher schools. Figure 1.3 displays schools according to their performance in the math section of the SIMCE test on a color-scale that goes from green for the highest-performing schools, to red for the lowest-performing schools.

These figures present a visual representation of the challenge faced by students from other parts of the city who want to attend a high-performing school. Without good public transportation, the costs of attending a better school in terms of travel time may be too high for many students. The construction of a faster mode of transport, such as the subway, plays a large role in diminishing the importance of these large distances and

Figure 1.2: Map of Distribution of Schools by Type in Santiago, 2003



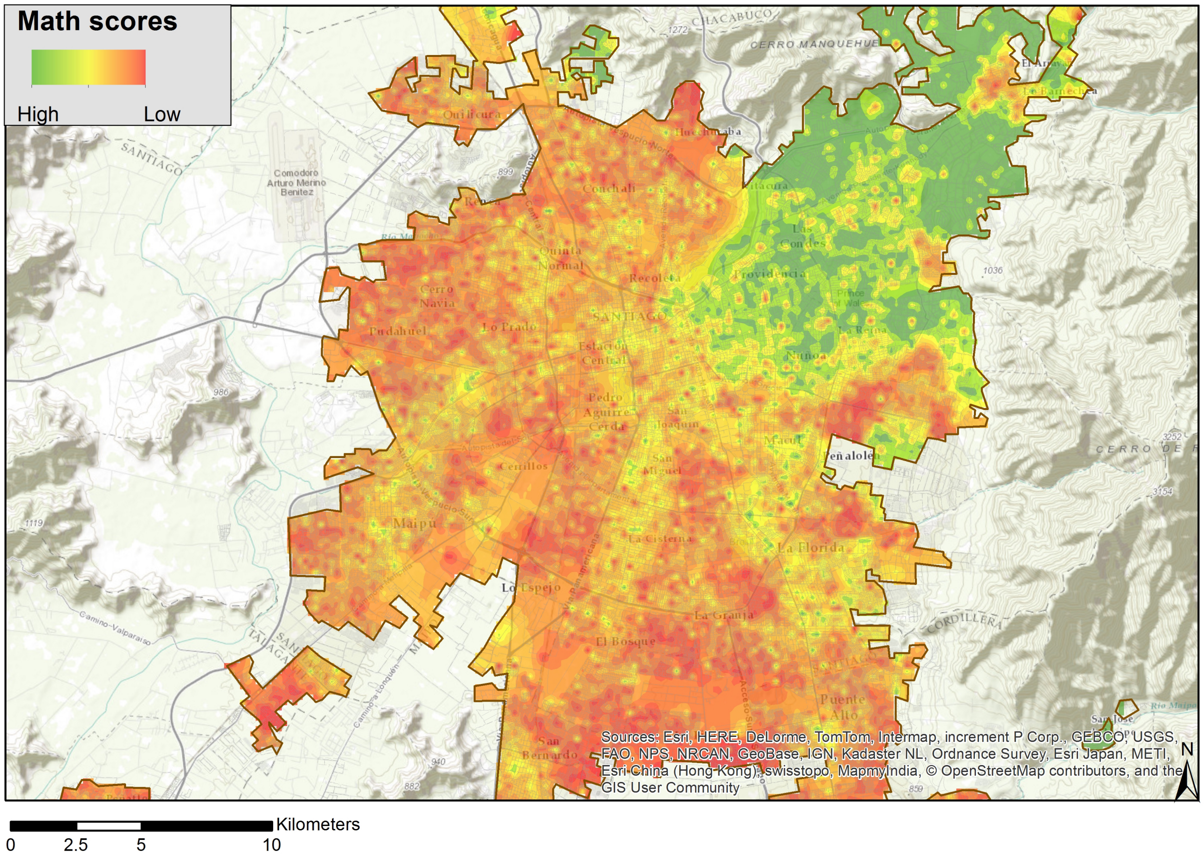
Note: This map was generated using school data from the SIMCE dataset. I geocoded every school in the city and then used GIS to perform an IDW interpolation, using school type as the point variable and the Metropolitan Region of Santiago as a polyline barrier. The 3 subway lines that existed in 2003 are displayed for reference.

expands the possible set of schools from which parents can choose. The subway line that was inaugurated in late 2005 reduced travel times from southern areas of the city to the north by more than 50% for subway users compared to bus users.[1]

The potential benefits of school choice rest on the assumption that students will actually choose schools in other areas of the city if there are no good schools close to their homes. This does not seem to be the case in several municipalities in Santiago. To

1. As an example, traveling from the southern end of the subway line to its northernmost station on a Monday morning at 8AM, takes about 1 hour and 40 minutes by bus, but only 35 minutes by subway, according to Google transit data.

Figure 1.3: Map of Distribution of Schools by Math Score Performance in Santiago, 2003



Note: This map was generated using math test score data from the SIMCE dataset. Test scores for all schools in the city are displayed using an IDW interpolation, using each school's math score as the point variable and the Metropolitan Region of Santiago as a polyline barrier.

show this, I analyze administrative records for the years 2003, 2005, and 2007 provided by the Ministry of Education for all students in the city. These records do not have students' exact addresses, but they do register the municipality the students reside in. Unlike the data I use to estimate the model, this dataset contains information about students from all grades. I use this information to show that a large proportion of students does not leave their own municipalities to attend schools in other parts of the city.

Table 1.1 shows the percentage of students who went to schools inside their own municipalities for the years 2003, 2005, and 2007 in Santiago. As can be seen, the percentage remains relatively stable at around 70%. That is, almost three out of four students did not leave their municipalities to attend school. This overall percentage however, hides large variations across municipalities.

Table 1.1: Percentage of Students that Study in Their Own Municipality Per Year Between 2003 and 2007

	%
2003	70.5
2005	71.8
2007	70.3

Note: This table is generated using data from the Ministry of Education for all students in Santiago during the years 2003, 2005 and 2007. I first calculate how many students are going to school in their own municipalities and show yearly averages of this value.

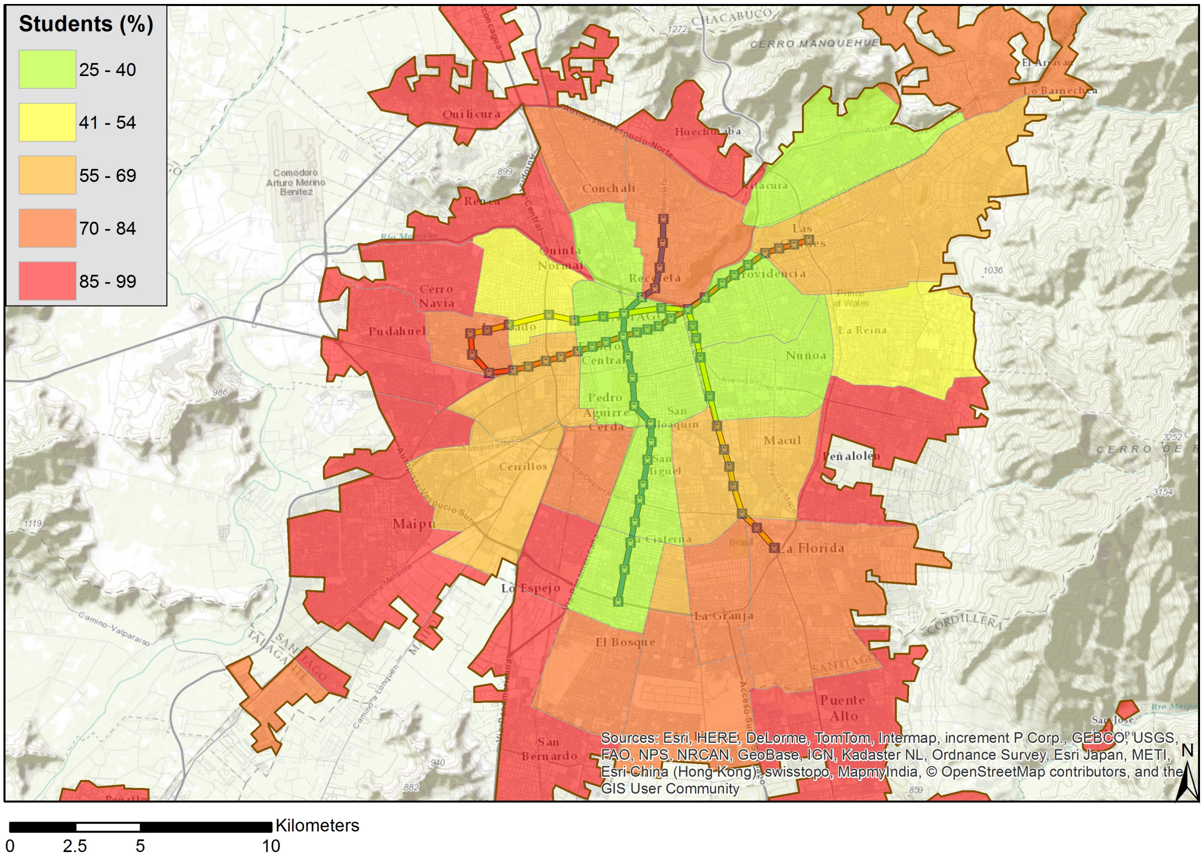
This becomes clear when examining the spatial distribution of students who do not leave their municipalities for school, as the percentages vary widely by city area. I generate a map to show the percentage of students attending school in the same municipality as they reside in, which is shown in Figure 1.4.

This map shows percentages for students attending schools in the same municipality as they live in, aggregated for each municipality. Municipalities in which most students travel to schools in other municipalities are displayed in green. Red areas show places in which more than 85% of students are living and attending schools inside the same municipality.

The map shown in Figure 1.4 reveals what might be an unexpected scenario for a city that in the year 2003 had been under a school-choice system for more than 20 years. In several municipalities, despite having mostly underperforming schools, high percentages of students did not leave their own municipality to find better-performing schools. In many cases, more than 90% of students living in a municipality were also studying there.

Without establishing any causality, this figure gives a first impression of the impor-

Figure 1.4: Map of Percentage of Students Attending School in Their Own Municipality in 2003



Note: This map was generated using official student registration data from the Ministry of Education. I aggregated data for each municipality and then calculated how many of the students living there were also attending schools inside the same area.

tance of access to transportation. Municipalities which had subway stations seem to also have been generally the ones that had the highest percentages of students traveling to take advantage of the possibility of attending schools in other parts of the city.

1.2.3 Descriptive statistics of schools in Santiago

Finally, to understand the context of school choice in Chile, it is relevant to observe the evolution of schools by type in Santiago over time, and the way students have distributed among these types of schools during the 2003-2007 period.

Table 1.2 shows the number of schools in Santiago per year by type of school. As this table shows, the number of private schools declines slightly during this period, while the number of public schools remains relatively stable. The number of voucher schools increases moderately but consistently every year, reflecting a trend that has been a permanent fixture since they were introduced. Overall, the number of total schools increased from 2,116 in 2003 to 2,332 in 2007.

Table 1.2: Number of Schools in Santiago Per Year Between 2003 and 2007

	Private		Public		Voucher		Total	
	N	%	N	%	N	%	N	%
2003	361	17.1	683	32.3	1,072	50.7	2,116	100.0
2004	325	15.2	681	31.8	1,138	53.1	2,144	100.0
2005	326	14.7	687	30.9	1,208	54.4	2,221	100.0
2006	326	14.4	678	29.9	1,267	55.8	2,271	100.0
2007	329	14.1	680	29.2	1,323	56.7	2,332	100.0

Note: Data used to construct this table comes from the SIMCE dataset. Each set of columns presents the number and percentage of schools for a particular year that corresponded to either private, public or voucher schools. The total number of schools in the city is displayed in the last set of columns.

Table 1.3 shows the number of high-school seniors by type of school each year. Overall enrollment gradually increased over time during the 2003-2007 period. Consistent with the previous table, the number of seniors in voucher schools increased more than the number of seniors in other types of schools.

Table 1.3: Number of High-School Seniors by Type of School in Santiago per Year Between 2003 and 2007

	Private		Public		Voucher		Total	
	N	%	N	%	N	%	N	%
2003	9,339	23.3	11,705	29.2	19,049	47.5	40,093	100.0
2004	9,176	22.3	11,543	28.1	20,349	49.5	41,068	100.0
2005	9,596	20.9	12,396	27.0	23,852	52.0	45,844	100.0
2006	9,811	17.9	15,765	28.7	29,296	53.4	54,872	100.0
2007	8,339	16.9	14,175	28.7	26,940	54.5	49,454	100.0

Note: These values correspond to the number of students that were accurately geocoded and thus used in this study, and do not correspond exactly to the total number of students in Santiago's school system. Data comes from PSU registration forms. Each set of columns presents the number of high-school seniors attending each type of school.The total number of high-school seniors in the city is displayed in the last set of columns.

As these tables show, the number of schools in the city increase by about 10% between 2003 and 2007. This is explained exclusively by the creation of new voucher schools. This increase in the number of schools would represent an expansion of the possible choice set of parents when choosing schools for their children, only if they have the means of actually attending those schools. As I will show, this seems to be in part determined by access to public transportation.

1.2.4 Subway network

Understanding the importance of access to transportation requires a brief overview of the subway network. Santiago is a city with a population of more than 6 million people, in which in the year 2013 more than 60% of trips in public transport were subway trips

(Metro de Santiago, 2013). The subway network construction began in 1969, and the first line (Line 1) was inaugurated in 1975. In 1978 the second line (Line 2) opened, followed by the third line (Line 5) in 1997.

Between November 2005 and March 2006, the fourth line (Line 4) of the subway system started operating. Line 4 is the longest line in the network, covering 24 km with 23 stations, and extending across 7 large municipalities. It transports around 500,000 people every day and it's connected to the other 3 lines in the subway network (Metro de Santiago, 2013; De Grange, 2010).

As I show in Table 3.3, the yearly ridership of the subway network in Santiago increased during the 2003-2007 period. Line 4 was inaugurated at the end of 2005, but some stations were not fully functional during part of 2006. This is reflected in the lower number of passengers during those years. By 2007 the line reached its maximum operating capacity with almost 115 million passengers per year.

Table 1.4: Number of Subway Users Per Year Between 2003 and 2007

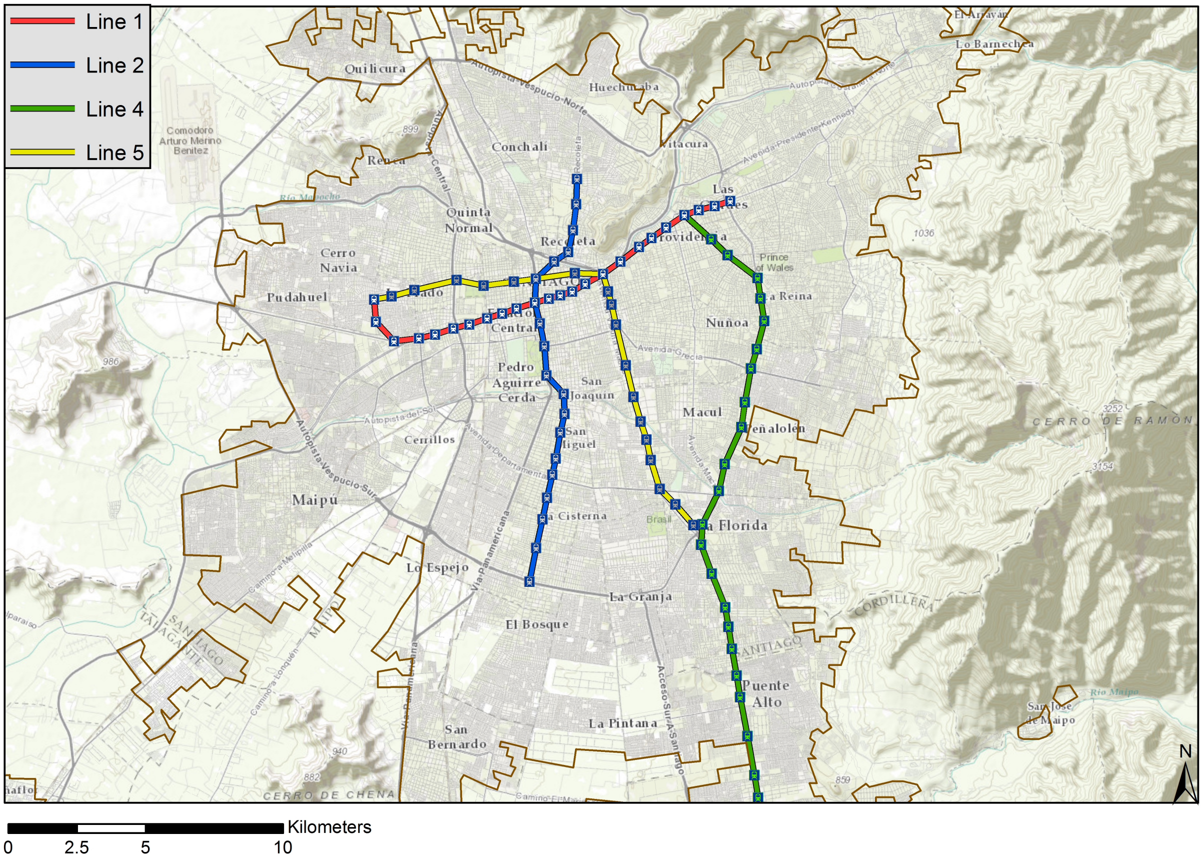
	2003	2004	2005	2006	2007
Line 1	130,749,000	145,192,992	158,254,000	167,192,000	256,036,992
Line 2	34,311,000	39,508,000	51,839,000	58,893,000	120,468,000
Line 4	0	0	2,692,000	48,419,000	114,008,000
Line 5	38,218,000	47,062,000	54,317,000	53,213,000	89,385,000

Note: There are 0 riders in Line 4 during 2003 and 2004 since it was inaugurated during the last months of 2005. Data comes from Metro de Santiago.

Before the expansion in 2005-2006, a vast area containing some of the most populated neighborhoods in the city was poorly connected to the transportation network, which consisted of Line 1, Line 2 and Line 5. Figure 3.2 shows what the subway networked looked like after the inauguration of the new subway line (Line 4) at the end of the year

2005. The network in the year 2016 transports more than 2 million passengers per day.

Figure 1.5: Subway Network in 2006 After the Inauguration of Line 4



Note: I generated this map by geocoding the exact location of each subway station that existed at the beginning of 2006, as informed by Metro de Santiago. The subway network has since been expanded in recent years with both the inauguration of new lines, and the extension of existing lines.

As I illustrate in Figure 3.2, Line 4 stretches towards the south of Santiago and is connected to the rest of the subway network. This is relevant because the new subway line connected a large area of the city that previously suffered from poor connectivity, greatly reducing travel times. Using the Google Distance Matrix, I performed several travel simulations of bus and subway trips between this southern area of the city and the northern end of Line 4, and found that travel time on average was reduced by about 50% when traveling by subway instead of by bus.

1.3 Data, variables, geocoding and distance calculations

To study the impact of the subway on school choice I use rich yearly administrative data which contains information on students' home addresses, school addresses and subway station addresses. I geocode every one of these addresses to find the coordinates that identify their exact location. Once students' homes have been geocoded, I identify which students live near the area of influence of the subway, and I also identify the students that live farther away. Then I calculate the distance between home and school for every student in the dataset. The details of this are explained in the sub-sections that follow.

1.3.1 Datasets

The main dataset I use for this study is generated from yearly Prueba de Selección Universitaria (PSU) registration forms from the years 2003 to 2007. During their high school senior year, Chilean students interested in attending higher education submit a registration form to take this test, which can be considered an SAT equivalent with a mathematics and language section, and other optional additional sections. The PSU test has a minimum score of 150 points, a maximum of 850, a mean of 500 points and a standard deviation of 110. I consider only students that are taking the test for the first time, and leave out students that are are registering to re-take the test since they are not high school seniors.

This dataset contains information on students' age, gender, test scores, GPA, the school they attended on their senior year, number of family members that are working, and their home address when registering for the test. Data is only available for high-school seniors each year, so I compare each year's cohort with the seniors from the other years in the 2003-2007 period.

It is important to note that students that do not register to take the PSU, cannot be included in this analysis. About 80% of high school seniors register for the PSU each year, with students from lower socio-economic backgrounds being the ones that usually don't register for the test (Consejo de Rectores de las Universidades Chilenas, 2015), which restricts the interpretation of the results of this study to students that registered for the test.

The second dataset I use is generated from SIMCE[2] test score results. This test, established in 1988, is applied yearly to all 4th graders in Chile, and on alternating years to 8th and 10th graders. It measures achievement in mathematics and verbal skills. Additional test subjects, such as an English language proficiency test, have been added over time. Possible scores range from 200 to 350 points, with a mean of 250 points and a standard deviation of 50.

The SIMCE data I use is aggregated by schools. In this study I look at average SIMCE math scores by school, and also take advantage of this dataset, using it as a census of all schools in the city, including the address of each school and the year it was established. It is relevant to point out that SIMCE results are widely reported on and the Ministry of Education devotes large amounts of resources to disseminate the results for each school in various ways.

Every year during the 2003-2007 period considered in this paper, between 50,000 and 60,000 high school seniors register to take the PSU for the first time. I am able to accurately geocode more than 90% of students' home addresses each year. I consider results as accurate if the coordinates obtained are either a precise street address or a named route. Addresses geocoded as sublocalities, neighborhoods or other administrative area levels are left out of the final sample.

2. SIMCE is the Spanish acronym for Quality of Education Measuring System

1.3.2 Variable definitions

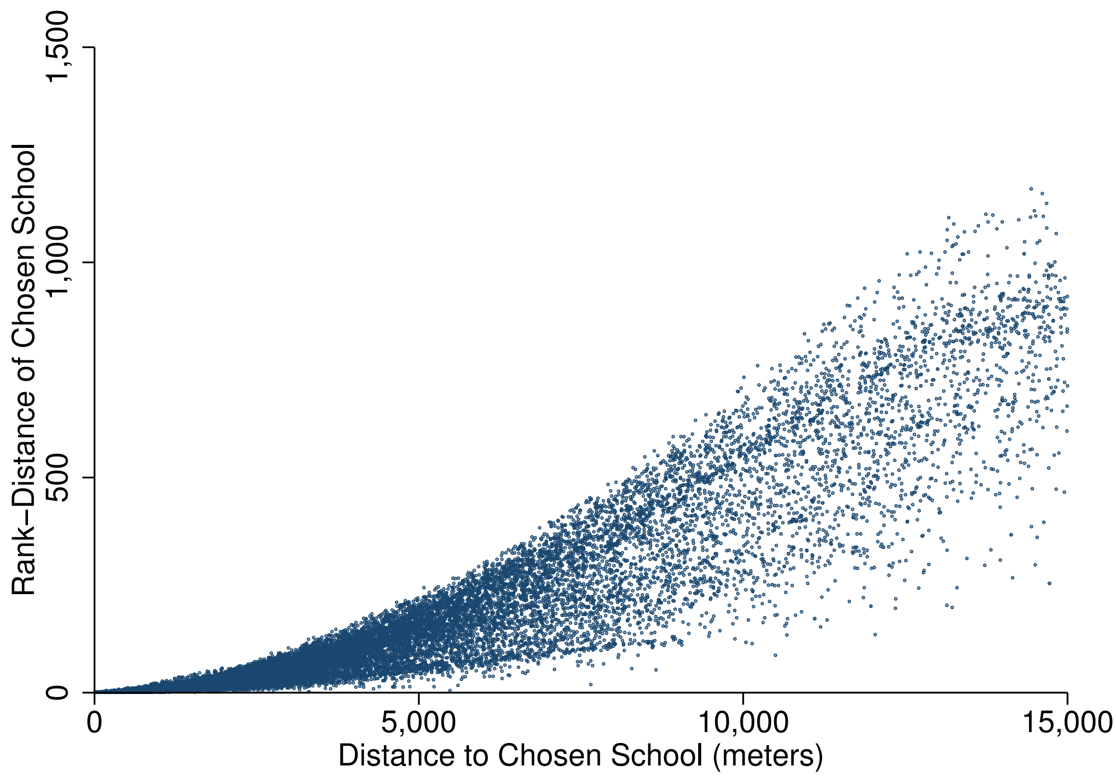
I consider three alternative outcome variables in this paper: home-school distance in meters, the logarithm of home-school distance, and rank-distance of schools. The first of these outcomes is the distance between home and school. Since the datasets contain home and school addresses, I geocode them to obtain their exact latitude and longitude with the objective of being able to calculate distances. It's important to note that this is not a measure of travel-time, which would require an assumption about the transport mode used by students, such as car, subway, bus, or walking. Calculating distances as I do in this paper approximates a realistic travel pattern for urban areas regardless of transportation mode. I explain this process in more detail in the next subsections.

The second outcome is the logarithm of the distance between home and school. This outcome allows for a more intuitive interpretation, in which the coefficient of interest indicates the percentage change in the distance traveled to school once the subway line is inaugurated.

The last outcome is the rank-distance of schools, which produces results which require an alternative interpretation. For this, I rank schools for each student according to their distance from that student's home. For example, the school closest to a student's home is ranked number one, the second school in distance to that house is ranked number two, and so on. In this way, a rank of 100 means that a student had 99 schools closer to her home than the school she chose to attend. This rank shows how many schools a student skips over in terms of distance before choosing the school she finally attends. The relationship between rank-distance and home-school distance in the year 2003 and 2007 is shown in Figure 1.6 and Figure 1.7, respectively.

The treatment variable is defined as students that live near a new subway station. That is, students living in a 2.5-km radius of a new subway station are considered as

Figure 1.6: Distance to Chosen School and Rank-Distance of Chosen School in 2003



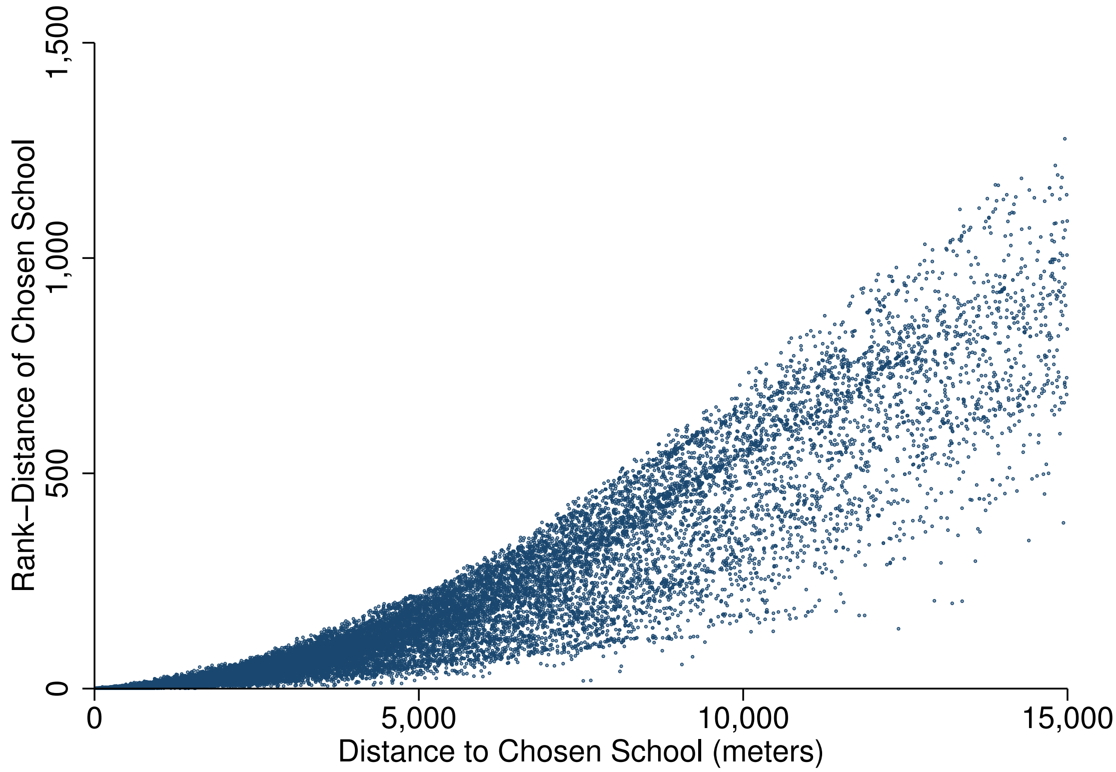
Note: This figure shows the relationship between rank-distance to school and distance to school in meters for the year 2003.

treated. Students living between 2.5 km and 4 km from a subway station are part of the control group.

The choice of these precise distances is somewhat arbitrary, although based on other studies of subway use. For example, a study of dutch rail stations shows that usage of stations largely declines as the distance between residence and station increases, and the largest declines seem to happen at around 3 kilometers from the station (Keijer and Rietveld, 2011). As shown in Section VI, slightly different distance definitions does not alter the results.

I also consider several control variables in the estimation. Among them, I look at the

Figure 1.7: Distance to Chosen School and Rank-Distance of Chosen School in 2007



Note: This figure shows the relationship between rank-distance to school and distance to school in meters for the year 2007.

number of schools in a 1-km and 5-km radius from the student's home (I also test several other distances, which does not alter the results). This allows me to control for the fact that students living near the subway also live closer to a larger number of schools, since understandably schools are more often located in larger avenues which coincide with the places the subway line is built through.

Other variables I consider as controls are the number of a student's family members that are currently working and the student's gender. Finally, I also consider the type of school the student attends, which can take 3 possible values: public, voucher, or private.

1.3.3 Geocoding

I geocoded approximately 200,000 student home addresses for this paper. For this purpose, I developed a custom script that first corrected errors in the addresses and then, over several months, used a combination of geocoding services provided by Google and ESRI to find the coordinates of homes and schools. This process made it possible to achieve a high level of accuracy (defined as either "rooftop" addresses, or routes), reflected by the fact that for each year of data, I was able to geocode about 90% of addresses accurately. This differs from other papers using Chilean data that have used distances measured from the center of municipalities, which may be several kilometers away from actual home addresses.

1.3.4 Home-school distance calculations

The first outcome of interest in the estimation is the distance between home and school. There are several possible ways to calculate these distances. For simplicity, a considerable number of previous research papers have used a "straight-line" distance calculation to measure the distance between 2 points. But since in this paper I am measuring distances within urban areas, it is much more accurate to use what is commonly referred to as "block" or "Manhattan block" distance (Boscoe, 2013).

The difference between these 2 types of measurements is non-trivial, since even when looking at a distance as short as a few city blocks, using a straight-line calculation will yield a distance considerably shorter than the more realistic block calculation.

As a simple example of this, note that if a student travels only two standard city blocks of 100 meters each to get to school, which include turning a corner, the straight-line distance calculation underestimates the distance traveled by 30% compared to the block-distance calculation.

The formula used to calculate block distance follows Gimpel and Schuknecht (2003):

$$d_i = |x_i - x_j| + |y_i - y_j|,$$

where x_i is the longitude of home i, and y_j is the latitude of school j. As stated, this formula assumes that students don't travel in a straight line between home and school, which is a more realistic representation of travel patterns within cities.

Table 1.5 shows the average distance traveled by students to get to their school by year. As can be seen, the average distance is about 5.5 km and it decreases slightly over time. One possible explanation for this is the increasing overall number of schools in the city as shown previously in Table 1.2.

Table 1.5: Average Distance Traveled to School by Year Between 2003 and 2007

	Distance (meters)
2003	5,665
2004	5,549
2005	5,137
2006	5,136
2007	5,199

Note: This table contains the average distance traveled by each student in Santiago from their home to school, calculated as block distance. I calculate these values after geocoding every home address from PSU registration forms and every school address from the SIMCE dataset.

1.4 Estimation Strategy

I use an independent cross-sample difference-in-difference specification to estimate the effect of subway access on school choice. In this approach, variation is determined by the distance between houses and the new subway stations, which in the short term should be arguably exogenous. The validity of this method would be threatened if families living near the subway would have changed the average distance they travel to school relative to families living farther away regardless of the construction of the subway line, or if certain types of families were more likely to locate closer to the subway than others. This is partly addressed by considering a period close to the opening of the subway line, comparing characteristics between groups before and after the subway opened, and examining the trends followed by each group.

Regarding this same issue, it is important to point out that although Line 4 was publicly announced during 2001, the initial announcement only revealed the general placement of the line with no information on specific stations. One way to address the possibility that families who cared more about education may have moved to homes near the subway is to look at changes in housing prices at the time the subway line was announced. Agostini and Palmucci (2005) do this and find that housing prices near the new subway line reacted in a very limited magnitude to the announcement, increasing around 3% with respect to locations farther away from the line. This would seem to indicate that a massive displacement of people living near the subway did not take place.

Additionally, in this paper I consider a period of time that begins shortly before the inauguration of the subway line and ends after it has been operating for only one year. This may also reduce concerns about the area surrounding the stations changing its social composition or infrastructure because of the new subway. Finally, in Section VI, I present results considering only the years before the subway line opened. This also provides evidence that supports the idea that students living near the area where the subway stations were constructed, were not traveling longer distances to school before the subway was inaugurated.

A different threat to the identification strategy proposed in this paper would be the possibility that schools are selecting students instead of families selecting schools. That is, if schools are the ones controlling which students are admitted, it would be possible for them to select students based on characteristics that would capture differences besides access to the subway.

This, however, does not seem to be the case in Chile. The evidence points in the opposite direction. That is, families are the one making these decisions, not schools. Gallego and Hernando (2008) show that sorting in the Chilean school system comes from the demand side and not from the supply side. They point to evidence from a 2006 survey, which found that 93% of parents actually said that their children were attending the school they had wanted to choose as their first option. Supporting this, parents applied on average to only 1.1 schools.

1.4.1 Treatment and Control Groups

For the analysis that follows, I consider that people who live close to a subway station are more likely to use the subway than people who live farther away. With this in mind, I define the treatment and control groups as follows: the treatment group is composed of high-school seniors living within a 2.5-km radius of a new subway station. The control group comprises high-school seniors living between 2.5 km and 4 km from a new subway station. Figure 3.4 presents a map I generated by using a fishnet that captures students in each group to show this more clearly.

Table 1.6 shows a comparison of several relevant variables between control and treatment groups before the subway line opened in the year 2003. In Table 1.7, I show this comparison for the year 2007, that is, after the subway line opened.

Table 1.6: Descriptive Statistics Pre-Subway (2003)

	Control	Treated
Student and Home characteristics		
Male student	0.5	0.5
Family members that work	1.5	1.5
Number of schools in 1km radius	4.6	7.2
Number of schools in 5km radius	117.2	143.3
School characteristics		
Average math score	275.1	277.3
Home-School distance	6247.2	5475.0
Observations	2689	8394

Note: This table is generated using data from the PSU registration forms. The first set of variables correspond to individual student characteristics, averaged for both the control and treatment groups in 2003. The second set corresponds to school characteristics for each group during the same year.

Table 1.7: Descriptive Statistics Post-Subway (2007)

	Control	Treated
Student and Home characteristics		
Male student	0.5	0.5
Family members that work	1.3	1.3
Number of schools in 1km radius	4.6	7.0
Number of schools in 5km radius	116.8	141.5
School characteristics		
Average math score	271.4	271.8
Home-School distance	5332.6	5152.8
Observations	3011	8854

Note: This table is generated using data from the PSU registration forms. The first set of variables correspond to individual student characteristics, averaged for both the control and treatment groups in 2007. The second set corresponds to school characteristics for each group during the same year.

The first of the variables shows the average distance traveled by students to their chosen school. Students in the control group in the year 2003 traveled longer distances on average to get to school than students in the treatment group: 6.2 km vs. 5.5 km, respectively. This could be explained by the fact that the control group had fewer schools near their homes, as shown by the average number of schools closer than 1 km and 5 km. Students in the control group had then on average 117 schools within a 5-km radius of their homes, while students in the treatment group had 143. It is important to point out that while there are differences in the values of these variables between groups, the differences remain similar over time except for the distance traveled to school, which is the outcome of interest.

These variables help address the possibility that schools may have reacted to the new subway line by positioning themselves near stations to attract more students. This could affect the estimation, since students in the treatment group would suddenly face an increase in the number of schools near their homes. This does not seem to be the case, as shown in Tables 1.6 and 1.7, since the number of schools within a 1-km radius from the students living close to subway stations was 7 in 2003 and 2007 and the number of schools within a 5-km radius changed from 143 to 142.

The percentage of students who are male also is similar between groups—close to 50% for the treatment and control groups—before and after the subway was built. Students in the control group attended schools that scored on average 275 points on the SIMCE math test in 2003 and 271 points in 2007. Meanwhile, schools attended by students in the treatment group scored 277 points in 2003 and 272 in 2007. Finally, the average number of family members who were working was 1.5 for both groups in 2003 and 1.3 for both groups in 2007. As can be seen in Tables 1.6 and 1.7, all of these variables have similar values and their differences remain practically constant over time.

1.4.2 Pre-treatment Trends

Defining the groups as explained in the previous section makes it likely that the treatment and the control groups will experience similar changes in their characteristics on average, although students in the treatment group will be more likely to use the subway as their means of transportation. These definitions are used to estimate a difference in difference specification, which assumes that both groups followed similar trends before the treatment. Visually this can be seen in Figure 1.8, which shows trends in covariates for each group before and after the inauguration of the subway line. There doesn't seem to be an important change in the characteristics of people living near the subway once the stations open. This helps address the possibility that once the subway opened, people with a specific set of characteristics may have immediately moved to areas surrounding the subway.

The panels in Figure 1.8 show several trends followed by the treatment and control groups. Panel (a) shows the percentage of students in each group that attends public schools. As can be seen, both groups follow relatively parallel trends over time, with about 25% of students attending a public school. As has been commonly observed in Chile, this percentage gradually declines over time. The relationship between the trends followed by both groups remains constant even after the subway line is operating in 2006 and 2007.

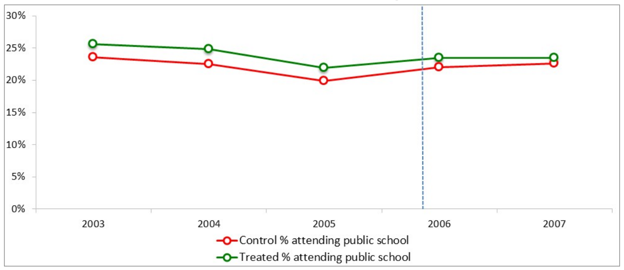
Panel (b) shows the senior-year grade average for both groups. This panel shows that there were no differences in the achievement of students in the treatment and control groups once the subway started operating and that they followed very similar trends between 2003 and 2007.

Panel (c) shows the average distance between the closest school and each student's home. It is important to point out that this is not the distance to the actual school the student attended but the distance to the school closest to their home. This figure reveals that students in the treatment group, on average, had a school closer to their home than students in the control group but also that the groups followed parallel trends over time. This did not seem to be affected by the opening of the subway line, which is consistent with the idea that schools did not immediately react to the construction of the subway line by changing their locations.

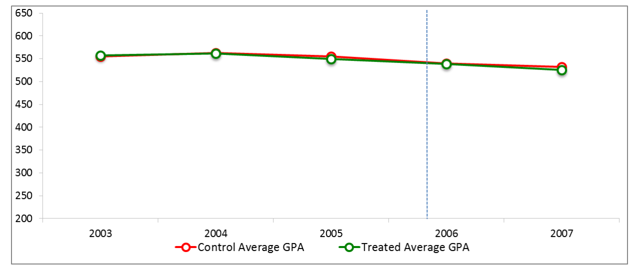
Finally, panel (d) shows the number of schools located within a 1-km radius of each student's home. As is the case with the other variables, even if there were more schools near students in the treatment group than in the control group (7 and 4, respectively), which is consistent with the results from panel (c), the groups still followed parallel trends between 2003 and 2007.

As can be seen from these tables and figures, the treatment and control groups were

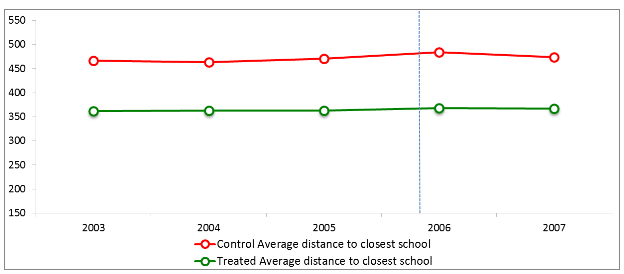
Figure 1.8: Trends of Covariates Between 2003 and 2007



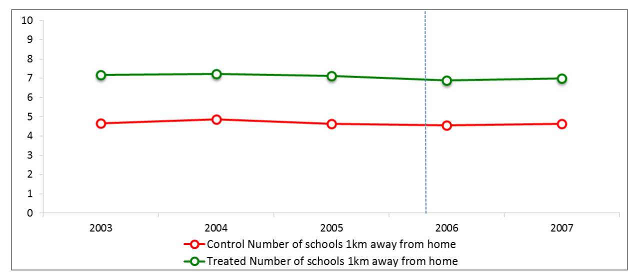
(a) Percentage Attending Public School



(b) Average GPA



(c) Average Distance to Closest School



(d) Number of Schools within a 1-km Radius From Home similar before and after the subway was built and they followed similar trends over time. There did not appear to be a significant event that affected one group and not the other besides the inauguration of the new subway line.

1.4.3 Effect of Subway Access on Distance to School

I first estimate the effect of subway access on the distance between home and school to see if families that got subway access sent their children to schools that were farther away than families that did not get a station nearby. After this, I include an interaction to estimate the quality of the chosen school to see if parents who lived near the subway were in fact choosing better schools.

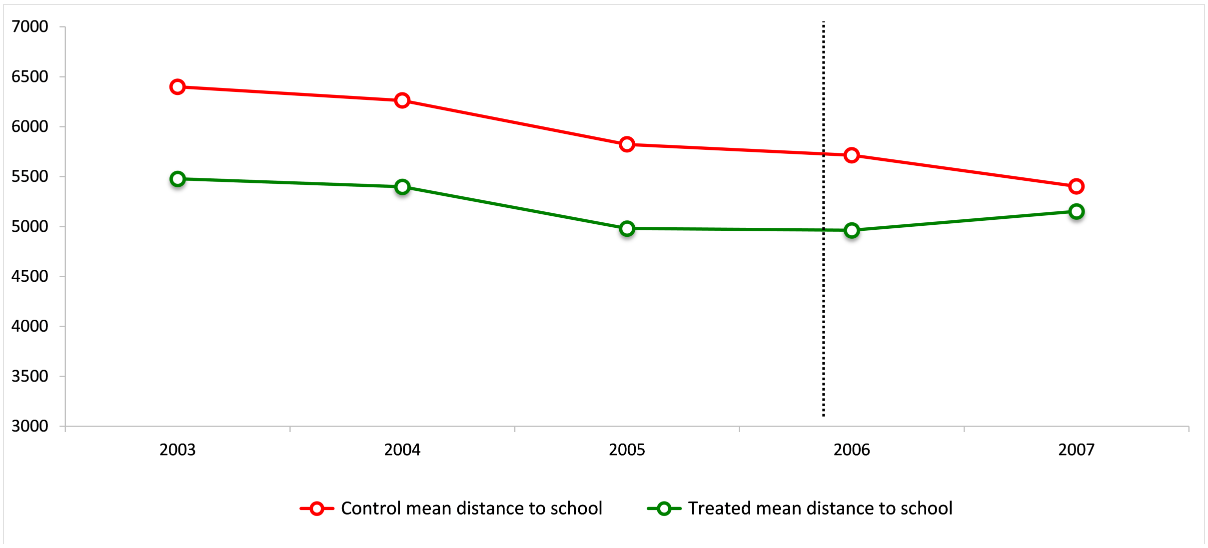
I consider that high-school seniors who lived within a 4-km radius of a subway station of the new line in the year 2003, before it started operating, are part of the "before" group. In turn, students who were high-school seniors in the year 2007 and who lived within a 4-km radius of a subway station of the new line are part of the "after" group.

The group that lived near the subway line (within a 2.5-km radius) after the subway line opened were the students most likely to be using the subway, so they will be considered as being in the treatment group, while the rest of high-school seniors, who lived between 2.5 km and 4 km away from the subway, will be considered as being in the control group.

The key identifying assumption of the difference-in-difference estimation is that both groups would have followed similar trends if the subway line had not been built. Even if it's not possible to test this assumption directly, by looking at graphs of the trends of each outcome variable it's possible to see whether both groups followed parallel trends until the subway line was inaugurated.

I show the parallel trends first for the distance between home and school in Figure 1.9. As the lines in the chart show, individuals in the treatment and control groups followed similar trends in the years before the subway opened. Students in the treatment group traveled on average about 5,500 meters to go to school, while students in the control group traveled almost 1,000 meters more. Both trends are parallel until the subway opens, which caused students living near the subway to start attending schools that were farther away from their homes.

Figure 1.9: Average Distance Traveled to School by Year Between 2003 and 2007



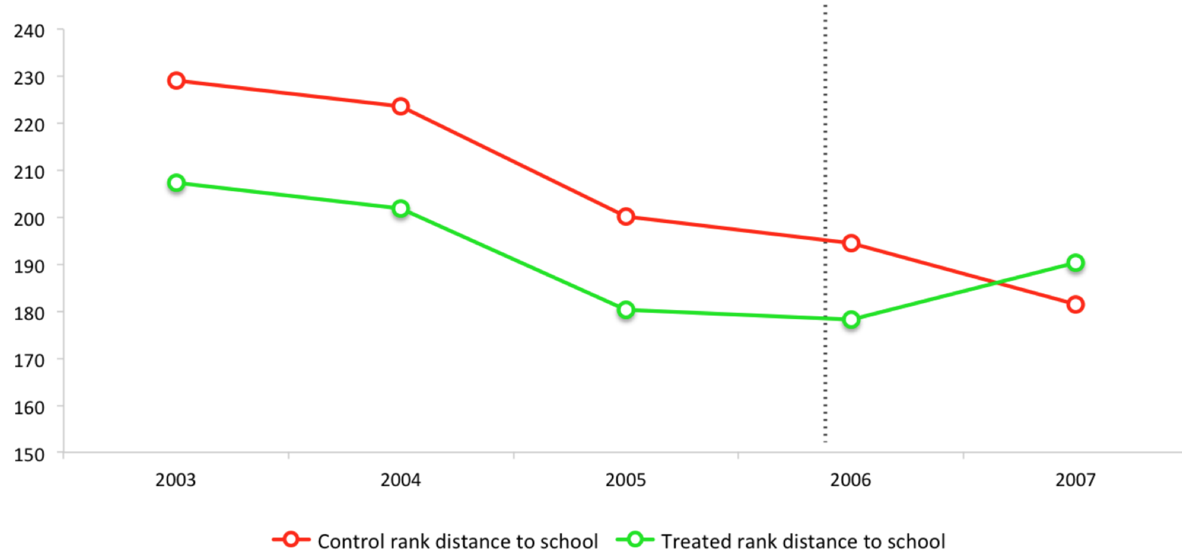
Note: This chart is generated using data from PSU registration forms and SIMCE datasets. Each data point represents the average distance traveled from home to school by students either in the treatment or the control group on a particular year.

The same pattern is present when looking at the alternative outcome, rank-distance to school, as presented in Figure 1.10. The figure shows that, for example, that students in the treatment group in 2003 on average, had almost 210 schools closer to their homes than the school they were actually attending. Each year, students attended schools closer to their homes in terms of rank-distance, but once the subway opened, students living near the subway started skipping over more schools than students in the control group.

The key equation that I estimate is the following:

$$\text{Distance}_{it} = \beta_0 + \beta_1 \text{Subway}_i + \beta_2 \text{Post}_t + \gamma(\text{Post}_t * \text{Subway}_i) + X_{it} + \epsilon_{it} \qquad (1.1)$$

Figure 1.10: Rank-Distance to School by Year Between 2003 and 2007



Note: This chart is generated using data from PSU registration forms and SIMCE datasets. Each data point represents the average rank-distance of the chosen school relative to a student's home for students either in the treatment or the control group on a particular year.

in which the outcome variable, *Distance*, is the block distance in meters between a student's home and their school, for student i in year t. I also estimate equation 1.1 using the logarithm of distance between home and school and the rank-distance of schools as alternative outcome variables. *Subway* is a binary variable that is set to one for students that live in houses that are within a 2.5-km radius of a new subway station. *Post* is a binary variable that is set to one if a student is a member of a cohort that graduated from high school after the new subway line was built.

The parameter γ, which corresponds to the difference-in-difference estimator, captures the effect of increased subway access on the distance between the school a student was attending and their home. It is important to note that this parameter captures the intention to treat, since the available data does not contain information that could be used to identify whether students were using the subway. Nonetheless, it seems reasonable to assume that students living closer to the subway were more likely to use it than students living farther away.[3]

If students who lived near the subway started attending schools that were farther away from their homes after the subway opened, we should expect to observe a positive and significant coefficient for this parameter.

X_{it} is a vector of control variables, which includes the gender of the student, a variable indicating if family members worked, the number of schools that were closer than 1 km to the student's home, the number of schools that were closer than 5 km to the student's home, and dummy variables for the type of school, where private schools are the omitted category. Since the identification strategy being used does not depend on the inclusion of these covariates, it is expected that they should not affect the results.

1.4.4 Effect of Subway Access on Distance to School by Test Scores

Additionally, I also estimate a model that includes the interaction between the treatment group and a measure of school performance: standardized math test scores. The idea behind this is to see if students living near the subway were traveling farther away to get to a better-performing school. There is some evidence that shows that families have not generally been aware of SIMCE results (Corvalán and Román, 2012). It is possible, though, that awareness increased over time, since scores are now widely reported on news coverage and the Ministry of Education at various times has released maps and other types of reports showing easy-to-understand representations of the test score of each school.

3. Information available from a large-scale transportation survey conducted in 2012 in Santiago shows that in the municipalities that surround the subway line studied in this paper, only about 35% of adults have a driver's license, and about half of the households don't own a car. About one third of all subway passengers are students (OSUAH, 2012).

The key equation in this case is the following:

$$\begin{aligned}\text{Distance}_{itj} = \beta_0 + \beta_1\text{Subway}_i + \beta_2\text{Post}_t + \beta_3\text{Test}_j + \beta_4(\text{Post}_t * \text{Subway}_i) + \\ \beta_5(\text{Post}_t * \text{Test}_j) + \gamma(\text{Post}_t * \text{Subway}_i * \text{Test}_j) + X_{it} + \epsilon_{itj}\end{aligned} \tag{1.2}$$

in which the outcome variable is the block distance in meters between a student's home and school, for student i and school j in year t. I also consider the alternative outcomes used previously: the logarithm of the distance between home and school and the rank-distance of schools.

The $Test_j$ variable corresponds to the SIMCE test math score of school j, which range from 200 to 350 points..

The parameter γ now captures the effect of increased subway access on the distance traveled to school when attending a school that scores higher on standardized tests. Thus, if students near the subway were willing to travel farther to attend better-performing schools, this coefficient should be positive and significant.

X_{it} is a vector of control variables, which includes the gender of the student, a variable indicating if family members worked, the number of schools that were closer than 1 km to the student's home, the number of schools that were closer than 5 km to the student's home, and dummy variables for the type of school, where private schools are the omitted category.

1.5 Data Analysis and Results

In this section, I look at the results from the estimation of the effect of subway access on the distance between home and school, its interaction with school performance, and results for sub-groups of students.

1.5.1 Results: Effect of Subway Access on Distance to Chosen School

The first set of results presented in Table 3.10 are from the estimation of equation 1.1, with and without covariates, using three different outcomes: the home-to-school distance in meters, the logarithm of the home-to-school distance, and the school rank-distance. Columns 1, 3 and 5 present estimations without covariates, while Columns 2, 4 and 5 include covariates.

The estimated effect of subway access on distance between the chosen school and students' homes is positive and significant, and it is practically unchanged when adding control variables, as was expected. The coefficients in Columns 1 and 2 show that having access to the subway results in attending schools about 600 meters farther away on average than students who did not live near the subway, which represents an almost 10% increase over the average distance traveled by all high-school seniors in the city, not just those in the control group.

Columns 3 and 4 in Table 3.10 also reveal interesting results. The coefficient of 0.2, using log distance as the outcome variable, can be interpreted as showing that having access to the subway allowed students to choose schools that were 20% farther away than students in the same neighborhood who did not get direct access to the subway. This result is important because it shows that students are willing to travel farther away to school if they have access to faster transportation. The new subway line reduces travel time, and this allows students living close to the stations to travel larger distances than they could before it opened.

Finally, Columns 5 and 6 in Table 3.10 present results using the alternative outcome of rank-distance to school. This allows for a complementary yet alternative interpretation. The positive and significant coefficient of 24.2 shows that students in the group treated by the subway skipped an additional 24 schools when choosing a school to attend, compared

Table 1.8: Regression Results (DID): The Effect of Subway Access on Distance Traveled to School

	Distance		Log distance		Rank-Distance	
	(1)	(2)	(3)	(4)	(5)	(6)
Access to subway	592.417***	609.691***	0.200***	0.200***	24.225**	27.118***
	(168.404)	(159.826)	(0.035)	(0.033)	(9.416)	(8.908)
Observations	22,046	22,046	22,046	22,046	22,046	22,046
R^2	0.004	0.103	0.004	0.098	0.002	0.108
Controls	No	Yes	No	Yes	No	Yes

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Columns (2), (4) and (6) present results controlling for the following variables: gender of the student, number of family members that are working, number of schools in a 1-km radius from the student's home, number of schools in a 5-km radius from the student's home, and dummy variables for the type of school.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

to students in the control group. That is, they went 24 schools farther away than they would have if they had not received access to the subway.

The full results, shown in Table 1.9, also reveal interesting results. Male students traveled almost 7% farther to school than females. This is in line with findings presented by Chumacero et al. (2008) who propose that parents are more willing to let their sons travel longer distances to school than their daughters.

The presence of more schools near a student's home (closer than 1 km and closer than 5 km) is associated with a decrease in the distance students travel to school, as would be expected. Having an additional school within a 1-km radius of home is associated with a reduction in the distance traveled to school of about 130 meters. The presence of an additional school within a 5-km radius has a much smaller but still negative effect on distance traveled to school.

Restricting the sample by using different definitions of the treatment and control groups does not alter the main results. Since there is no data on which students were using the subway, it is relevant to note that defining the treatment group as a wider or narrower area around the new station produces similar results in terms of the direction and magnitude of the effect. This is shown in more detail in Section VI.

1.5.2 Results: Effect of Subway Access on Test Scores of Chosen School

The results presented in the previous section show that subway access had a relevant effect on the home-school distance traveled by students living near the new subway, but these results do not address the question about what is the performance in tests of chosen schools. Table 1.10 examines the results of the estimation of equation 1.2, which introduces the interaction with school test scores. These results are relevant because once it is established that families living near the subway choose schools that are farther away, it is important to know if the schools they are choosing are better-performing than the ones chosen by families without subway access.

The coefficients in Columns 1 and 2 in Table 1.10 are positive and significant and do not change substantially when control variables are included in the estimation. These coefficients were obtained using distance between home and school as the outcome variable, and they correspond to the interaction between the treated group in the period after the subway was inaugurated and school test scores. The coefficients can be interpreted as showing that students living near the subway traveled 12 additional meters to attend a school that scored one point higher on a standardized math test in which scores ranged from 200 to 350 points. This is consistent with an estimation done by Chumacero et al. (2008) who also found that students were willing to travel an extra 12 meters for an extra point in this test.

This result is particularly important because it reflects the value assigned by parents and students to school performance. Families were willing to send their children farther away if they were attending better-performing schools, indicating that travel restrictions play a relevant role in school-choice policies. Parents seem to value school performance, but it's possible that students face too many difficulties when trying to access better-performing schools and end up attending schools closer to their homes that are not as high-performing as the ones they would like to attend.

These results are similar to those derived using the log of the distance between home and school, which are presented in Columns 3 and 4. The latter results show that students that had access to the subway would travel 0.2% farther to attend a school that scored one point higher on the SIMCE test. The estimations with and without covariates produce similar coefficients, which are positive and significant.

Columns 5 and 6 show results using rank-distance to schools as the outcome variable. Again, both coefficients are positive but only the coefficient of the estimation that includes control variables is significant. This coefficient of 4.4 reflects that a student would skip over four schools closer to her home to attend a school that performed one point higher on the test.

In summary, these results give clear indications that once transportation restrictions are eased, students are willing to choose higher-performing schools even if these are farther away. On the aggregate level, if transportation was improved for a larger number of students, we could expect an increase in competition between schools to attract students that are now able to reach a wider set of schools. This, in turn, could lead to overall increases in the quality of the school-system, as school-choice policies originally intended.

1.5.3 Results: Effect of Subway Access by Subgroups

Finally, I examine results from the estimation by subgroups. I separate students first by the type of school they attended and then by how they performed on the PSU test.

First, I show that having access to the subway did not affect the type of school that a student attended, be it public, voucher, or private. For this, I estimate a variation of equation 1.1 in which the outcome is the type of school. I do this one time for each possible type of school. This way, it's possible to estimate the impact of subway access on the type of school chosen by students. As can be seen in Table 1.11, Columns 1, 2 and 3 have coefficients that are close to 0 and not significant. This can be interpreted as showing that subway access did not affect the type of school a student chose. Considering this result along with the previously established results, it's possible to state that students who got access to the subway chose schools that performed better and that were farther away, but they did not choose a specific type of school.

Along with this, to see if results vary according to the type of school students were attending, I again estimate equation 1.1, this time separately for students according to the type of school they attended. That is, I estimate the equation three times: once for public-school students, once for voucher-school students, and once for private-school students. The results of this estimation are presented in Table 1.12. These results show that students from private and voucher schools traveled more than 700 meters farther when they had access to the subway. Students from public schools that had access to the subway did not attend schools that were significantly farther away compared to students in public schools without subway access.

The second subgroup corresponds to students divided into three quantiles by their PSU math test scores, classified as Low, Medium, or High. The goal of this is to estimate if having access to the subway had an effect over a student's performance on university entrance exams. Table 1.13 shows that having access to the subway did not have an effect on test scores. The coefficients in Columns 1 and 2 are not significant and close to 0. The coefficient in Column 3 is significant at $p < 0.1$ and close to 0. Even if students with subway access were attending better-performing schools, their tests scores did not improve. One possible explanation for this is that the administrative records used in this estimation corresponded only to high-school seniors. It's likely that students were switching schools in their last year of high school to prepare for the PSU, but this doesn't seem to have had an effect on actual test scores.

1.6 Robustness Checks

1.6.1 Alternative Treatment and Control Group Definitions

In this section I examine how robust the results are to changes in the definition of the treatment and control groups. I replicate the estimation strategy by estimating equation 1.1 under different definitions of the treatment and control groups with the goal of showing that slightly different definitions do not alter the main results of this paper. Since the definition of treatment and control groups is somewhat arbitrary, despite being based on available data, it is important to show that results do not depend on the exact definitions used.

For this, I generate 3 alternative treatment and control groups. The first one considers that students living within a 1.5-km radius of the new subway stations are treated and students living within a 1.5-km and 3-km radius are in the control group. The second definition considers that students living within a 2-km radius of the new subway stations are treated and students living within a 2-km and 5-km radius are in the control group. Finally, the third definition considers that students living within a 3-km radius of the new subway stations are treated and students living within a 3-km and 5-km radius are in the control group. The results of the estimations using these definitions are shown in Table 1.14.

Column 1 and Column 2 replicate the main results restricting the treated group only to students that live in closer proximity to the subway, without and with covariates. This tighter 1.5-km radius assumes that only students living very close to the subway are using the new stations. The coefficients under this specification are smaller than the ones obtained when the original treatment and control groups are used, but they still are positive and significant, and can be interpreted as saying that students living within a 1.5-km radius of the subway travel about 430 meters more than students living within 1.5-km and 3-km away from the subway. Column 3 and Column 4 present results of the estimation using a slightly larger treatment group (2 km away from the subway) and then considering a larger control group of students living within 2-km and 5-km away from the new subway stations. Again, the coefficients are positive and significant and show that changing the definitions slightly does not alter the main conclusions of this paper. Finally, Column 5 and Column 6 present results using an even larger treatment group, which considers students living within a 3-km radius of the subway as being treated. These results are similar to the ones obtained using the definitions considered in the main estimation. In summary, the results presented in this paper do not depend on the exact distance ranges used to define the treatment and control groups.

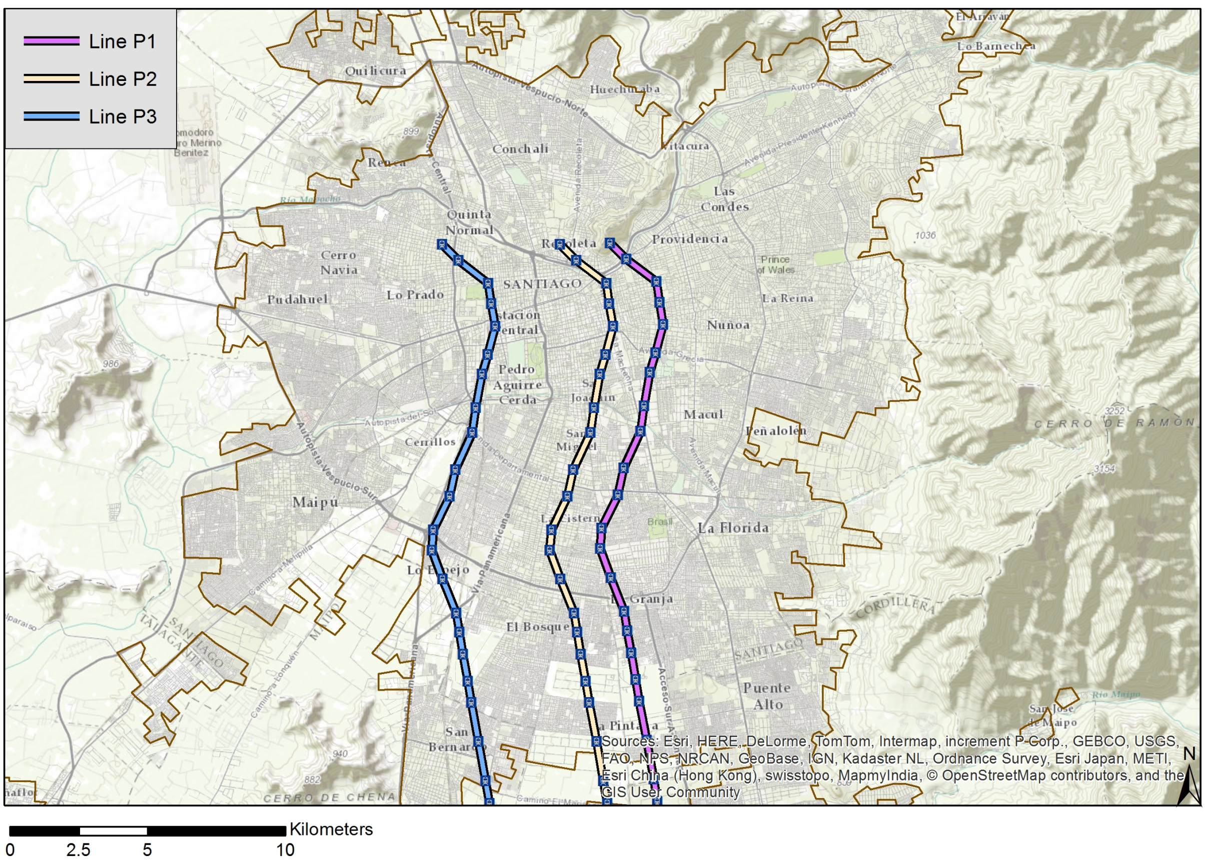
1.6.2 Falsification Tests

In an attempt to provide evidence that the estimation is capturing the impact of subway access on distance traveled to school, I estimate equation 1.1 using only the years before the subway line was built. If the subway is causing students to attend schools that are farther away, we should expect to see no significant effect before the subway line was constructed. The results of this estimation are shown in Table 1.15.

Column 1 and Column 2, show the results of estimating equation 1.1 using only information for the years 2003 and 2004 using distance between home and school as the outcome. As expected, the results are not significant. That is, students living near where the subway was going to be constructed, did not travel significantly longer distances than students living farther away. Column 3 and Column 4 show the same results using the logarithm of home-to-school distance as the outcome, and Column 5 and Column 6 show the results using rank-distance between home and school. The results as before, are not significant.

As a complement to this strategy, I generate placebo subway lines and reproduce the estimation using these non-existent lines. For this, I generate subway lines that are the same shape and length than Line 4. I spatially offset these subway lines by altering the latitude of the original subway stations and maintaining their longitude constant. This results in a series of subway lines that are parallel to Line 4, which are restricted to inhabited areas in Santiago. A map with 3 of these placebo lines is shown in Figure 1.11.

Figure 1.11: Map of Placebo Subway Lines



Note: I generated this map by offsetting the latitude of the subway line that was inaugurated in 2006. I did this several times to generate alternative placebo subway lines to use in falsification tests. This map displays three of those subway lines.

Using these new geographic points, I reproduce the entire procedure by calculating the distances between every student's home, school and subway station. I then estimate equation 1.1 using these values. Since there are no real subway stations going through the areas being considered, we should expect to see no effect of being treated by the placebo subway lines.

As expected, living near these placebo subway stations has no effect on distance traveled to school. Results of these estimations using each of the 3 placebo subway lines are presented in Table 1.16, Table 1.17 and Table 1.18. In each table, the first 2 Columns present results using home-to-school distance, without and with covariates. Column 3 and Column 4 show the same results using the logarithm of home-to-school distance as the outcome, and Column 5 and Column 6 show the results using rank-distance between home and school. None of the coefficients are significant.

1.7 Conclusion

The main results from the difference-in-difference estimation show that households that gained access to the subway started sending their children to schools that were 20% farther away, on average, and that the students with close access were willing to travel a longer distance to attend better-performing schools than students who lived farther away from the subway. These results do not change when control variables are included in the estimation.

Another way of to illustrate this using an alternative outcome is to state that, after the subway was built, parents living near the new subway stations sent their children to schools that were on average about 24 schools farther from their homes than parents who did not have easy access to the subway line. That is, they skipped over 24 schools that were closer to their homes and chose schools that performed better in standardized tests.

Parents face a restriction created by problems with access to transportation when choosing schools. When this restriction is relaxed by the construction of a new subway line, parents with access to the subway start choosing higher-performing schools even if these are farther away from their homes. Thus, policies that seek to improve school quality via creating a system that increases competition between schools must necessarily consider the relationship between access to transport systems and school choice. This set of results is particularly informative for the ongoing school-choice debate. It is commonly claimed that school choice has not produced system-wide improvements in Chile because parents do not care enough about school quality when choosing schools. I show that parents care enough to send their children to better-performing schools even if they are farther away as long as it's a realistic possibility in terms of access. A system in which schools are highly spatially segregated will be unlikely to produce quality improvements if students have difficulties accessing the schools they would like to attend.

Schooling systems that foster competition among schools may still result in system-wide improvements. In this case, the construction of a subway line reveals that competition might eventually bring higher quality into the system but that it might not have an effect on its own as long as other challenges have not been addressed.

Table 1.9: Regression Results (DID): The Effect of Subway Access on Distance Traveled to School With Full Covariates

	Distance	Log distance	Rank-Distance
	(1)	(2)	(3)
Access to subway	609.691***	0.200***	27.118***
	(159.826)	(0.033)	(8.908)
Year after 2006 dummy	-903.823***	-0.221***	-45.090***
	(138.642)	(0.029)	(7.728)
Treatment dummy	-341.590***	-0.072***	-25.142***
	(122.206)	(0.025)	(6.811)
Male student	442.328***	0.066***	25.813***
	(69.304)	(0.014)	(3.863)
Family works	-55.022	0.002	-4.245*
	(45.670)	(0.010)	(2.546)
Schools close (1km)	-131.383***	-0.055***	-5.027***
	(10.813)	(0.002)	(0.603)
Schools close (5km)	-7.169***	-0.000	0.755***
	(1.350)	(0.000)	(0.075)
Public	3921.191***	0.679***	229.027***
	(100.525)	(0.021)	(5.603)
Voucher	229.078***	-0.002	13.058***
	(86.428)	(0.018)	(4.817)
Constant	6538.790***	8.353***	104.177***
	(204.759)	(0.043)	(11.413)
Observations	22,046	22,046	22,046
R^2	0.103	0.098	0.108

Note: This table displays the coefficients of all covariates included when estimating equation 1.1. Each column presents the results of a separate regression, one for each outcome variable.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

Table 1.10: Regression Results (DID): The Effect of Subway Access on School Test Scores

	Distance		Log distance		Rank-Distance	
	(1)	(2)	(3)	(4)	(5)	(6)
School quality	12.129**	12.165***	0.002**	0.002**	0.441	0.441*
	(4.862)	(4.575)	(0.001)	(0.001)	(0.271)	(0.254)
Observations	22,044	22,044	22,044	22,044	22,044	22,044
R^2	0.012	0.125	0.007	0.107	0.013	0.135
Controls	No	Yes	No	Yes	No	Yes

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Columns (2), (4) and (6) present results controlling for the following variables: gender of the student, number of family members that are working, number of schools in a 1-km radius from the student's home, number of schools in a 5-km radius from the student's home, and dummy variables for the type of school.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

Table 1.11: Regression Results (DID): The Effect of Subway Access on School Type

	Public	Voucher	Private
	(1)	(2)	(3)
Access to subway	-0.008	0.021	-0.012
	(0.013)	(0.015)	(0.013)
Observations	22,948	22,948	22,948
R^2	0.000	0.006	0.006

Note: Each set of columns presents the results of a separate regression, using each type of school as the outcome variable. The estimation does not include control variables, but results do no change when they are included.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

Table 1.12: Regression Results (DID): The Effect of Subway Access on Distance Traveled to School by School Type

	Distance			Log Distance			Rank-Distance		
	(1) Public	(2) Voucher	(3) Private	(4) Public	(5) Voucher	(6) Private	(7) Public	(8) Voucher	(9) Private
Access to subway	43.5	760.2***	821.1***	0.0	0.2***	0.2***	-6.0	39.0***	33.8**
	(426.5)	(205.7)	(267.3)	(0.1)	(0.0)	(0.1)	(24.7)	(10.9)	(15.0)
Observations	5,323	11,573	5,150	5,323	11,573	5,150	5,323	11,573	5,150
R^2	0.006	0.003	0.010	0.005	0.003	0.016	0.005	0.003	0.001

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Within each set, separate results are presented for each type of school: public, voucher or private. The estimation does not include control variables, but results do no change when they are included.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

Table 1.13: Regression Results (DID): The Effect of Subway Access on PSU Math Score

	Low	Medium	High
	(1)	(2)	(3)
Access to subway	0.020	0.004	-0.024*
	(0.014)	(0.015)	(0.015)
Observations	22,948	22,948	22,948
R^2	0.004	0.002	0.001

Note: Each set of columns presents the results of a separate regression, using performance in the PSU math test as the outcome variable, classified as either low, medium or high. The estimation does not include control variables, but results do no change when they are included.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

Table 1.14: Regression Results (DID): The Effect of Subway Access on Distance Traveled to School Using Various Treatment and Control Group Definitions

	Distance		Distance		Distance	
	(1)	(2)	(3)	(4)	(5)	(6)
Access to subway 1.5 km	414.4***	429.5***				
	(155.9)	(148.0)				
Access to subway 2 km			450.8***	484.6***		
			(143.4)	(136.2)		
Access to subway 3 km					663.5***	767.4***
					(181.4)	(172.2)
Observations	18,952	18,952	23,344	23,344	23,344	23,344
R^2	0.004	0.103	0.005	0.103	0.004	0.103

Note: Each set of columns presents the results of a separate regression, using different definitions of the treatment and control groups, with distance from home to school as the outcome variable. Columns (2), (4) and (6) present results controlling for the following variables: gender of the student, number of family members that are working, number of schools in a 1-km radius from the student's home, number of schools in a 5-km radius from the student's home, and dummy variables for the type of school.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

Table 1.15: Regression Results (DID): The Effect of Subway Access on Distance Traveled to School Using Data Before the Subway Line Opened

	Distance		Log distance		Rank-Distance	
	(1)	(2)	(3)	(4)	(5)	(6)
Access to subway	144.956	92.504	0.021	0.007	4.321	3.283
	(176.365)	(164.424)	(0.037)	(0.034)	(9.986)	(9.299)
Observations	21,208	21,208	21,208	21,208	21,208	21,208
R^2	0.003	0.134	0.005	0.135	0.000	0.134
	(1)	(2)	(3)	(4)	(5)	(6)
Controls	No	Yes	No	Yes	No	Yes

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Only the years before the subway line was constructed are considered. Columns (2), (4) and (6) present results controlling for the following variables: gender of the student, number of family members that are working, number of schools in a 1-km radius from the student's home, number of schools in a 5-km radius from the student's home, and dummy variables for the type of school.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

Table 1.16: Regression Results (DID): The Effect of Subway Access on Distance Traveled to School (Placebo Line 1)

	Distance		Log distance		Rank-Distance	
	(1)	(2)	(3)	(4)	(5)	(6)
Access to subway	-74.009	-86.941	-0.031	-0.029	-1.688	-1.286
	(126.510)	(122.160)	(0.029)	(0.028)	(7.746)	(7.573)
Observations	22,185	22,185	22,185	22,185	22,185	22,185
R^2	0.000	0.068	0.000	0.083	0.000	0.045
Controls	No	Yes	No	Yes	No	Yes

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Columns (2), (4) and (6) present results controlling for the following variables: gender of the student, number of family members that are working, number of schools in a 1-km radius from the student's home, number of schools in a 5-km radius from the student's home, and dummy variables for the type of school.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

Table 1.17: Regression Results (DID): The Effect of Subway Access on Distance Traveled to School (Placebo Line 2)

	Distance		Log distance		Rank-Distance	
	(1)	(2)	(3)	(4)	(5)	(6)
Access to subway	138.523	177.367	0.036	0.040	0.590	1.799
	(173.687)	(168.809)	(0.041)	(0.039)	(11.469)	(11.257)
Observations	11,630	11,630	11,630	11,630	11,630	11,630
R^2	0.033	0.089	0.033	0.103	0.020	0.058
Controls	No	Yes	No	Yes	No	Yes

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Columns (2), (4) and (6) present results controlling for the following variables: gender of the student, number of family members that are working, number of schools in a 1-km radius from the student's home, number of schools in a 5-km radius from the student's home, and dummy variables for the type of school.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

Table 1.18: Regression Results (DID): The Effect of Subway Access on Distance Traveled to School (Placebo Line 3)

	Distance		Log distance		Rank-Distance	
	(1)	(2)	(3)	(4)	(5)	(6)
Access to subway	223.712	174.774	0.089	0.098	1.965	2.693
	(392.237)	(375.350)	(0.098)	(0.093)	(25.498)	(25.039)
Observations	4,732	4,732	4,732	4,732	4,732	4,732
R^2	0.004	0.091	0.004	0.103	0.000	0.039
Controls	No	Yes	No	Yes	No	Yes

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Columns (2), (4) and (6) present results controlling for the following variables: gender of the student, number of family members that are working, number of schools in a 1-km radius from the student's home, number of schools in a 5-km radius from the student's home, and dummy variables for the type of school.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

CHAPTER 2

THE EFFECT OF CELL PHONE USE ON TRAFFIC ACCIDENTS

2.1 Introduction

In Santiago, the capital of Chile, about 25,000 traffic accidents are routinely recorded in police reports every year. Cell phone use has long been thought to be a significant cause of traffic accidents (GHSA, 2011). Because of this, public resources—money for awareness campaigns and to pay for the time of police personnel—are allocated to curbing cell phone use by drivers. As stated by the National Safety Council (2013): "Currently there is no reliable method to accurately determine how many crashes involve cell phone use; therefore, it is impossible to know the true scope of the problem" (NSC, 2013, pg. 3). This paper aims to provide a causal estimate of the effect of cell phone use on traffic accidents.

Police reports, which are the main source of traffic accident information, do not generally record the use of cell phones, and drivers have incentives to underreport this activity, considering it is usually illegal (McCartt et al., 2006). Accidents caused by cell phone use tend to be, although not always, classified simply as "distracted driving" or "reckless driving" accidents. However, this category also includes many other causes of accidents. Roughly 50% of traffic accidents in Santiago are attributed to distracted driving. In the US, it is estimated that between 15% and 30% of accidents are caused by driver distraction (GHSA, 2011).

In contrast, other causes of accidents that have received much more attention both from the public and lawmakers, such as drunk driving, are linked to only about 5% of all traffic accidents. In Chile, laws have been passed regulating driver behavior and car safety features in many areas, such as the use of seat belts, installation of air bags, tougher driving exams, and banning of alcohol consumption, to name a few. For example, Chile recently approved a zero-tolerance law against drinking and driving with severe jail penalties, while cell phone use is punished only with a small fine.

According to US data, the frequency with which individuals use their cell-phones while driving has been high and increasing for most of the last decade (Nikolaev et al., 2010). A 2015 survey revealed that 68% of teens in the US admitted to using applications on their mobile phones while driving, and despite being the focus of several public awareness campaigns, 27% of teens admit to texting and driving (Liberty Mutual Insurance, 2016).

Additionally, direct observation studies seem to confirm the high frequency of cell phone use by drivers. The National Highway Traffic Safety Administration conducts a yearly study in which 50,000 vehicles stopped at intersections are observed. They estimate that about 8% of all drivers in the US were using either a handheld or hands-free phone in a typical daylight moment in 2014 (NHTSA, 2015).

In Chile, as well as in the US, laws that attempt to curb cell phone use while driving have been implemented, but it is difficult to evaluate those measures or find the most effective ones if the magnitude of the effect is not really known (NSC, 2013). Knowing the extent of the problem is important because, as stated by the National Safety Council, traffic accident data can influence policymakers in several ways, such as how accident prevention campaigns, funding, legislation, and vehicle and road design are determined (NSC, 2013).

The causal effect of cell phones on traffic accidents has been particularly difficult to estimate, and the question of whether using a cell phone while driving increases the risk of having an accident has not been definitively answered (Nikolaev et al., 2010). Most of the evidence has its origins on either self-report surveys or studies conducted in driving simulators, and both of these sources may have problems: in the first case, drivers have incentives to underreport illegal conduct, and the extent to which simulators actually reproduce real driving conditions is unknown (McCartt et al., 2006).

An important section of the literature regarding this topic corresponds to "laboratory" studies, which have generally found that driving while using a cell phone impairs driving performance (McCartt et al., 2006). Strayer et al. (2006) used a driving simulator to compare drivers using a cell phone with drunk drivers in a controlled setting. They found that drivers talking on their cell phones braked 9% slower and took 19% longer to recover their normal speed after braking compared to baseline drivers. They found similar results for drunk drivers.

Dingus et al. (2006) used data generated by videotaping drivers in their cars and found that the risk of an accident was 1.3 times higher when drivers were talking on their cell phones. They also found that the risk of having an accident was almost three times higher when drivers were dialing their cell phones.

Other studies have relied on police reports and cell phone company data. Out of these, one of the most frequently cited findings is that of Redelmeier and Tibshirani (1997) , who surveyed drivers who were involved in traffic accidents in the years 1994 and 1995, and examined their cell phone records to see if they were talking on their phones at the time of their accidents. The authors found that the use of cell-phones was associated with a 4.3 times increase in the likelihood of having a traffic accident, compared to the same drivers not using their phones. However, as they point out, the non-experimental nature of this study makes it difficult to rule out other factors, such as the possibility that drivers who are emotionally distressed could be more likely to talk on their phones, but are also more likely to have decreased driving ability at the same time.

Overcoming these problems caused by confounding variables requires finding sources of exogenous variation that allow for the estimation of the causal effect of cell phones on traffic accidents. The combination of exogenous variation and available accident data has proven difficult to find.

Bhargava and Pathania (2013) estimated the causal link between cell phones and traffic accidents by taking advantage of a discontinuity in cell phone plan pricing. They found no evidence that an increase in call volume, associated with a decrease in the cost of cell phone calls, resulted in a higher number of traffic accidents. Their study constituted the first, and, to the best of my knowledge, only credible attempt at estimating this causal effect, but it is possible to argue that cell phone usage has changed substantially since their study was published. Their estimation uses data from 2002 to 2005, before smartphones were widely adopted (for example, Apple's iPhone was introduced in 2007). At the time, the main concern was that drivers could be distracted by talking on their phones. Currently, cell phone usage often involves much more direct interaction with the phone, which means longer and more intense periods of distracted driving.

In this paper, I employ two alternative estimation strategies to attempt to calculate the causal impact of cell phone use on traffic accidents. For both methods, I take advantage of a natural experiment generated by a government program implemented in Chile, in which the government constructed cell phone towers in areas previously not connected to the mobile network. Because GIS software has limited geographic accuracy in remote regions of Chile, I focus on the administrative region—called Metropolitan Region—in which the capital city of Santiago is located.

The first estimation strategy consists of a difference-in-difference specification, in which I take advantage of the fact that the government program constructed cell phone towers in separate phases for different groups of localities. This allows me to compare areas where the main difference is the stage of the program in which they received cell phone coverage.

For the second estimation strategy, I use GIS to create a control group by generating and selecting random geographic areas in which there was consistent cell phone service throughout the period being studied. I then use this control group to estimate a difference-in-difference specification, comparing the change in the number of accidents in areas affected by the government program with random areas that already had cell phone coverage.

Both methods lead to similar conclusions: the availability of cell phone coverage results in an almost 30% increase in the number of traffic accidents. More specifically, using the phased-in implementation of the program, I find a 30% increase in traffic accidents when cell phone service becomes available. In concordance with this, using the alternative random control polygons method, I find a 31% increase in the number of accidents.

Section II provides additional background on the links between cell phone use and distracted driving, and on the government program. Section III describes the data set and geographic calculations used in this paper. Section IV introduces the two alternative estimation strategies. Section V presents estimates of the effect of cell phone use on traffic accidents and discusses the results. Section VI concludes the paper.

2.2 Background

2.2.1 Cell Phones and Distracted Driving

To understand why there is no good official data on accidents caused by cell phone use, it is important to observe the limitations in the way accidents are classified by police.

The cause of a traffic accident is not always clear. In a collision involving multiple cars, for example, versions may not coincide. In a single-car accident, such as a car crashing into a stationary object, the driver may have incentives to hide the actual cause of the accident if it involved illegal behavior. Regarding cell phone use, the National Safety Council (2013) concludes: "Police must often rely on drivers to admit to cell phone use. This is not possible when drivers are not forthcoming or are seriously injured or deceased"(NSC, 2013, pg. 3).

Despite these limitations, the best source for traffic accident information is still official accident reports. These reports classify accidents into several categories according to the cause, as determined by the police. The reports also register factors such as the age and gender of the driver and some of the conditions under which the accident took place, such as the time of day. Some of these statistics for the period between 2007 and 2014 in the Metropolitan Region are shown in Table 2.1.

As Table 2.1 shows, the characteristics of accidents remain stable over time. Drivers involved in accidents are, on average, about 40 years old, and almost 80% of them are male (73% of driver's license holders in Chile are men) (Registro Civil, 2011). Almost one-third of accidents happen during morning hours, while 15% happen between 6 p.m. and midnight.

Table 2.1: Descriptive statistics: 2007-2014 Metropolitan Region

	Year							
	2007	2008	2009	2010	2011	2012	2013	2014
Age (mean)	38.68	38.73	39.31	39.65	39.57	39.72	39.70	39.79
Female (pct.)	22.48	22.35	22.46	22.60	22.18	22.95	22.16	23.40
Morning (pct.)	29.28	28.45	27.65	27.94	27.83	29.22	29.05	28.84
Night (pct.)	15.56	15.74	15.71	15.96	16.34	15.96	15.71	15.78
Day of week (0-7)	3.158	3.101	3.118	3.150	3.137	3.120	3.096	3.122
Hour of day (0-24)	13.89	13.94	13.94	13.96	13.95	14.03	13.95	13.94
Total accidents	22499	23205	21680	21675	23520	23240	25216	25892

Source: Data from Ministerio de Transportes y Telecomunicaciones

In Chile, police classify traffic accidents into 14 main categories, which are further subdivided into more specific groups. The main categories are: reckless driving, reckless passenger, reckless pedestrian, not following signals, speeding, drunk driver, drunk passenger, drunk pedestrian, drugs or fatigue, lost control of car, mechanical issues, road issues, other causes, and undetermined causes.

The distribution of accidents by reported cause in the year 2014 in the Metropolitan Region is shown in Table 2.2.

Table 2.2: Number of accidents by cause in 2014

	Number	Percent
Disobeying road signs	2,444	9.44
Drugs or fatigue	107	0.41
Drunk driver	1,201	4.64
Drunk passenger	5	0.02
Drunk pedestrian	65	0.25
Lost control of vehicle	1,459	5.63
Mechanical issues	273	1.05
Other causes	3,421	13.21
Reckless driver	12,561	48.51
Reckless passenger	116	0.45
Reckless pedestrian	893	3.45
Road issues	140	0.54
Speeding	186	0.72
Undetermined causes	3,021	11.67
Total	25,892	100.00

Source: Data from Ministerio de Transportes y Telecomunicaciones

It is not obvious how an accident is classified when it is caused by multiple factors. For example, if a driver crashes while recklessly passing another vehicle at high speed, it will be up to the police officer filing the report to classify that accident as caused by either reckless driving or speeding.

As stated in a National Safety Council (2013) report on the underreporting of cell phone use as a factor in accidents: “this project’s review of fatal crashes uncovered cases where drivers using cell phones crossed over center lines resulting in head-on crashes, but the crash reports did not mention cell phone use. These omissions limit the usefulness of these data for prevention. There are many reasons why a driver could cross over the center line including: attempting to pass, reaching for something in the vehicle, experiencing a medical problem, alcohol or other drug impairment, as well as using a cell phone. Each of these root factors would likely be addressed by different prevention strategies” (NSC, 2013, pg. 4).

Police reports register the date and time of each accident. The number of accidents doesn’t vary greatly by day of the week, except for a marked decrease on Sundays, when there is arguably a smaller number of cars on the road. The distribution of accidents by day of the week in the year 2014 in the Metropolitan Region is shown in Table 2.3. During an average weekday, accidents are higher around peak commute hours, especially between 7 a.m. and 8 a.m. and between 6 p.m. and 8 p.m. An aggregated table of how accidents are distributed during the day is shown in Table 2.4.

Table 2.3: Number of accidents by day of the week in 2014

	Number	Percent
Sunday	2,814	10.87
Monday	3,796	14.66
Tuesday	3,943	15.23
Wednesday	3,822	14.76
Thursday	3,704	14.31
Friday	4,009	15.48
Saturday	3,804	14.69
Total	25,892	100.00

Source: Data from Ministerio de Transportes y Telecomunicaciones

Table 2.4: Number of accidents by time of day in 2014

	Number	Percent
12 am to 6 am	1,966	7.59
6 am to 12 pm	7,467	28.84
12 pm to 6 pm	8,644	33.38
6 pm to 12 am	7,815	30.18
Total	25,892	100.00

Source: Data from Ministerio de Transportes y Telecomunicaciones

These tables give a general description of traffic accidents in the Metropolitan Region. Looking at how these variables interact with each other, it is possible to see that certain types of accidents happen much more often at specific moments. For example, accidents classified as being caused by drunk driving predominantly occur after 9 p.m. on a Friday or Saturday. Other accidents, such as ones caused by disobeying road signs, seem to be spread out across days of the week and times of day.

While accidents caused by cell phones are not directly classified as such, they usually fall under the broader "distracted driving" category. The Governors Highway Safety Association (GHSA) distinguishes between four types of distracted driving: visual, auditory, manual, and cognitive (GHSA, 2011). These are defined as follows:

- Visual: looking at something other than the road
- Auditory: hearing something not related to driving
- Manual: manipulating something other than the wheel
- Cognitive: thinking about something other than driving

Following these definitions, it's possible to argue that using a smartphone while driving can create all four types of distraction, depending on the how the phone is used. A regular phone call would possibly be associated with auditory distractions, while use of applications in a smartphone could easily lead to all four types of distractions at the same time.

Several self-report surveys have addressed the issue of using cell phones or other electronic devices while driving in recent years. In 2014, the AAA Foundation for Traffic Safety (AAAFTS) conducted a nationally representative survey of drivers 16 years of age and older. The survey found that almost 70% of drivers admitted to having used their cell phones while driving within the past month, most of them on more than one occasion. One-third of drivers admitted to doing this fairly often or regularly (AAAFTS, 2015).

As mentioned above, the use of cell phone has expanded far beyond simply making phone calls since the introduction of smartphones. While the AAAFTS survey found that drivers seem to be aware of the distraction involved in typing text messages or emails while driving—96% of respondents consider it unacceptable behavior—surprisingly, more than a third of the same respondents admitted to reading text messages or emails while driving during the past month, and more than one in four drivers admitted to typing text messages or emails while driving. One-third of drivers between 25 and 39 years old report that they use the Internet while driving.

The Insurance Institute for Highway Safety (IIHS) surveyed drivers 18 years and older in 2010 and focused specifically on cell phone use. Asked about their most recent day driving, respondents had spent about 7% of their driving time talking on their cell phones (Braitman and McCartt, 2010).

Data from a self-report survey of 700 drivers conducted in the Metropolitan Region in Chile in 2015 reveals a similar pattern. Eighty-three percent of drivers admitted to frequently using their cell phone to write emails while driving, 80% frequently check Facebook, 68% frequently make phone calls, 62% send and read text messages, and 55% check the Internet (Automóvil Club, 2015).

2.2.2 Government Program: Todo Chile Comunicado

In 2010 the Chilean government began a program with the goal of providing rural communities with wireless services, as a way to address concerns of a perceived digital divide between urban and rural areas. At a total cost of US$110 million, mobile phone towers were constructed, and thus brought cell phone and Internet service to previously unconnected areas. Although 42 localities were intervened in the Metropolitan Region, due to geographic data limitations for some of the most remote rural areas, I only consider 14 of these areas in this paper (ENTEL, 2014).

The design phase of the program focused on identifying areas not covered by cell phone service providers. The majority of these areas corresponded with small, inhabited, rural communities that didn't have cell phone coverage because of the mountainous geography of the country, which blocks radio signals, and populations too low to justify the investment required to construct a cell tower site by private companies.

After areas with no coverage were identified, the program's focus was on the construction of cell phone sites, through a partnership with a telecommunications provider. The program was originally meant to be implemented in a staggered manner, with groups of areas being affected in three consecutive stages until the program's completion. In practice, stages 2 and 3 were implemented almost simultaneously, so I consider them to be the same stage for the estimation. Results do not change when they are considered as separate stages. Considering this, the first stage began in September 2010, and the second stage in the third quarter of 2012 (MTT, 2008).

Areas treated by the program at different stages seem to have similar characteristics in terms of the drivers involved in accidents inside those areas and the conditions when these accidents happened. The details of this are shown in Table 2.5.

Table 2.5: Descriptive statistics: phased-in localities

	Phase 1	***Phase 2***
Driver characteristics		
Age (mean)	36.56	34.54
Female (pct.)	23.79	13.74
Accident characteristics		
Morning (pct.)	24.83	18.30
Night (pct.)	18.95	27.69
Day of the week (0-7)	3.12	3.02
Hour of the day (0-24)	14.26	15.27
Observations	56	140

Note: Number of observations corresponds to localities by quarter

2.3 Data and Geocoding

Data used in this paper comes from two separate data sets. The main data source is the official traffic accidents statistics data set, compiled by Chile's National Traffic Safety Commission and the Ministry of Transportation and Telecommunications. This data set contains every traffic accident that that was reported to or registered by the police between 2007 and 2014. This means that certain types of accidents—for example, where a single driver crashes into private or public property and drives away—are not recorded in this dataset.

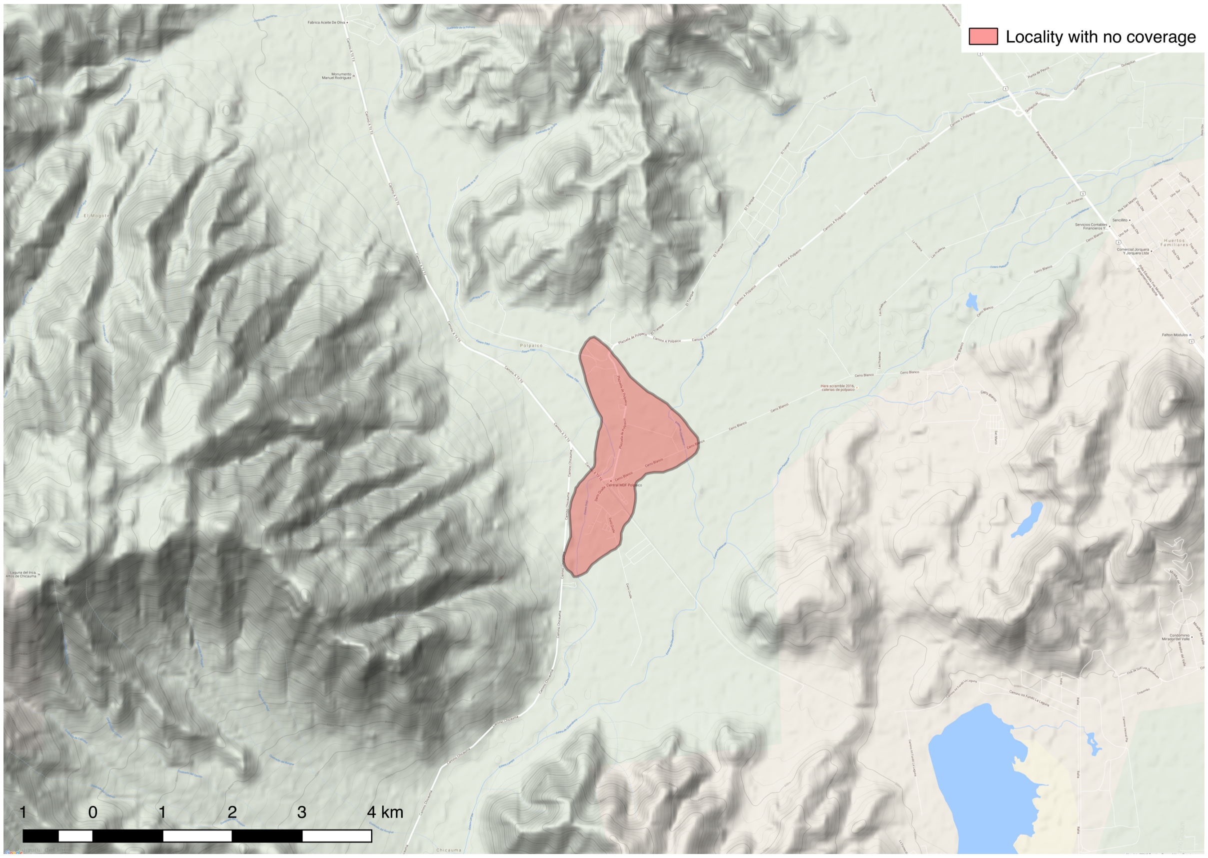
From this source I use locality-level aggregated data on the number of traffic accidents, cause of accidents, day and time of accidents, and age and gender of the drivers involved. I also use the exact address or the intersection where the accident occurred. To estimate the impact of cell phone use on traffic accidents, I consider two outcome variables: total accidents and log total accidents. I use the log (total accidents +1) to provide an easier interpretation of the effect of the government program.

The first identification strategy I employ relies on the phasing in of the government program in different localities. Data on the location of new cell phone sites constructed as part of the Todo Chile Comunicado program was provided by the Ministry of Transportation and Telecommunications, and includes the stage of the program in which a cell phone site was constructed and its location. It also includes the geographic information—as polygons—for the areas identified by the government as having no cell phone signal. An example area is illustrated in Figure 2.1.

2.3.1 Geocoding

I geocode every accident that happened inside the Metropolitan Region between 2007 and 2014. By calculating the intersection between the polygon identified as a locality

Figure 2.1: Locality identified as having no cell-phone coverage



Note: This figure shows the geographic extent of a locality that was treated by the program.

treated by the program and the geocoded accident, I identify accidents that occurred in the specific areas covered by the new cell phone towers. The accuracy of geographic data in rural Chile is very precise for certain localities, but not usable for others. For this paper I use only localities in which I am able to accurately determine the locations of accidents. This forces me to leave several localities out of the analysis. It is likely that more rural and remote localities are left out because of geographic data limitations, which must be considered when interpreting the results presented in this paper.

2.4 Estimation Strategy

I use two separate methods to estimate the impact of cell phones on traffic accidents. First, I use the program's phased-in implementation across localities in the Metropolitan Region. That is, I compare localities in which cell phone towers were constructed in an earlier stage of the program with control areas that were treated in a later stage, by estimating a differences-in-differences specification.

As an alternative approach, I use geographic information system (GIS) software to generate random areas that had cell phone coverage for the duration of the program, and compare the number of accidents inside those areas with the number of accidents inside areas treated by the program. In this way, I generate a random control group that should be unaffected by the construction of cell phone towers in other areas.

Since it is not possible to know if drivers inside the treated localities are actually using their cell phones, the estimation produces an intent-to-treat estimator. A more detailed description of each method follows.

2.4.1 Estimation Strategy: Phasing-In of the Program

Localities affected by the program were treated in two distinct phases. Localities treated by the program during stage 1 received cell phone coverage after the third quarter of 2010. Localities treated during stage 2 of the program did so after the third quarter of 2012. The telecommunications company in charge of the project decided how to assign localities to each stage. It is not possible to know how exactly this assignment took place, but it seems reasonable to assume that the decision was not related to the number of traffic accidents taking place in each locality. It is likely that larger areas were treated first, and smaller areas were left for the second stage.

It is thus possible to take advantage of this phasing-in of the program and use areas that were treated in the second stage of the program as controls for areas that were treated in the first stage. I exploit this phasing-in of the program by using a differences-in-differences specification to calculate the effect of the construction of cell phone sites on traffic accidents. For this I first estimate variations of the following equation:

$$\text{Accidents}_{lt} = \beta_0 + \gamma \text{CellPhone}_{lt} + \theta_l + \delta_{lt} + \epsilon_{lt} \tag{2.1}$$

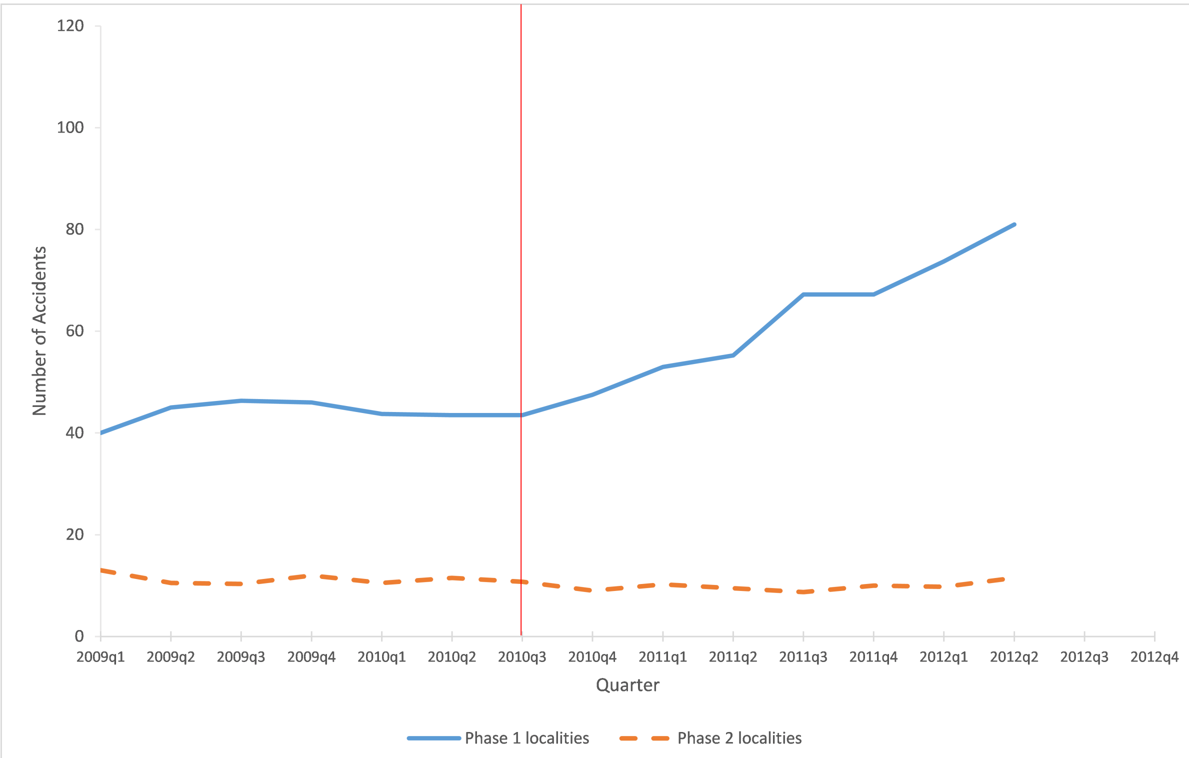
in which the outcome variable, $Accidents_{lt}$, are traffic accidents (total number, logs) in locality l in quarter t. $CellPhone_{lt}$ indicates that a cell-phone tower was constructed in locality l in quarter t; as such, it takes a value of 1 if locality l had cell phone service in quarter t, and a value of 0 otherwise. The coefficient γ captures the average effect of the treatment on the number of accidents in a locality.

I include locality fixed effects (ϑl) to attempt to capture any locality-level factors that remain constant over time, and that I may be unable to control for. Locality by quarter fixed effects (δlt) are included to account for shocks that may impact individual localities differently. In some specifications, instead of locality by quarter, I include quarter fixed effects (λt) to attempt to control for external shocks that may have happened and that may have affected all localities equally in each quarter. This does not alter the main results. Standard errors are clustered at the locality level.

This differences-in-differences specification relies on the assumption that localities would have followed similar trends if there had been no intervention. Even if it is not possible to test this assumption directly, in Figure 2.2 I show that the trends in the number of accidents in localities treated in phase 1 and phase 2 appears to be similar, until localities belonging to phase 1 receive cell phone coverage.

It is also assumed that the assignment of localities to a specific phase of the program was unrelated to the number of traffic accidents in a locality. Since the program was

Figure 2.2: Trend in number of accidents by phase



Note: This figure shows the number of accidents by quarter for Phase 1 and Phase 2 localities.

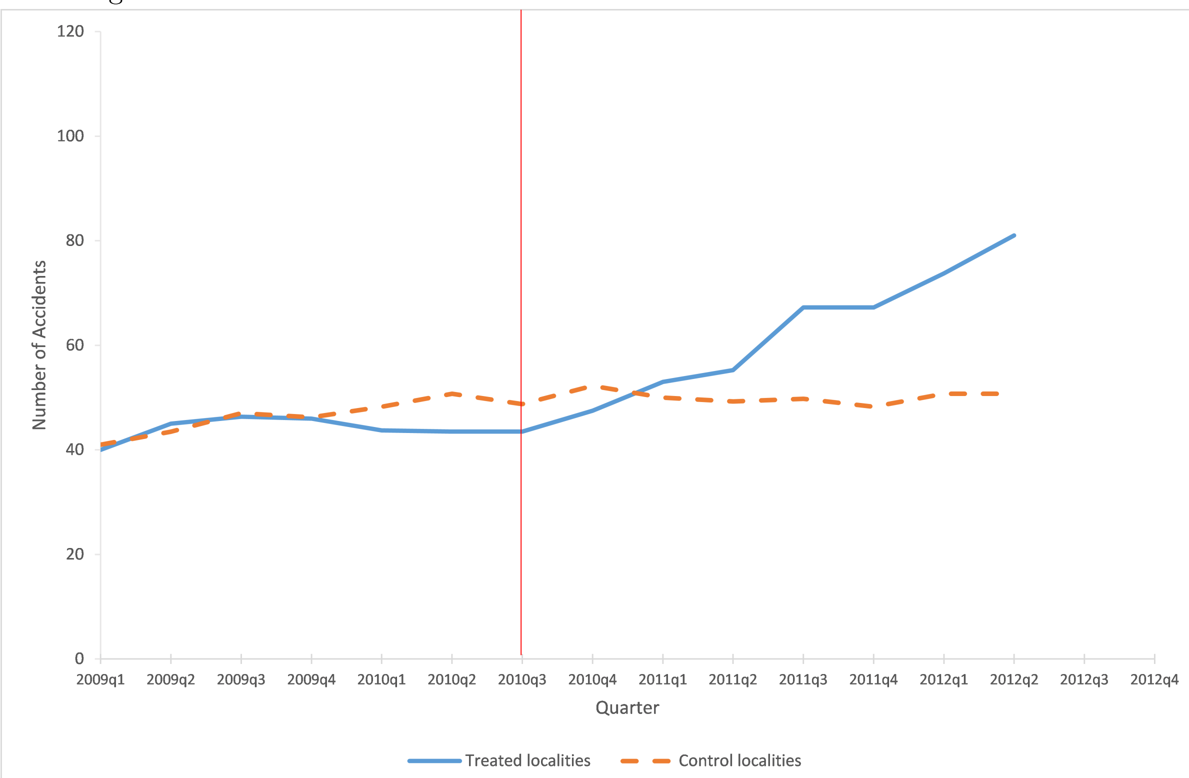
intended to address a perceived digital access gap between areas, it seems unlikely that the number of traffic accidents would have played a role in the assignment of localities to treatment phases.

2.4.2 Estimation Strategy: Random Control Polygons

As an alternative to the previous approach, I use GIS tools to generate random polygons that serve as control groups. For this, I consider all localities that were not affected by the government policy (since they already had cell phone service) and randomly select polygons inside those areas to serve as controls. As the treatment group, I only consider areas treated in phase 1 of the program, leaving out the period that corresponds to phase 2.

The intuition behind this approach is that the number of accidents inside areas that already had cell phone coverage should not be affected by the government program. That is, if for example cell phones did increase the number of accidents, accidents should increase in areas treated by the program but remain unchanged in these control areas that already had cell phone service. Figure 2.3 shows the trends in the number of accidents inside treated and control localities over time.

Figure 2.3: Trend in number of accidents in treated and random control localities



Note: This figure shows the number of accidents by quarter for treated and control localities.

Since the government program was mostly focused on small rural communities that didn't have cell phone service, it's highly likely that areas that already had cell-phone service were less rural and more developed. Because of this, it would not be unexpected to observe some differences in the characteristics of accidents that happen in the control and treated areas. However, areas treated by the program and areas selected randomly seem to have similar characteristics in terms of the drivers involved and the conditions under which these accidents happened. The details of this are shown in Table 2.6.

Table 2.6: Descriptive statistics: random control polygons

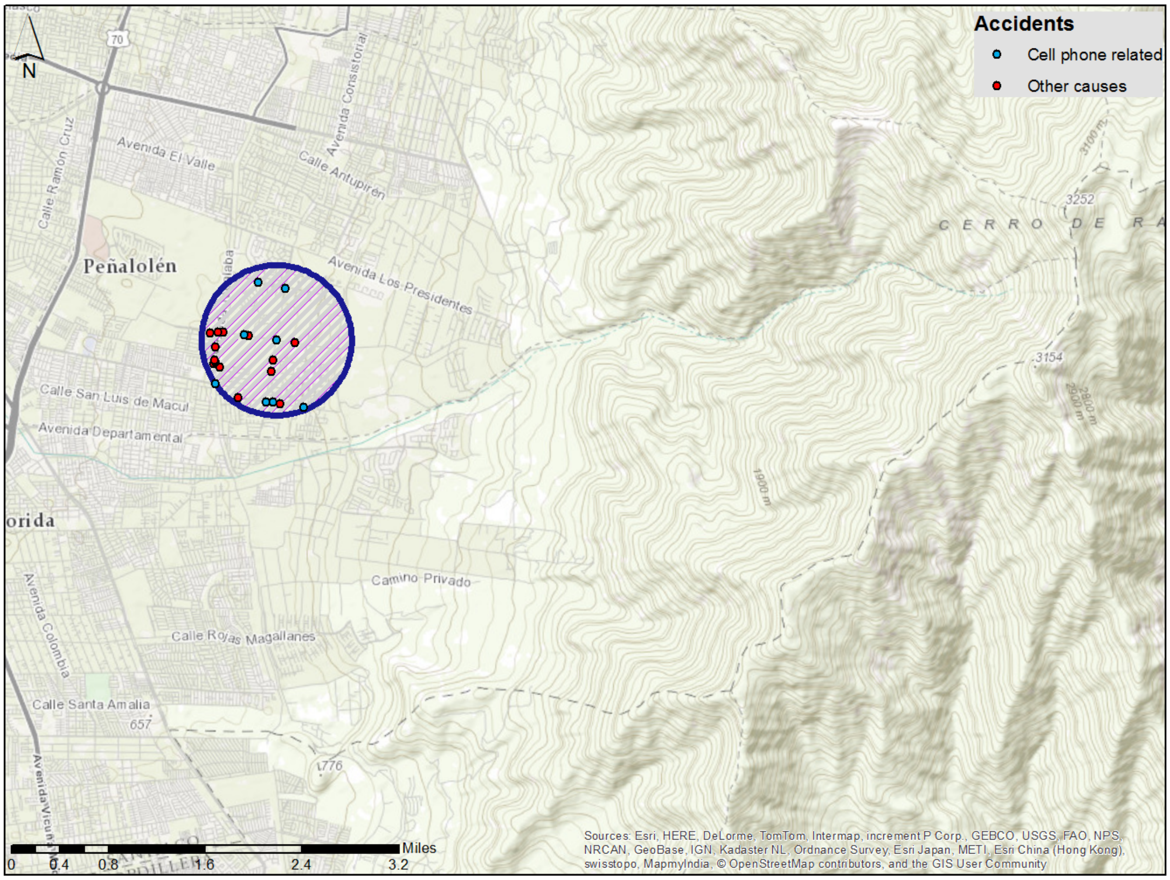
	Control	Treated
Driver characteristics		
Age (mean)	38.82	36.56
Female (pct.)	24.36	23.79
Accident characteristics		
Morning (pct.)	32.53	24.83
Night (pct.)	14.05	18.95
Day of the week (0-7)	3.31	3.12
Hour of the day (0-24)	13.26	14.26
Observations	504	56

Note: Number of observations corresponds to localities by quarter

In practice, I select roughly 30 random localities that always had cell phone coverage as a control group. To avoid selecting areas in which there are no roads, I restrict areas that may be selected to localities in which at least one accident has happened in a 1-kilometer radius. That is, I randomly generate 30 polygons of similar total area to the localities treated by the program, and then sum the number of accidents that happened inside those polygons during each quarter for the period being studied. An example of a randomly selected locality is illustrated in figure 2.4.

To ensure that the results are not dependent on a particular random selection of localities, I repeat this procedure 10,000 times, selecting about 30 random polygons each time. Finally, I calculate the average number of accidents inside these random localities. I then use these polygons to estimate a differences-in-differences model, using variations

Figure 2.4: Randomly selected polygon



Note: This figure shows the geographic extent of a random locality used as part of the control group.

of the following equation:

$$\text{Accidents}_{lt} = \beta_0 + \gamma\text{Treated}_{lt} + \theta_l + \delta_{lt} + \epsilon_{lt} \tag{2.2}$$

in which the outcome variable, $Accidents_{lt}$, are traffic accidents (total number, logs) in locality l in quarter t. $Treated_{lt}$ indicates that locality l was treated by the program in quarter t. The parameter γ captures the average effect on traffic accidents from a locality being treated by the program.

I include locality fixed effects (ϑl) to attempt to capture any locality-level factors that remain constant over time, and that I may be unable to control for. Finally, locality by quarter fixed effects (δlt) are included to account for shocks that may impact individual localities differently. Quarter fixed effects (λt) are included in some specifications to attempt to control for external shocks that may have happened and that may have affected all localities equally in each quarter. Standard errors are clustered at the locality level.

2.5 Results

In this section, I present results for each method separately. First, for the estimation that takes advantage of the phasing-in of the program, and then for the method using random polygons as a control group. Results for both estimations are similar and show that when a locality receives cell phone coverage, the number of accidents that happen inside that locality tends to increase.

2.5.1 Results: Phasing-In of the Program

This estimation takes advantage of the phasing-in of the program, comparing the number of accidents from the first phase with the number of accidents in the second phase. Results from estimating equation 2.1 are presented in Table 2.7. The results of estimating each specification, and for every outcome, are presented in separate columns.

Column (1) shows results of the estimation using log total accidents as the outcome variable, without including locality by quarter fixed effects.

Column (2) shows results using log total accidents as the outcome variable, including locality by quarter fixed effects. The treatment coefficient is positive and statistically significant in every specification, except the one in column (3).

Table 2.7: Regression results: phasing-in of the program

	Log Total Accidents		Total Accidents	
	(1)	(2)	(3)	(4)
Treated	0.297**	0.304**	6.493	6.626*
	(0.128)	(0.130)	(3.687)	(3.585)
Observations	196	196	196	196
Adjusted R^2	0.862	0.864	0.855	0.850
Locality-Quarter FE	No	Yes	No	Yes

Note: Columns (2) and (4) include locality by quarter fixed effects.
* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

The log coefficient of 0.297 presented in column (1) allows for a more intuitive interpretation, showing that introducing cell phone service to an area causes an increase of approximately 30% in the number of total accidents during a quarter in that area, compared to an area without cell-phone coverage. The positive and significant coefficient is robust to including locality by quarter fixed effects, as shown in column (2).

Column (3) shows the result of estimating equation 2.1 without including locality by quarter fixed effects and using total accidents as the outcome. Although the coefficient in column (3) is not statistically significant, its sign and magnitude are consistent with the previous outcome, showing that accidents increase when a locality receives cell phone coverage. Finally, in column (4) locality by quarter fixed effects are included and the outcome corresponds to total accidents. Again the coefficient is positive and indicates that accidents increase once cell phone towers are built in a locality.

These results can be more clearly seen in Figure 2.2, presented earlier. In the pre-program quarters, localities belonging to phase 1 and phase 2 follow relatively parallel trends, without noticeable changes in the number of accidents that happen each quarter. However, once phase 1 of the program is implemented, there is a sharp increase in the number of accidents in treated localities. In contrast, localities belonging to phase

2, which have not yet been treated during the period under consideration, maintain a constant level of accidents.

2.5.2 Results: Random Control Polygons

This estimation uses polygons generated randomly inside areas that had cell phone coverage before the program was implemented as controls for areas treated during the first phase of the government program.

Results from estimating equation 2.2 are presented in Table 2.8. This table follows the same format as Table 3.10.

Table 2.8: Regression results: random control polygons

	Log Total Accidents		Total Accidents	
	(1)	(2)	(3)	(4)
Treated	0.312**	0.343**	6.464*	6.283*
	(0.145)	(0.139)	(3.538)	(3.421)
Observations	560	560	560	560
Adjusted R^2	0.599	0.595	0.756	0.756
Locality-Quarter FE	No	Yes	No	Yes

Note: Columns (2) and (4) include locality by quarter fixed effects.
* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

Again, I focus on log total accidents as the outcome variable since it provides for a more intuitive interpretation. Column (1) shows results using log total accidents, without including locality by quarter fixed effects. This coefficient of 0.312 can be interpreted as showing that introducing cell phone service to an area causes an increase of approximately 31% in the number of total accidents during a quarter in that area, compared to an area without cell phone coverage. The positive and significant coefficient is robust to including quarter fixed effects, as shown in column (2).

Column (3) shows the result of estimating equation 2.2 without including locality by quarter fixed effects and using total accidents as the outcome. In column (4) locality by quarter fixed effects are included and the outcome corresponds to total accidents. This time both coefficients are positive and significant, indicating that the number of accidents increases once cell phone towers are built in a locality.

Graphically, the results can be more clearly seen in Figure 2.3, presented earlier. Before the first phase of the program is implemented, both the random control localities and localities that are yet to be treated follow relatively similar trends in terms of the number of accidents each quarter. But once cell phone sites are constructed, accidents consistently increase inside treated localities while remaining constant inside control localities.

2.6 Conclusion

In this paper, I provide evidence of the impact of cell phone use on traffic accidents, which has not been causally estimated under current cell phone usage conditions. I show that when a locality receives cell phone coverage, the number of accidents increases by about 30%.

These results are congruent with evidence from diverse sources. Drivers are frequently distracted, and one of the main sources of distraction seems to be cell phones. The 100-Car Naturalistic Driving Study (2006) showed that drivers were involved in a secondary task more than 50% of the time while they were driving. Olson et al. (2009) found a similar result: commercial drivers were distracted performing a secondary task about 55% of the time.

The data informing this analysis comes from traffic accidents in rural regions of Chile. Despite this, given reported cell phone usage patterns by drivers in both countries, it is possible to imagine that results for the US would not be dramatically different. Drivers in Chile admit to using their cell phones while driving more frequently than US drivers and using them for purposes that demand elevated amounts of attention. Therefore, it is possible that the effect in the US might be smaller than the one found in Chile, but it seems unlikely that cell phones have not caused an increase in traffic accidents in the US. as they have in Chile.

While cell phones generate more possibilities of distraction for drivers, it is also important to acknowledge that cell phones are precisely what allow drivers to call for help when they are in an accident. Information from cell phones also allows drivers to reduce commute times and pollution by finding more efficient routes to their destinations.

Even so, the impact of cell phones on traffic accidents seems to be high enough to merit larger public awareness campaigns and stricter laws that discourage cell phone use while driving. With more clarity about the magnitude of the problem should come more focus on finding its solution.

CHAPTER 3

THE EFFECT OF SUBWAY ACCESS ON EMPLOYMENT

3.1 Introduction

Subway construction accounts for some of the most costly transportation projects for cities across the world (Meyer et al., 1965). Despite this, remarkably little is known about the effects of building new subway lines in major cities. Most studies have focused mainly on how property prices react to the opening of new transit facilities, such as light rail or subway stations, but it seems reasonable to suggest that the effects of increased access to public transportation might be much broader.

Subways and urban rail connect people to each other, to firms, to services, and to amenities. It seems plausible that the opening of new subway stations could cause changes in the neighborhoods surrounding the stations. For example, firms located in an area in which a station opens might see an increased flow of people into the area, as those firms become more easily accessible for customers and workers in other parts of the city. Would the opening of the subway, then. cause changes to the number of workers in areas surrounding the stations? On the other hand, it is also possible that, in certain contexts, public transit may generate incentives for urban areas to remain underdeveloped (Glaeser, 2011).

In this paper, I analyze the effect of the 2005 opening of a new subway line in Santiago, Chile, on the number of employees working at firms located close to the subway stations. Additionally, I analyze the effects on firms related to real estate and construction, that could be more affected by the proximity of the subway line.

The construction of urban rail systems usually involves large and complex projects that can bring years of undesirable consequences for people and firms located near construction sites. The systems are expensive to build and incur high operating costs. Thus, authors such as Winston and Maheshri (2007) argue that urban rail systems are socially undesirable but continue to be built because they have unrestricted support from urban planners and policymakers.

But subway lines do offer some benefits to those nearby beyond improved transportation access. Most studies so far have looked at the impact of transportation on property values, usually finding a positive association between access to transportation and property values, although this seems to depend on the initial income level of the neighborhood (Bowes and Ihlanfeldt, 2001; Agostini and Palmucci, 2005; Debrezion et al., 2007; Hess and Almeida, 2007; Duncan, 2011; Hewitt and Hewitt, 2012; Yan et al., 2012).

The two main exceptions to this general focus are Schuetz's (2014) study, which examines the effects of rail stations on neighborhood retail activity in California, and the study by Bollinger and Ihlanfeldt (1997), which examines the impact of rapid rail transit in Atlanta on population density and employment.

Schuetz's (2014) work most directly addresses the effect of rail transit stations on employment; this study examined the effect of investments in rail transportation on the quantity of retail offerings in neighborhoods surrounding the stations, in the setting of a rail expansion in California's four largest metropolitan areas—Los Angeles, Sacramento, San Diego, and San Francisco-San Jose. The author compared areas surrounding new rail stations with areas around older stations to determine whether employment in retail firms increased around the new stations. Results showed that considering all cities and all neighborhoods, the opening of new rail stations was not associated with a significant change in retail employment.

Bollinger and Ihlanfeldt (1997) also examine the relationship between transportation and employment, reporting on the effects of Atlanta's rail service (MARTA) on population density and employment in areas surrounding the stations. The authors found that the rail stations "had neither a positive nor a negative impact on total population and total employment in station areas, and MARTA has altered the composition of employment in favor of the public sector, but only in those areas with high levels of commercial activity" (Bollinger and Ihlanfeldt, 1997, pg. 180).

Other studies look at the spatial aspects of employment more generally, examining the importance of access in employment outcomes. Since employment requires matching workers with firms that need them, geography and access to transportation play a relevant role in employment patterns. For instance, O'Regan and Quigley (1996) found support for the idea that access to jobs plays an important role in employment, especially for minority youth.

Stoll et al. (2000) examining the spatial distributions of jobs and population across several North American cities, found that less-educated and low-income people mostly live in areas in which the availability of less-skilled jobs is low, meaning that they must have transportation in order to sustain employment. An important part of the literature on the spatial distribution of jobs has focused on minority populations' access to employment opportunities; for instance, Holzer et al. (2003) examined the effects of an expansion of the San Francisco Bay Area's heavy rail system on minority employment, by conducting two waves of a firm survey. The authors found that firms closer to the rail stations hired significantly more Latino workers, but found no changes to overall employment levels in the region.

The lack of studies of the effect of transportation on other outcomes may be partly explained by a lack of data detailed enough to allow study of those outcomes in combination with essential spatial information. In this paper, which follows Schuetz's (2014) leading work, I use a detailed administrative panel dataset containing information on all firms operating in Chile, including accurate geographic information about firm locations.

To test whether the opening of subway stations has an effect on employment in the areas surrounding new stations, I take advantage of a natural experiment generated by the opening of a new subway line in Santiago, Chile, in late 2005. The inauguration of this subway line and the use of GIS tools, alongside a rich administrative panel dataset with detailed geographic information, allows me to estimate the effect of improved access to transportation infrastructure on employment.

This estimation strategy is based on the hypothesis that improved transportation accessibility should increase the desirability of areas near new facilities for firms, since those areas are now accessible to a larger number of customers and better-qualified workers (Holl, 2004). This enhanced desirability could be reflected in a higher number of employees, if firms located near the subway grow as a result of the new transportation infrastructure or if new firms decide to locate in the area. In other words, if firms benefit from, or expect to benefit from, a positive effect of the opening of a subway station, we would expect to see an increase in the number of workers in firms close to the station compared to those located farther away.

Using a difference-in-difference estimation, I find that firms located closer to the subway station didn't have significantly more workers than firms located farther away. This is true across most types of firms, with the exception of real estate firms located close to the subway, which hired more workers than those firms with less convenient access to the subway network.

This paper differs from previous studies in several ways. First, I examine the impact of building a new subway line in the context of a major city in a developing country; most studies so far have focused on transit construction in the United States and Europe. Second, I focus on a subway line built across several already-developed residential neighborhoods and not in a business district. And finally, I examine outcomes for firms that perform activities related to real estate and construction, which could be in theory more affected by changes caused in surrounding areas by the opening of the subway line.

The paper is structured as follows. Section II provides additional background on the links between access to transportation and employment. Section III describes the dataset and geographic calculations used in this paper. Section IV introduces the estimation strategy. Section V presents estimates of the effect of access to the subway on employment and discusses the results. Section VI concludes the paper.

3.2 Background

The following subsections briefly discuss the relationship between firm location and transportation, provide descriptive statistics for these firms, and present an overview of the subway network and its expansion.

3.2.1 Firm location and transportation

Although location might be more important for a retail firm than, for example, a firm that sells online advertising, location is important for all firms in terms of the location's attractiveness to potential employees. Researchers in urban economics have generated several models that attempt to explain the determinants of firm location, but "the role of transport investment in influencing the location of economic activity is an important subject that has received much less attention" (Holl, 2004).

Distance to public transportation may be a relevant variable for firms in deciding on a location, as convenient transportation will yield access to a larger set of consumers. It is also possible that firms may benefit from the influx of a larger number of customers who may visit the neighborhood for other purposes. Finally, improved transportation connectivity may allow better matches between workers and firms. All of these factors together could result in increases in the number of firms and the number of workers in areas around subway stations. But it is also important to consider that "an increase in retail establishments near a newly built train stop could represent either a net increase in retail activity through new store creation or redistribution from other, less accessible sites, as stores relocate closer to the rail station" (Schuetz, 2014).

3.2.2 Descriptive statistics for firms in Santiago

Since this paper focuses on areas surrounding the subway stations opened as part of the line inaugurated in 2005, I only consider firms located in the nine municipalities of Santiago through which the subway line crosses. For these areas, I generate a yearly panel of firms between the years 2005 and 2009.

Table 3.1 shows the number of firms in the nine selected municipalities per year. As this table shows, the total number of firms in this area gradually increases over the period of study, from 55,591 in 2005 to 88,950 in 2009.

Table 3.1: Number of Firms in Selected Areas: 2005-2009

	Number	Percent
2005	55,591	15.04
2006	63,815	17.26
2007	76,820	20.78
2008	84,445	22.85
2009	88,950	24.07
Total	369,621	100.00

Source: Data from Servicio de Impuestos Internos

Table 3.2 shows the average number of workers per firm each year. Overall, the number of workers remains relatively stable over time during the 2005–2009 period, after an initial increase between 2005 and 2006, from 47.6 to 51.8 workers per firm. For the remaining years in the study, the number of workers per firm varies between 55 and 51.

The standard deviation is large, as many firms are composed of only one worker and there are a few firms with thousands of workers each. Removing these large firms from the data used in the estimations does not significantly alter the results.

Table 3.2: Average and Total Number of Employees by Firms in Selected Areas: 2005-2009

	Mean	N
2005	47.6	1,045,537
2006	51.8	1,328,743
2007	53.9	1,686,266
2008	55.4	1,888,473
2009	51.0	1,791,497

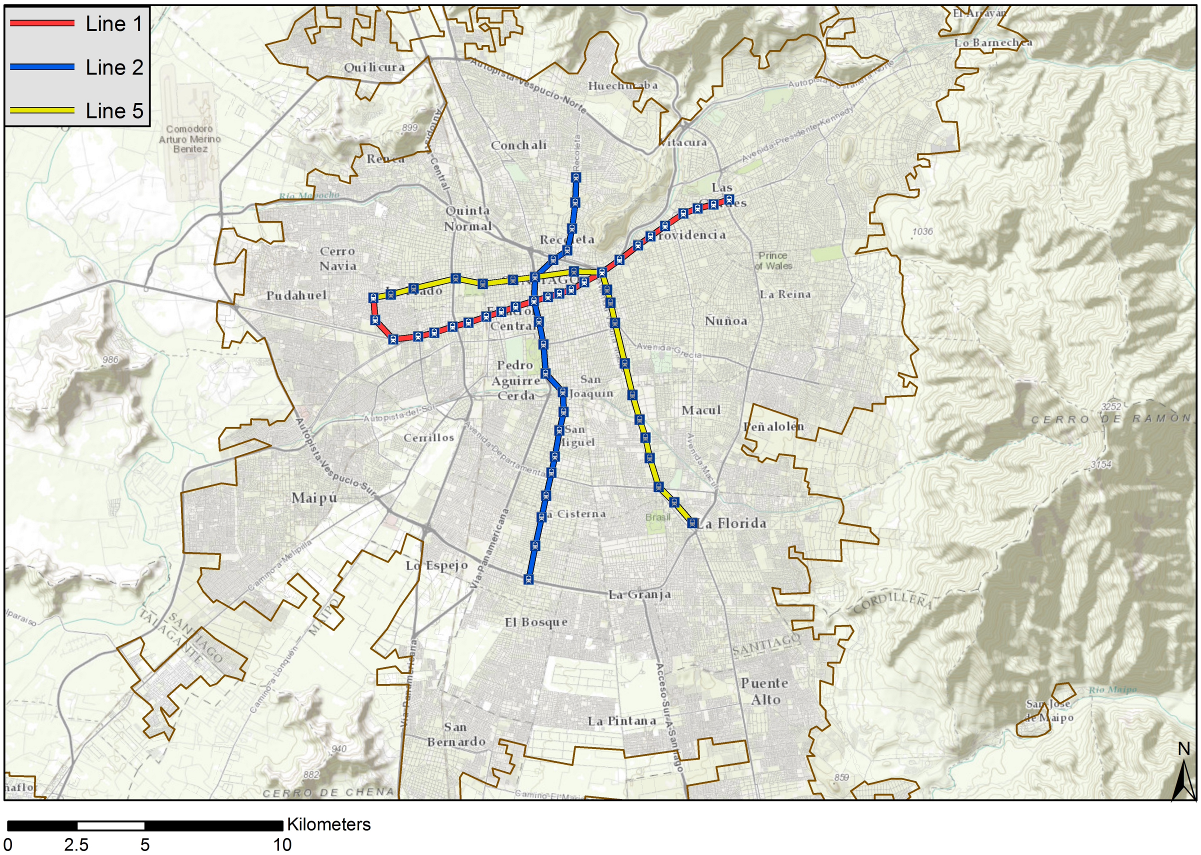
Source: Data from Servicio de Impuestos Internos

3.2.3 The Santiago subway network

Understanding the importance of access to transportation in Santiago requires a brief overview of the subway network. Santiago is a city with a population of more than 6 million people and a complex transportation network, of which the subway is a key element—in 2013, more than 60% of trips on public transport were subway trips (Metro de Santiago, 2013). Construction of the subway network began in 1969, and the first line (Line 1) was inaugurated in 1975. The second line (Line 2) opened in 1978, followed by the third line (Line 5) in 1997. Figure 3.1 shows what the network looked like in the year 2004, before further line construction took place.

After a long delay in expansion, the fourth line (Line 4) started operating between November 2005 and March 2006. Line 4 is the longest line in the network, covering 24 km with 23 stations and extending across seven large municipalities. This line, which connects to the other three lines in the network, carries around 500,000 people each day

Figure 3.1: Subway Network in 2004 Before the Inauguration of Line 4



Note: I generated this map by geocoding the exact location of each subway station that existed at the beginning of 2004, as informed by Metro de Santiago. The subway network has since been expanded in recent years with both the inauguration of new lines, and the extension of existing lines.

(Metro de Santiago, 2013; De Grange, 2010).

As I show in Table 3.3, yearly ridership of Santiago's subway network in Santiago increased during the period 2005–2009. Line 4 was inaugurated at the end of 2005, but some stations were not fully functional until later in 2006. This is reflected in the lower numbers of passengers in those years. By 2007, the line reached its maximum operating capacity of almost 115 million passengers per year. Because of this, 2007 can be considered as the first year that the subway is in full effect. In 2016, the network as a whole transported more than 2 million passengers per day.

Table 3.3: Number of Subway Users Per Year Between 2005 and 2009

	2005	2006	2007	2008	2009
Line 1	158,254,000	167,192,000	256,036,992	272,104,000	258,174,000
Line 2	51,839,000	58,893,000	120,468,000	128,830,000	121,836,000
Line 4	2,692,000	48,419,000	114,008,000	123,921,000	117,509,000
Line 5	54,317,000	53,213,000	89,385,000	95,885,000	90,735,000

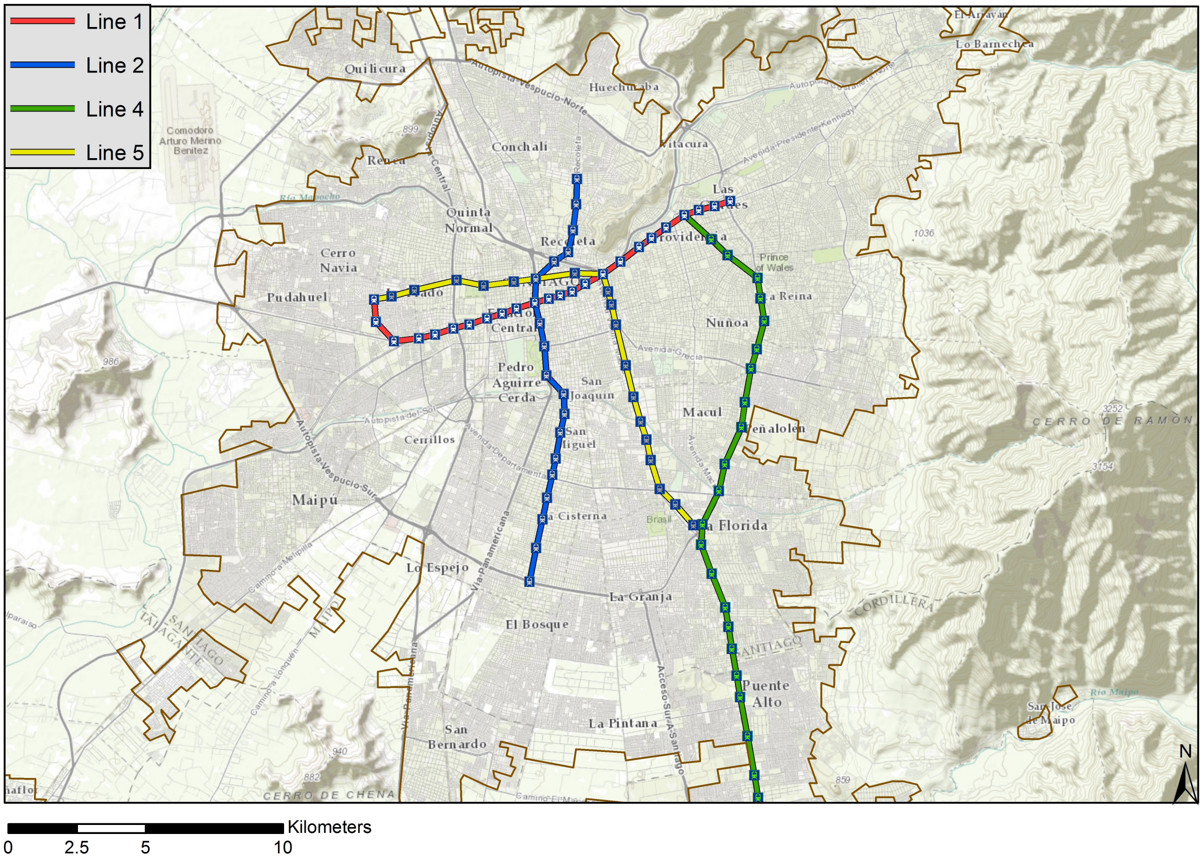
Note: Line 4 was inaugurated during the last months of 2005 and was fully operational in 2007. Data comes from Metro de Santiago.

Before the addition of Line 4 in 2005–2006, a vast area that included some of the most populated neighborhoods in the city was poorly connected to the transportation network. Figure 3.2 shows what the subway network looked like after the inauguration of Line 4 at the end of 2005. Line 4 stretches toward the south of Santiago, connecting a large area of the city that previously suffered from poor connectivity to the rest of the network and greatly reducing travel times to and from that area. Using the Google Distance Matrix API, I performed several travel simulations of bus and subway trips between this southern area of the city and the northern end of Line 4; travel time on average was reduced by about 50% when traveling by subway rather than by bus—the only public transportation available in that area before the opening of Line 4.

3.3 Data, Variables, Geocoding and Distance Calculations

To study the impact of the subway line on employment, I use data on subway station addresses and yearly administrative panel data that contains information on firms' addresses. Each of these addresses is geocoded to identify coordinates for its exact location. Firms are then identified by their location relative to the subway line, to separate those within the area of influence of the subway from those located farther away. The details

Figure 3.2: Subway Network in 2006 After the Inauguration of Line 4



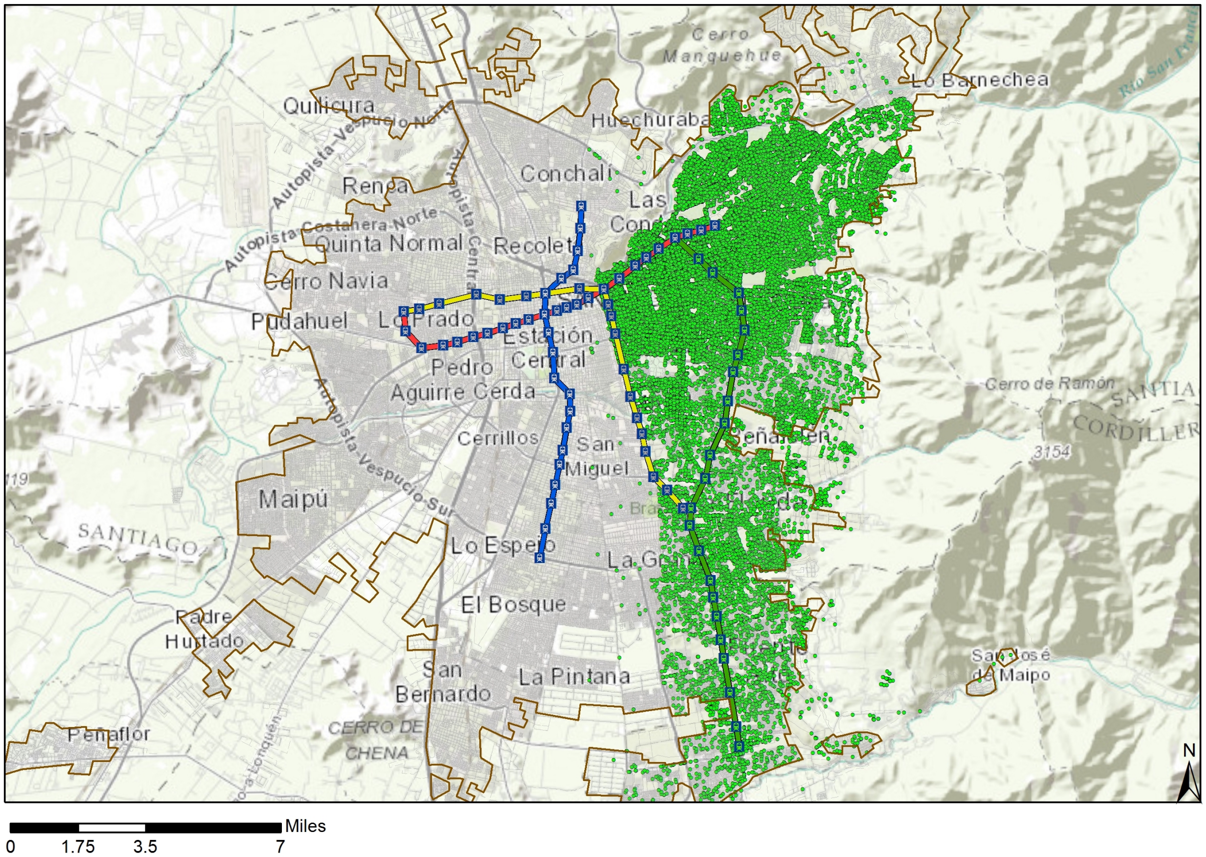
Note: I generated this map by geocoding the exact location of each subway station that existed at the beginning of 2006, as informed by Metro de Santiago. The subway network has since been expanded in recent years with both the inauguration of new lines, and the extension of existing lines.

of this analysis are explained in the subsections that follow.

3.3.1 Dataset

The main dataset used for this study is generated from yearly reports compiled by Chile's Internal Revenue Service (Servicio de Impuestos Internos); the dataset contains a census of all firms legally operating in the country. Using unique firm identifiers, I create a panel that allows me to follow individual firms over the period 2005–2009 and track changes to the number of workers in specific firms, focusing on firms in the municipalities surrounding the new subway line, as shown in Figure 3.3.

Figure 3.3: Firms in Selected Municipalities - 2009



Note: I generated this map by geocoding the exact location of all firms.

In the year 2005, 55,591 firms operated in the area under consideration. During that year, 12,643 new firms started operating and 4,419 firms closed. These numbers track with averages for the period: each year of the period of study, about 8% of firms stopped operating and about 20% of started operating.

3.3.2 Definitions of variables

The main outcome variable considered in this study is the number of employees in each firm. I also consider the logarithm of the number of workers in each firm, since it provides a more intuitive interpretation of the results.

The treatment variable is defined as firms' location relative to a new subway station. Firms located within a 2.5-km radius of a new station are considered as treated; firms located between 2.5 km and 4 km from a subway station make up the control group. These distances are based on other studies of subway use. For example, a study of Dutch rail stations shows that a person's usage of a station declines as the distance between his or her residence and the station increases, and the largest decline seems to happen at about 3 kilometers from the station (Keijer and Rietveld, 2011).

The dataset on which this study relies contains extensive information on firms operating in the country. For each firm, the dataset includes the firm's address, the type of activity the firm performs (e.g., agriculture, transportation, retail), the number of employees, the date the firm started operating, the date the firm ended operations, and the firm's annual sales range.

A firm's primary activity is defined when the firm registers with the Internal Revenue Service at the beginning of its operations. Activity is divided into 18 main categories and hundreds of subcategories; the distribution of economic activity in the relevant area of the city for the year 2009 is presented in Table 3.4. As the table shows, the three most common firm activities are real estate, financial intermediation, and retail/wholesale. These three activities account for about two-thirds of all firms in the selected area.

Table 3.4: Type of Economic Activity - 2009

	Number	Percent
Agriculture	2,154	2.42
Building Administration	232	0.26
Construction	6,581	7.40
Education	1,393	1.57
Financial Intermediation	18,696	21.02
Fishing	110	0.12
Foreign Organization	16	0.02
Hotels and Restaurants	2,320	2.61
Metallic Manufacturing Industry	1,477	1.66
Mining	525	0.59
Non-metallic Manufacturing Industry	2,791	3.14
Other Social or Community Activities	3,383	3.80
Public Administration and Defense	22	0.02
Real Estate	22,645	25.46
Retail and Wholesale	17,448	19.62
Social and Health Services	5,037	5.66
Transportation, Storage, Communications	3,837	4.31
Utilities Supplier	283	0.32
Total	88,950	100.00

Source: Data from Servicio de Impuestos Internos

In 2009, real estate firms had 47 workers, on average; financial intermediation firms had 27 workers, and retail/wholesale firms 31 workers. The average number of workers for each activity category can be seen in Table 3.5.

Table 3.5: Average Number of Workers by Activity - 2009

	Mean	N
Agriculture	89.7	94,416
Building Administration	10.8	2,366
Construction	128.2	392,783
Education	84.1	79,789
Financial Intermediation	26.5	92,489
Fishing	163.0	6,683
Foreign Organization	8.6	121
Hotels and Restaurants	81.4	100,632
Metallic Manufacturing Industry	38.3	28,502
Mining	134.1	27,617
Non-metallic Manufacturing Industry	91.3	119,458
Other Social or Community Activities	31.0	60,051
Public Administration and Defense	738.8	14,776
Real Estate	46.6	411,966
Retail and Wholesale	31.0	242,068
Social and Health Services	19.4	44,141
Transportation, Storage, Communications	36.9	67,627
Utilities Supplier	48.5	6,012

Source: Data from Servicio de Impuestos Internos

The annual sales range calculated by the Internal Revenue Service, which classifies firms into ranges without representing any firm's actual economic value.[1] The ranges are defined as follows:

1. First Range Micro Firm: 0.01 to 200 Annual UF
2. Second Range Micro Firm: 200 to 600 Annual UF
3. Third Range Micro Firm: 600 to 2,400 Annual UF
4. First Range Small Firm: 2,400 to 5,000 Annual UF
5. Second Range Small Firm: 5,000 to 10,000 Annual UF
6. Third Range Small Firm: 10,000 to 25,000 Annual UF
7. First Range Medium Firm: 25,000 to 50,000 Annual UF
8. Second Range Medium Firm: 50,000 to 100,000 Annual UF
9. First Range Large Firm: 100,000 to 200,000 Annual UF
10. Second Range Large Firm: 200,000 to 600,000 Annual UF
11. Third Range Large Firm: 600,000 to 1,000,000 Annual UF
12. Fourth Range Large Firm: more than 1,000,000 Annual UF

The distribution across sales ranges of firms in the area of interest for 2009 is presented in Table 3.6. As the table shows, the vast majority of firms are classified as either micro or small firms. Firms in those sales ranges account for almost two-thirds of all firms in the area of interest.

1. Values are in UF, where 1 UF is equivalent to approximately $40 during 2017.

Table 3.6: Sales Range - 2009

	Number	Percent
First Range Micro Firm	9,673	10.87
Second Range Micro Firm	8,160	9.17
Third Range Micro Firm	16,834	18.93
First Range Small Firm	9,801	11.02
Second Range Small Firm	7,465	8.39
Third Range Small Firm	6,410	7.21
First Range Medium Firm	3,147	3.54
Second Range Medium Firm	2,117	2.38
First Range Large Firm	1,375	1.55
Second Range Large Firm	1,166	1.31
Third Range Large Firm	295	0.33
Fourth Range Large Firm	629	0.71
No information	21,878	24.60
Total	88,950	100.00

Source: Data from Servicio de Impuestos Internos

The number of years each firm in the panel is operating is presented in Table 3.7. As can be seen, during the five-year period under consideration, 35% of firms are present every year, and about 18% of firms appear only during one year.

Table 3.7: Number of Years a Firm Operates

	Number	Percent
1	20,334	18.15
2	19,160	17.10
3	19,122	17.07
4	13,549	12.09
5	39,881	35.59

Source: Data from Servicio de Impuestos Internos

3.3.3 Geocoding

Approximately 370,000 firm addresses were geocoded for this study using a custom script that first corrected errors in the addresses and then used a combination of geocoding services provided by Google and ESRI to find the coordinates of firms. This process achieved a high level of accuracy: for each year in the dataset, about 95% of addresses were geocoded accurately.

3.3.4 Firm-Station distance calculations

The main distance calculation of interest in the estimation is the distance between firms and subway stations. There are several possible ways to calculate these distances. For simplicity, much previous work has used a straight-line distance calculation. However, given that the object of this study is an urban area, what is commonly referred to as "block" or "Manhattan block" distance (Boscoe, 2013) is more relevant and likely more accurate.

The difference between these two types of measurements is not trivial; a straight-line calculation will yield a distance considerably shorter than the more realistic block calculation even for a distance as short as a few city blocks. As a simple example, if a person travels two standard city blocks of 100 meters each to get to her destination, including turning a corner, the straight-line distance calculation will underestimate the distance traveled by 30% compared to the block-distance calculation.

The formula used to calculate block distance follows Gimpel and Schuknecht (2003):

$$d_i = |x_i - x_j| + |y_i - y_j|,$$

where x_i is the longitude of firm i, and y_j is the latitude of station j. This formula assumes that individuals don't travel in a straight line between two points, which is a more realistic representation of travel patterns within cities.

3.4 Estimation Strategy

I use a difference-in-difference approach to estimate the effect of subway access on employment. In this approach, variation is determined by the distance between firms and new subway stations, which in the short term should be exogenous. The validity of this method would be threatened if firms located near the subway would have hired more or fewer workers relative to firms located farther away regardless of the construction of the subway line. This is partly addressed by considering a period close to the opening of the line, comparing characteristics between groups before and after the subway opened, and examining the trends in each group. This approach may also reduce concerns about the subway producing changes in social composition or infrastructure in the area surrounding the station. Finally, I present results considering only the years before the subway line opened. This analysis provides evidence to support the idea that firms located near the area where the subway stations were constructed were not hiring significantly different numbers of workers before the opening of the subway.

3.4.1 Treatment and control groups

Table 3.8 shows a comparison of several relevant variables between control and treatment groups before the subway line opened in the year 2005. In Table 3.9, I show this comparison for the year 2009—that is, after the subway line had opened. The first of the variables is the average number of workers per firm. Firms in the control group in the year 2005 hired a similar number of workers to firms in the treatment group: 28.9 vs. 29.4, respectively. By the year 2009, firms in the control group were hiring fewer workers, 26.1 on average, while firms in the treatment group were hiring more workers—32.7, on average.

Table 3.8: Descriptive Statistics Before Treatment - 2005

	Control	Treated
Workers		
Average Number of Workers	28.9	29.4
Total Number of Workers	2384.0	8421.0
Type of Activity		
Agriculture (pct.)	2.1	2.5
Building Administration (pct.)	0.0	0.1
Construction (pct.)	10.9	9.8
Education (pct.)	3.1	3.3
Financial Intermediation (pct.)	10.9	10.2
Fishing (pct.)	0.0	0.1
Foreign Organization (pct.)	0.0	0.0
Hotels and Restaurants (pct.)	3.7	3.0
Metallic Manufacturing Industry (pct.)	2.7	2.6
Mining (pct.)	0.5	0.3
Non-metallic Manufacturing Industry (pct.)	5.5	4.4
Other Social or Community Activities (pct.)	5.2	6.0
Public Administration and Defense (pct.)	0.0	0.0
Real Estate (pct.)	21.3	21.8
Retail and Wholesale (pct.)	22.1	24.9
Social and Health Services (pct.)	5.7	5.2
Transportation, Storage, Communications (pct.)	6.2	5.4
Utilities Supplier (pct.)	0.0	0.1
Sales Range		
First Range Micro Firm (pct.)	14.0	14.5
Second Range Micro Firm (pct.)	11.4	12.2
Third Range Micro Firm (pct.)	22.4	21.3
First Range Small Firm (pct.)	11.0	10.8
Second Range Small Firm (pct.)	6.8	6.9
Third Range Small Firm (pct.)	6.5	6.1
First Range Medium Firm (pct.)	2.7	2.2
Second Range Medium Firm (pct.)	2.1	1.3
First Range Large Firm (pct.)	0.8	0.7
Second Range Large Firm (pct.)	0.6	0.5
Third Range Large Firm (pct.)	0.1	0.2
Fourth Range Large Firm (pct.)	0.1	0.1

Source: Data from Servicio de Impuestos Internos

Table 3.9: Descriptive Statistics After Treatment - 2009

	Control	Treated
Workers		
Average Number of Workers	26.1	32.7
Total Number of Workers	3777.0	12940.0
Type of Activity		
Agriculture (pct.)	1.5	1.6
Building Administration (pct.)	0.3	0.3
Construction (pct.)	11.1	10.7
Education (pct.)	2.4	3.0
Financial Intermediation (pct.)	12.1	11.2
Fishing (pct.)	0.1	0.0
Foreign Organization (pct.)	0.0	0.0
Hotels and Restaurants (pct.)	3.4	3.4
Metallic Manufacturing Industry (pct.)	3.6	3.2
Mining (pct.)	0.5	0.4
Non-metallic Manufacturing Industry (pct.)	5.0	4.2
Other Social or Community Activities (pct.)	3.3	4.6
Public Administration and Defense (pct.)	0.0	0.0
Real Estate (pct.)	21.6	21.9
Retail and Wholesale (pct.)	22.2	23.9
Social and Health Services (pct.)	5.9	5.3
Transportation, Storage, Communications (pct.)	6.8	6.1
Utilities Supplier (pct.)	0.1	0.1
Sales Range		
First Range Micro Firm (pct.)	13.9	13.5
Second Range Micro Firm (pct.)	10.4	11.6
Third Range Micro Firm (pct.)	22.5	21.9
First Range Small Firm (pct.)	11.1	11.2
Second Range Small Firm (pct.)	7.3	7.6
Third Range Small Firm (pct.)	5.9	6.2
First Range Medium Firm (pct.)	2.6	2.7
Second Range Medium Firm (pct.)	1.5	1.2
First Range Large Firm (pct.)	0.7	0.5
Second Range Large Firm (pct.)	0.5	0.5
Third Range Large Firm (pct.)	0.2	0.1
Fourth Range Large Firm (pct.)	0.1	0.2

Source: Data from Servicio de Impuestos Internos

The next set of variables shows that the types of activity performed by firms both close to and farther away from the subway were very similar. This seems to be the case both in 2005 and 2009, without significant change over time.

Finally, the percentage of firms in each sales range is similar between the two groups—close to 50% of both the treatment and control groups are micro firms—before and after the subway was built. As can be seen in Tables 3.8 and 3.9, all of these variables have similar values and their differences remain practically constant over time.

Figure 3.4 shows a close-up image of an area close to the new subway line, to illustrate how firms are classified into the treatment or control group based on their distance from the subway station. The analysis of the available information presented here shows that firms in both groups were reasonably similar and their differences remained relatively constant over time.

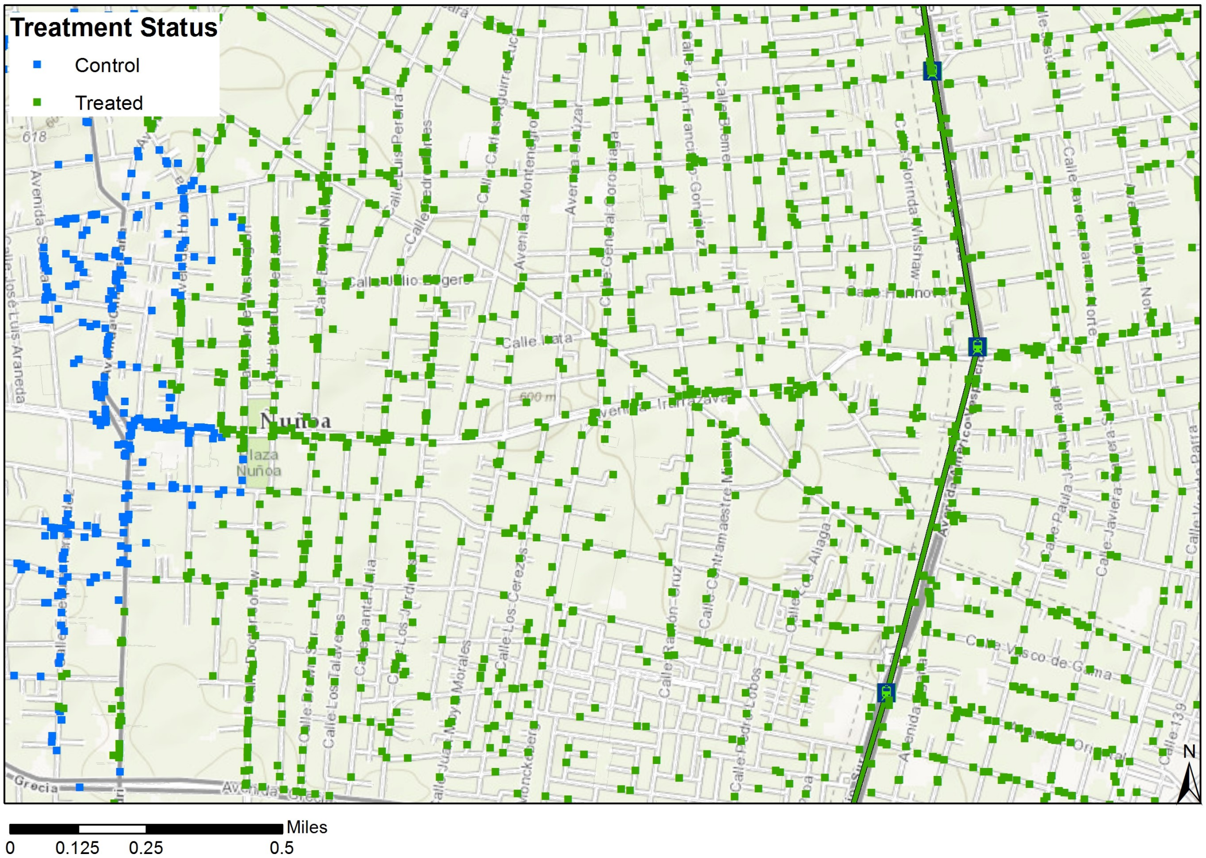
3.4.2 Effect of subway access on employment: Differences-in-differences

I estimate the effect of subway access on the number of workers per firm to see if firms close to a new station hired more workers than firms that were not close to a new station. Firms located within a 4-km radius of a new subway station in 2005, before the line started operating, are considered part of the "before" group. In turn, firms located within a 4-km radius of a new subway station after 2007 are part of the "after" group.

The group of firms located near the new subway line (within a 2.5-km radius) after the line opened were most likely to be affected by the subway, so these firms are considered as being in the treatment group, while the rest of the firms, those located between 2.5 km and 4 km away from a new subway station, will be considered as being in the control group.

The key identifying assumption of the difference-in-difference estimation is that both groups of firms would have followed similar trends in hiring if the subway line had not

Figure 3.4: Firms in Treatment and Control Groups - 2009



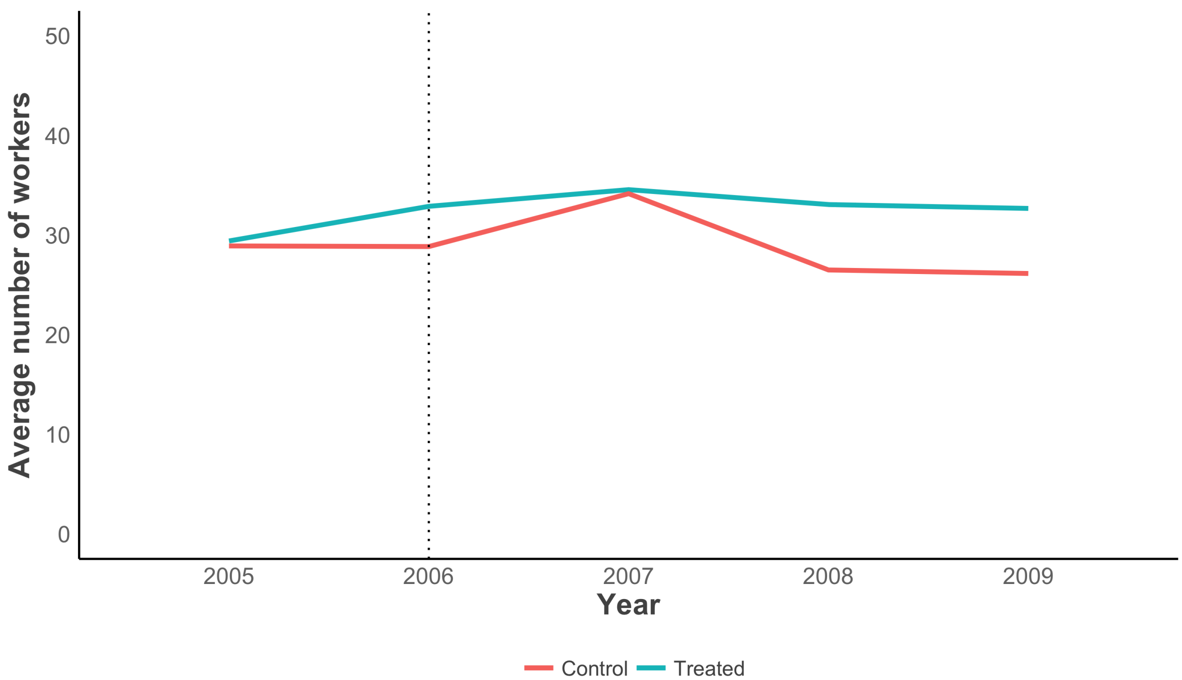
Note: I generated this map by geocoding the exact location of all firms, and then identifying firms that were located within the area of influence of a subway station.

been built. Although it is not possible to test this assumption directly, by looking at graphs of the trends for the outcome variable it is possible to visually inspect whether the two groups followed parallel trends until the subway line was inaugurated.

Defining the groups in terms of their distance from a new subway station makes it likely that the treatment and control groups will experience similar changes in their characteristics on average, in the absence of the subway station, as the primary difference between the two groups is their distance from the new subway stations. These definitions are then used to estimate a difference-in-difference specification.

I show the trends for the average number of workers in Figure 3.5. As the chart shows, firms in the treatment and control groups don't seem to follow parallel trends in the years before the subway opened. Over the period being studied, firms in the treatment group had on average about 32 workers, while firms in the control group had on average 28 workers. Even if the series do not appear to follow parallel trends, estimating a regression considering only the periods before the subway was fully operating does not reveal significantly different trends.

Figure 3.5: Average Number of Workers In Treated and Control Groups: 2005-2009



Note: This chart shows the average number of workers each year, for firms in the treatment and control groups.

The key equation that I estimate is:

$$\text{Workers}_{it} = \beta_0 + \beta_1 \text{Subway}_i + \beta_2 \text{Post}_t + \gamma(\text{Post}_t * \text{Subway}_i) + X_{it} + \epsilon_{it} \tag{3.1}$$

where the outcome variable, *Workers*, is the number of workers in firm i in year t. I also estimate equation 3.1 using the logarithm of the number of workers as an alternative outcome variable. *Subway* is a binary variable that is set to one for firms within a 2.5-km radius of a new subway station. *Post* is a binary variable that is set to one in years after the new subway line was built, which in this case are 2007, 2008 and 2009.

The parameter γ, which corresponds to the difference-in-difference estimator, captures the effect of increased subway access on employment. It seems reasonable to assume that firms located closer to the subway were more likely to benefit from it than firms located farther away. If firms located closer to the subway started hiring more workers than firms farther away after the subway opened, we should expect to observe a positive and significant coefficient for this parameter.

X_{it} is a vector of control variables that includes the activity of the firm, the sales range of the firm, and dummy variables for each municipality. Since the identification strategy being used does not depend on the inclusion of these covariates, it is expected that they should not affect the results.

As the dataset includes all years between 2005 and 2009, it is possible to see if the effect of having access to the subway changed over time, and if there were significant effects before the actual intervention. For this, I estimate an equation like:

$$\text{Workers}_{it} = \beta_0 + \lambda_t + \phi_i + \sum_{\tau=0}^{m} \beta_{-\tau} D_{i,t-\tau} + \sum_{\tau=1}^{q} \beta_{+\tau} D_{i,t+\tau} + X_{it} + \epsilon_{it} \tag{3.2}$$

in which the sums allow for post-treatment and anticipatory effects (Angrist and Pischke, 2008). This equation includes year fixed effects, firm fixed effects, and the treatment variable Dit represents access to the subway.

3.5 Results

In this section, I look at the results from the estimation of the effect of subway access on employment and then at the particular cases of employment in real estate firms and in construction firms.

3.5.1 Results: Effect of subway access on employment

The first set of results, presented in Table 3.10, is from the estimation of equation 3.1, with and without covariates, using workers per firm as the outcome in two forms: the number of workers per firm and the logarithm of the number of workers per firm. Columns 1 and 3 present estimations without covariates; Columns 2 and 4 include covariates.

Table 3.10: Regression Results (DID): The Effect of Subway Access on Number of Workers in All Firms

	Workers		Log Workers	
	(1)	(2)	(3)	(4)
Access to Subway	6.033	3.962	0.064	0.019
	(8.699)	(8.399)	(0.073)	(0.050)
Observations	10,561	9,711	10,561	9,711
R^2	0.000	0.192	0.000	0.571

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Columns (2) and (4) present results controlling for the following variables: municipality, sales range, and type of activity.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

As can be seen in Table 3.10, the estimated effect of subway access on employment is positive but not significant, and it decreases when control variables are added. The coefficients in Columns 1 and 2 show that access to the subway doesn't cause a significant change in the number of workers employed compared to firms located farther from the subway.

Columns 3 and 4 in Table 3.10 confirm these results. The coefficient of 0.004, using the log of the number of workers as the outcome variable, shows that having access to the subway doesn't cause a significant increase in employment. The new subway line may increase connectivity for firms located nearby, but that increased connectivity is not reflected in changes in employment during the period of study.

Restricting the sample by using different definitions of the treatment and control groups does not alter the main results. Defining the treatment group to include firms in a wider or narrower area around the new station produces similar results in terms of the direction and magnitude of the effect.

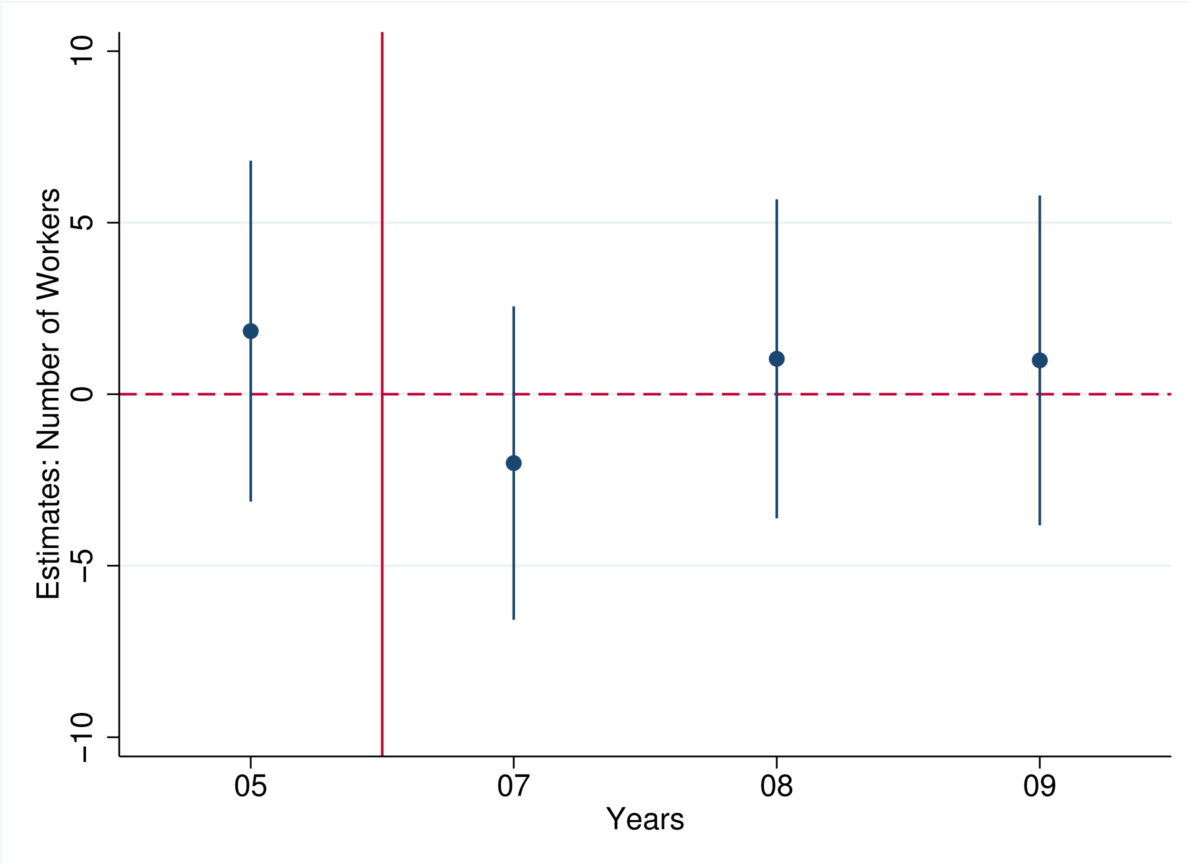
Finally, I look at the results of estimating equation 3.2, to see if employment in firms changed before or after the treatment in specific years for firms in the treatment group and firms in the control group. As Figure 3.6 shows, there were no significant changes in the period before or after the subway was inaugurated.

The same result can be seen in Table 3.11, which presents results from the estimation of equation 3.2, with and without covariates, using workers per firm as the outcome in two forms: the number of workers per firm and the logarithm of the number of workers per firm. Columns 1 and 3 present estimations without covariates; Columns 2 and 4 include covariates. Here too, the effect of the subway line on employment was not significant for any of the years under consideration.

3.5.2 Results: Effect of subway access on employment in real estate firms

Real estate firms are of special interest because research shows that subway or rail station openings affect property values. To examine whether the number of workers in this type of firm is affected by station openings, I estimate equations 3.1 and 3.2, this time considering

Figure 3.6: Regression Estimates by Year: All Firms



Note: This figure shows yearly estimates of the effect of the opening of the new subway line on the number of workers per firm.

only firms engaged in real estate activity.

As can be seen in Table 3.12, the estimated effect of subway access on employment in real estate firms is positive but not significant and remains practically unchanged when control variables are added. The coefficients in Columns 1 and 2 show that having access to the subway doesn't cause a significant change in the number of workers employed in real estate, compared to firms of the same type located farther from the subway station.

The results of estimating equation 3.2 reveal that there was a positive and significant increase in employment for real estate firms located close to the subway, compared to real estate firms in the control group, in 2009. However, as Figure 3.7, shows, there were no

Table 3.11: Regression Results (DID): The Effect of Subway Access on Number of Workers in All Firms: Yearly Effects

	Total Workers		Log Workers	
	(1)	(2)	(3)	(4)
dd05	1.838	2.199	0.004	0.011
	(2.535)	(2.566)	(0.027)	(0.026)
dd07	-2.008	-1.774	-0.026	-0.013
	(2.331)	(2.363)	(0.025)	(0.024)
dd08	1.031	1.334	0.005	0.006
	(2.373)	(2.412)	(0.025)	(0.024)
dd09	0.986	0.467	0.017	0.013
	(2.453)	(2.505)	(0.026)	(0.025)
Observations	27084	24987	27084	24987
Adjusted R^2	0.926	0.933	0.893	0.910
Controls	No	Yes	No	Yes

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Columns (2) and (4) present results controlling for the following variables: municipality, sales range, and type of activity

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

significant changes either before the subway opened or during its first years of operation. The same result can be seen in Table 3.13, which presents results from the estimation of equation 3.2, with and without covariates.

3.5.3 Results: Effect of subway access on employment in construction firms

Construction firms are another interesting category, as the opening of a subway station seems likely to generate other changes in the infrastructure of the surrounding areas, and thus to lead to a growth in construction activity. Because of this, construction firms located in that area could find themselves in need of more workers. To examine whether the number of workers in this type of firm are affected by a station opening, I estimate equations 3.1 and 3.2, this time considering only construction-related firms.

Table 3.12: Regression Results (DID): The Effect of Subway Access on Number of Real Estate Workers

	Workers		Log Workers	
	(1)	(2)	(3)	(4)
Access to Subway	22.104	21.757	0.004	0.053
	(29.556)	(30.261)	(0.163)	(0.126)
Observations	1,965	1,859	1,965	1,859
R^2	0.001	0.085	0.001	0.454

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Columns (2) and (4) present results controlling for the following variables: municipality and sales range.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

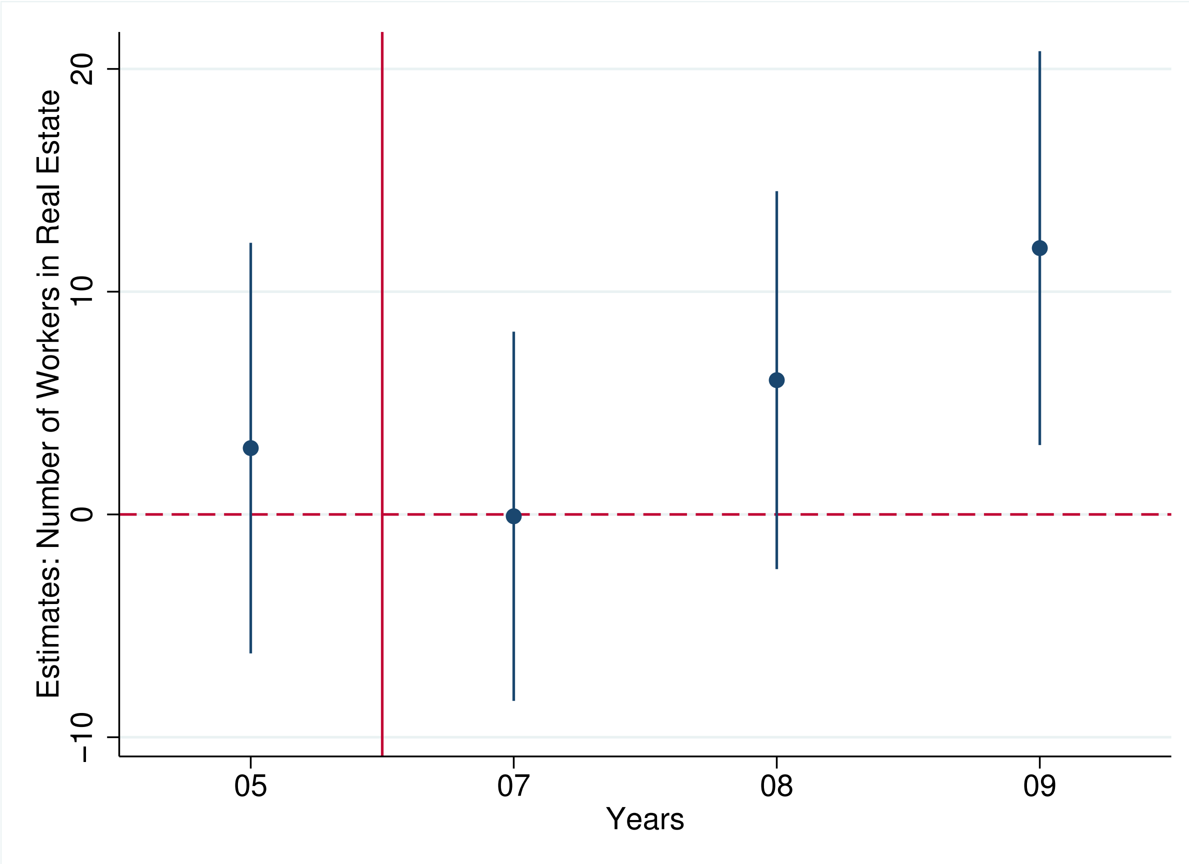
As can be seen in Table 3.14, the estimated effect of subway access on employment in construction firms is negative but not significant. Having access to the subway doesn't cause a significant change in the number of workers employed in construction compared to firms of the same type located farther from the subway.

The results of estimating equation 3.2 reveal that the effect of the introduction of the subway on employment in construction firms was not significant for any of the years under consideration. As Figure 3.8 shows, there were no significant changes either before or after the subway opened. The same result can be seen in Table 3.15, which presents results from the estimation of equation 3.2, with and without covariates.

3.6 Conclusion

Subway networks constitute major investments for cities and their success is generally evaluated by the number of passengers they transport. But although the effects of connecting geographically separate urban areas could be much broader, there have been few studies aimed at examining these effects. Determining the impact of improved transportation on employment is relevant from a policy perspective, since, as Haughwout (1999),

Figure 3.7: Regression Estimates by Year: Real Estate Firms



Note: This figure shows yearly estimates of the effect of the opening of the new subway line on the number of workers per real estate firm.

states, "While employment growth is not a useful proxy for the value of public investment, it is an important state policy objective" (Haughwout, 1999, pg. 550). Having a better understanding of the tools available to help increase employment in urban areas is an important objective; thus, it is important to establish whether expanded public transportation is such a tool.

The main results of this paper show no significant changes in employment between firms close to a new subway station and firms located farther away after the opening of the station. Additionally, there is no evidence of effects on employment growing over time. An examination of the data for specific types of firms shows that employment

Table 3.13: Regression Results (DID): The Effect of Subway Access on Number of Real Estate Workers: Yearly Effects

	Total Workers		Log Workers	
	(1)	(2)	(3)	(4)
dd05	2.978	3.002	-0.040	-0.022
	(4.700)	(4.817)	(0.062)	(0.060)
dd07	-0.082	1.488	0.084	0.107**
	(4.227)	(4.293)	(0.056)	(0.053)
dd08	6.029	7.269*	0.079	0.104*
	(4.327)	(4.406)	(0.057)	(0.055)
dd09	11.956***	14.299***	0.160***	0.206***
	(4.509)	(4.611)	(0.059)	(0.057)
Observations	5063	4791	5063	4791
Adjusted R^2	0.959	0.962	0.882	0.900
Controls	No	Yes	No	Yes

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Columns (2) and (4) present results controlling for the following variables: municipality and sales range.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

in firms related to real estate did increase significantly in areas surrounding the new subway stations, although only two years after the subway opened, even though overall employment did not change.

These findings provide relevant information for policymakers and urban planners as to the limits of infrastructure projects as tools to revitalize urban areas. Investments in public transportation should be a consequence of city growth and should be aimed at reducing congestion, long commutes, and pollution. As pointed out by Glaeser (2011), if economic development is needed in urban areas, investments in human capital may do much more than infrastructure spending.

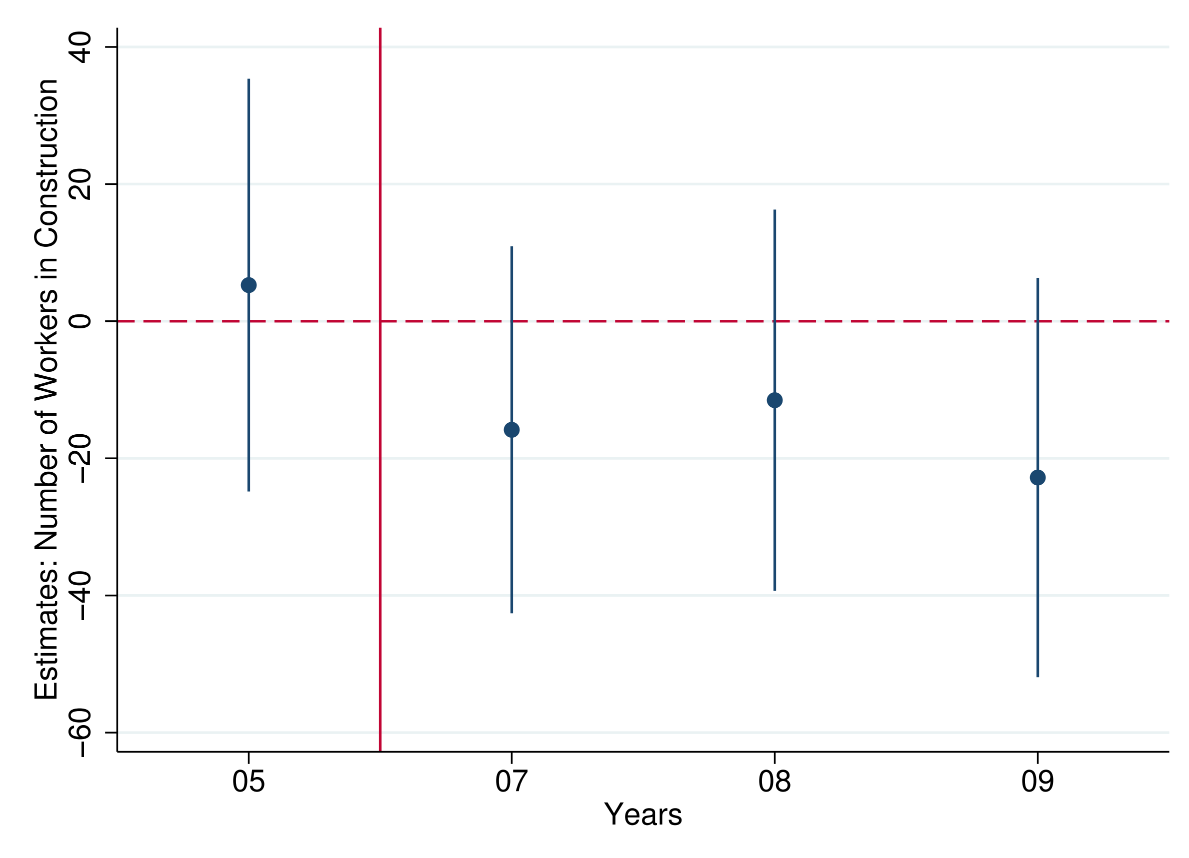
Table 3.14: Regression Results (DID): The Effect of Subway Access on Number of Construction Workers

	Workers		Log Workers	
	(1)	(2)	(3)	(4)
Access to Subway	-17.877	-23.157	0.001	-0.058
	(23.984)	(16.564)	(0.209)	(0.151)
Observations	1,315	1,282	1,315	1,282
R^2	0.001	0.554	0.000	0.500

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Columns (2) and (4) present results controlling for the following variables: municipality and sales range.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

Figure 3.8: Regression Estimates by Year: Construction Firms



Note: This figure shows yearly estimates of the effect of the opening of the new subway line on the number of workers per construction firm.

Table 3.15: Regression Results (DID): The Effect of Subway Access on Number of Construction Workers: Yearly Effects

	Total Workers		Log Workers	
	(1)	(2)	(3)	(4)
dd05	5.266	-0.129	0.025	0.032
	(15.348)	(12.053)	(0.120)	(0.103)
dd07	-15.838	-1.466	-0.173	-0.078
	(13.646)	(10.850)	(0.107)	(0.093)
dd08	-11.521	3.931	-0.071	-0.043
	(14.176)	(11.252)	(0.111)	(0.096)
dd09	-22.801	-10.863	-0.074	-0.073
	(14.852)	(11.917)	(0.116)	(0.102)
Observations	3407	3338	3407	3338
Adjusted R^2	0.803	0.885	0.766	0.832
Controls	No	Yes	No	Yes

Note: Each set of columns presents the results of a separate regression, one for each outcome variable. Columns (2) and (4) present results controlling for the following variables: municipality and sales range.

* p <0.1, ** p <0.05, *** p <0.01. Standard errors in parentheses.

BIBLIOGRAPHY

AAAFTS (2015). 2014 traffic safety culture index. Technical report, AAA Foundation for Traffic Safety.

Agostini, C. and Palmucci, G. (2005). Capitalizacion anticipada del metro de santiago en el precio de las viviendas. *ILADES-Georgetown University Working Papers*.

Alderman, H., Orazem, P. F., and Paterno, E. M. (2001). School quality, school cost and the public/private school choices of low-income households in pakistan. *The Journal of Human Resources*, 36(2):304–326.

Andre-Bechely, L. (2007). Finding space and managing distance: Public school choice in an urban california district. *Urban Studies*, 44(7):1355–1376.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Asahi, K. (2014). The impact of better school accessibility on student outcomes. *Spatial Economic Research Centre Discussion Paper*, 156.

Automóvil Club (2015). Survey of drivers in santiago. Technical report, Unidad de Estudios Automóvil Club de Chile.

Bae, C.-H. C., Jun, M.-J., and Park, H. (2003). The impact of seoul's subway line 5 on residential property values. *Transport Policy*, 10:85–94.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, pages 775–805.

Bhargava, S. and Pathania, V. S. (2013). Driving under the (cellular) influence. *American Economic Journal: Economic Policy*, 5(3):92–125.

Bollinger, C. R. and Ihlanfeldt, K. R. (1997). The impact of rapid rail transit on economic development: The case of atlanta's marta. *Journal of Urban Economics*, 42(2):179 – 204.

Boscoe, F. (2013). *Geographic Health Data: Fundamental Techniques for Analysis*. CAB International.

Bowes, D. R. and Ihlanfeldt, K. R. (2001). Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics*, 50(1):1 – 25.

Braitman, K. A. and McCartt, A. T. (2010). National reported patterns of driver cell phone use in the united states. *Traffic Injury Prevention*, 11(6):543–548. PMID: 21128181.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. *NBER Working Paper Series*, (4483).

Chumacero, R., Caorsi, D. G., and Paredes, R. (2008). I would walk 500 miles (if it paid). *Munich Personal RePEc Archive (MPRA).*

Consejo de Rectores de las Universidades Chilenas (2015). Proceso de admisión 2015.

Corvalán, J. and Román, M. (2012). The permanence of low-performance schools in the chilean educational quasi market. *Revista Uruguaya de Ciencia Política.*

De Grange, L. (2010). El gran impacto del metro. *EURE*, 36(27):125–131.

Debrezion, G., Pels, E., and Rietveld, P. (2007). The impact of railway stations on residential and commercial property value: A meta-analysis. *The Journal of Real Estate Finance and Economics*, 35(2):161–180.

Denzler, S. and Wolter, S. C. (2011). Too far to go? does distance determine study choices? *IZA Discussion Paper*, (5712).

Dickerson, A. and McIntosh, S. (2010). The impact of distance to nearest education institution on the post-compulsory education participation decision. *Sheffield Economic Research Paper Series*, (2010007).

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S., Sudweeks, J., Perez, M., Hankey, J., Ramsey, D., Gupta, S., et al. (2006). The 100-car naturalistic driving study, phase ii-results of the 100-car field experiment. Technical report.

Duncan, M. (2011). The impact of transit-oriented development on housing prices in san diego, ca. *Urban Studies*, 48(1):101–127.

Elacqua, G. and Fabrega, R. (2004). El consumidor de la educación: El actor olvidado de la libre elección de escuelas en chile. *PREAL, Universidad Adolfo Ibáñez.*

Elacqua, G., Schneider, M., and Buckley, J. (2006). School choice in chile: Is it class or the classroom? *Journal of Policy Analysis and Management*, 25(3):577–601.

ENTEL (2014). Todo chile comunicado. Technical report.

Falch, T., Lujala, P., and Strom, B. (2013). Geographical constraints and educational attachment. *Regional Science and Urban Economics*, 43:164–176.

Frenette, M. (2004). Access to college and university: Does distance to school matter? *Canadian Public Policy / Analyse de Politiques*, 30(4):427–443.

Friedman, M. (1962). *Capitalism and Freedom.* Chicago University Press.

Friedman, M. (1997). Public schools: Make them private. *Education Economics*, 5(3):341–344.

Gallego, F. and Hernando, A. (2009). School choice in chile: Looking at the demand side. *Working Paper - Pontificia Universidad Católica de Chile, Instituto de Economía.*

Gallego, F. and Hernando, A. E. (2008). On the determinants and implications of school choice: Semi-structural simulations for chile. Documentos de trabajo, Instituto de Economia. Pontificia Universidad Católica de Chile.

Gertler, P. and Glewwe, P. (1989). The willingness to pay for education in developing countries: Evidence from rural peru. *LSMS Working Paper*, (54).

GHSA (2011). Distracted driving - what research shows and what states can do. Technical report, Governors Highway Safety Association.

Gibbons, S. and Vignoles, A. (2009). Access, choice and participation in higher education. *Working Paper - Centre for the Economics of Education.*

Gimpel, J. and Schuknecht, J. (2003). Political participation and the accessibility of the ballot box. *Political Geography*, 22:471–488.

Glaeser, E. (2011). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier*. Penguin.

Griffith, A. L. and Rothstein, D. S. (2007). Can't get here from there: The decision to apply to a selective institution. *Working Paper - Cornell University, School of Industrial and Labor Relations.*

Hastings, J. S., Kane, T. J., and Staiger, D. O. (2007). Parental preferences and school competition: Evidence from a public school choice program. *Working Paper.*

Haughwout, A. F. (1999). State infrastructure and the geography of employment. *Growth and Change*, 30(4):549–566.

Hess, D. B. and Almeida, T. M. (2007). Impact of proximity to light rail rapid transit on station-area property values in buffalo, new york. *Urban Studies*, 44(5-6):1041–1068.

Hewitt, C. M. and Hewitt, W. (2012). The effect of proximity to urban rail on housing prices in ottawa. *Journal of Public Transportation*, 15(4).

Holl, A. (2004). Transport infrastructure, agglomeration economies, and firm birth: Empirical evidence from portugal*. *Journal of Regional Science*, 44(4):693–712.

Holzer, H. J., Quigley, J. M., and Raphael, S. (2003). Public transit and the spatial distribution of minority employment: Evidence from a natural experiment. *Journal of Policy Analysis and Management*, 22(3):415–441.

Hsieh, C.-T. and Urquiola, M. (2002). When schools compete how do they compete? an assessment of chile's nationwide school voucher program. *Working Paper*.

jae Lee, M. and Kang, C. (2006). Identification for difference in differences with cross-section and panel data. *Economics Letters*, 92:270–276.

Keijer, M. and Rietveld, P. (2011). How do people get to the railway station? the dutch experience. *Transportation Planning and Technology*, 23(3):215–235.

Kjellstrom, C. and Regnér, H. (1999). The effects of geographical distance on the decision to enroll in university education. *Scandinavian Journal of Educational Research*, 43(4):335–348.

Liberty Mutual Insurance (2016). Teen driving study reveals "app and drive" is new danger among teens, new worry for parents.

McCartt, A. T., Hellinga, L. A., and Bratiman, K. A. (2006). Cell phones and driving: review of research. *Traffic Inj Prev*, 7(2):89–106.

McEwan, P. and Carnoy, M. (2000). The effectiveness and efficiency of private schools in chile's voucher system. *Educational Evaluation and Policy Analysis*, 22(3):213–239.

McEwan, P., Urquiola, M., and Vegas, E. (2008). School choice, stratification, and information on school performance: Lessons from chile. *Economía*.

Metro de Santiago (2013). Memoria anual metro de santiago.

Meyer, J., Kain, J., and Wohl, M. (1965). *The urban transportation problem*. Rand Corporation research study. Harvard University Press.

MTT (2008). Bases específicas concurso público para la asignación del proyecto infraestructura digital para la competitividad e innovación y su respectivo subsidio correspondiente al programa anual de proyectos subsidiables del año 2008 del fondo de desarrollo de las telecomunicaciones. Technical report, Ministerio de Transportes y Telecomunicaciones.

NHTSA (2015). Driver electronic device use in 2014. Summary of statistical findings, National Highway Traffic Safety Administration.

Nikolaev, A. G., Robbins, M. J., and Jacobson, S. H. (2010). Evaluating the impact of legislation prohibiting hand-held cell phone use while driving. *Transportation Research Part A: Policy and Practice*, 44(3):182 – 193.

NSC (2013). Crashes involving cell phones - challenges of collecting and reporting reliable crash data. Technical report, National Safety Council.

Olson, R. L., Hanowski, R. J., Hickman, J. S., and Bocanegra, J. L. (2009). Driver distraction in commercial vehicle operations. Technical report.

O'Regan, K. M. and Quigley, J. M. (1996). Spatial effects upon employment outcomes: The case of new jersey teenagers.

OSUAH (2012). Encuesta origen destino de viajes. Technical report, Universidad Alberto Hurtado.

Phillips, D. C. and Sandler, D. (2015). Does public transit spread crime? evidence from temporary rail station closures. *Working paper*.

Redelmeier, D. A. and Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine*, 336(7):453–458. PMID: 9017937.

Registro Civil (2011). Estadísticas licencias de conducir. Technical report, Registro Civil de Chile.

Sá, C., Florax, R. J., and Rietveld, P. (2004). Does accessibility to higher education matter? *Tinbergen Institute Discussion Paper*.

Schuetz, J. (2014). Do rail transit stations encourage neighborhood retail activity? *Working Paper*.

Stoll, M. A., Holzer, H. J., and Ihlanfeldt, K. R. (2000). Within cities and suburbs: Racial residential concentration and the spatial distribution of employment opportunities across sub-metropolitan areas. *Journal of Policy Analysis and Management*, 19(2):207–231.

Strayer, D. L., Drews, F. A., and Crouch, D. J. (2006). A comparison of the cell phone driver and the drunk driver. *Hum Factors*, 48(2):381–391.

TRB (2002). Transit-oriented development and joint development in the united states: A literature review. Technical Report 52, Transportation Research Board.

Winston, C. and Maheshri, V. (2007). On the social desirability of urban rail transit systems. *Journal of Urban Economics*, 62(2):362 – 382.

Yan, S., Delmelle, E., and Duncan, M. (2012). The impact of a new light rail system on single-family property values in charlotte, north carolina. *Journal of Transport and Land Use*, 5(2).