Highway to Hell or Stairway to Heaven? *

The Economic Consequences of Setting Foot in a College in Colombia

Author:

Luis Omar Herrera-Prada†

Universität Göttingen and International Monetary Fund

May 19, 2023

Keywords: Returns to Education, Sheepskin Effect, Higher Education, Education, Graduation, Graduation on Time

JEL Classification Numbers: I23, I24, I26, J24, J31

*I am thankful to Inmaculada Martínez-Zarzoso, Sarah Belford, Juan Esteban Saavedra, Andrés Córdoba, Pablo Druck, Óscar Bernal, Paula Triana, Diana Durán, Luz Emilse Rincón, Clara Paola Pava, Nadmy Daniela Valbuena González, Pelin Berkmen, Dmitry Gershenson, Julia Faltermeier, Sidonia Mckenzie, Hannah Winterberg, Román David Zárate, SPADIES team at Universidad de los Andes (2005-2011), Ministry of Education of Colombia, Ministry of Health and Social Protection of Colombia, and all the very useful comments received in the Development Economics Seminar - Universität Göttingen; Western Hemisphere Department Seminar at the International Monetary Fund, Public Policy Seminar - Universidad de los Andes (Bogotá - Colombia), Jornada NIP 2020 - Universidad de la República (Montevideo-Uruguay), AEFP meeting 2021, MEA-SOLE meeting 2021, and the Public Policy Seminar at Colombian Central Bank in 2021. An online appendix with more detailed charts is available in `https://herreraprada.com` . *The views expressed in this paper are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.*

†Corresponding Author: l.herreraprada@stud.uni-goettingen.de

Abstract

This paper investigates the impact of higher education on wages in Colombia, where the labor market is characterized by high informality, youth unemployment, and self-employment. Using administrative data for over 5.4 million secondary school graduates, our analysis reveals significant average earnings increases for college enrollees compared to non-enrollees: 40.2% (ATT for total post-treatment), 46.7% (LATE), and 80.6% (ATT for post-college period). Additionally, for those who graduate, there is an average premium of 143.5% during the 5 to 10 years after enrolling as freshmen, increasing to 220.4% in the 10th year. Furthermore, a wage difference of 92.4% is observed between graduates and those who dropped out with over 90% coursework completed, which rises to 158.2% in the 10th year. These findings highlight the potential of higher education in reducing inequality and the wage gender gap while leading to higher future income, irrespective of academic skill level, program type, or field of study. Policymakers can utilize these results to promote higher education enrollment and graduation, thereby improving labor market outcomes.

1 Introduction

The decision to pursue higher education can have significant implications for an individual's access to quality employment opportunities in the future. In developing countries, where data availability has improved, examining the relationship between earnings and educational attendance and achievement is particularly important for economists studying development. The case of Colombia is particularly fascinating due to the scale and quality of data available, and the country's challenging economic and labor circumstances. After facing a devastating economic crisis in the early 2000s, Colombia grappled with income and education inequality, high levels of informality in the labor market, and a high rate of self-employment that pushed many to start working at a young age. To address these challenges, the government reduced barriers to college enrollment, increased college attendance rates, and changed the life paths of millions of secondary school graduates. However, barriers still remain, with one of the biggest being the college dropout rate.This paper contributes to the literature by showing how the lives of students who attended college changed for the better; attending college increases future earnings for graduates and dropouts, with the exception of long extended active students who do not see any difference in income compared to their peers who did not attend college. The results indicate that the sheepskin effect[1] and returns to higher education are significant, with the magnitude of the effect increasing with work experience.

A large body of literature examines the implications of pursuing higher education, with key contributions from economists such as Becker (1962), Mincer (1974) , Spence (1973), and Hungerford and Solon (1987). The Mincer equation (Mincer, 1974), which estimates the relationship between education and income, is frequently used in this literature, often based on household surveys. While higher levels of education generally lead to higher income, there is still an ongoing debate regarding whether this relationship holds in developing countries, where access to education is more unevenly distributed (Card, 2001; Duflo, 2001). Our paper finds that income disparities persist for high-income and private secondary school students. Nevertheless, recent research shows that returns on education are generally comparable across developed and developing countries, although outcomes vary depending on factors such as region, race, and educational sector (Peet et al., 2015; Patrinos

[1]The sheepskin effect is a term used to denote when people with an academic degree earn a higher income than those with an equivalent schooling level but without the credential. This effect was first described by Hungerford and Solon (1987), and analyzed in Colombia by (Mora, 2003; Mora and Muro, 2008)

and Psacharopoulos, 2020). In Brazil, Chile, and Colombia, longitudinal surveys have been conducted to provide more accurate measurements of these returns, with Brazil and Colombia being of particular interest due to their exit exams in secondary and tertiary education (Manacorda et al., 2007; Melguizo and Wainer, 2016; MacLeod et al., 2017).

Based on the revised literature, three main gaps in the research on returns to education in Colombia have been identified. Firstly, more tracking of secondary school graduates is required to evaluate if attending college made a difference, despite the fact that the attendance rate is just 52%. Secondly, previous studies have not included self-employed workers in their analysis of returns to education, even though, according to OECD, they represent about 53.1% of the Colombian formal labor force. Finally, the sheepskin effect has yet to be adequately studied in previous approaches, as they were developed with household surveys comparing students with lower levels of education.

In this paper, a set of administrative databases is merged to generate a novel panel dataset that contains comprehensive records for 5.4 million students who graduated from secondary school between 2002 and 2012. These records include information on college attainment, non-formal education attainment, and performance in the formal Colombian labor market, including self-employed workers. The main contribution of this research is that the dataset enables us to compare every high school graduate across the country to their closest peers, i.e., classmates in their high school cohort and classmates in their college cohort. This identification of comparable pupils allows us to measure the impact of both enrollment in higher education and attainment of a college degree. Our main research questions are twofold: first, whether attending higher education makes a difference in future formal earnings, particularly in a country with a high level of informality in the labor market; and second, we aim to evaluate the real value of a college degree, including the so-called sheepskin effect.

The main empirical strategy consists of estimating a modified version of the Mincer equation. This equation includes a binary variable that takes the value of one for students who attended college and zero for those who did not. In addition to pooled estimations, we also utilize more sophisticated estimation techniques to address potential econometric issues, such as sample selection and endogeneity. Specifically, we employ the Callaway and Sant'Anna (2021) framework to estimate the Average Treatment Effect on the Treated (ATT), as well as an instrumental variables (IV) approach that utilizes the distance between high school and college as an instrument for attending college. We use a non-linear two-stage least squares (NL2SLS) estimator to obtain the Local Average Treatment Effect (LATE) of attending college. While the pooled estimations provide important context, the ATT and LATE estimations allow for more precise estimation of the causal effects of attending college on earnings.

To address the first question, we utilized the ATT and LATE results. For the second question, we compared the ATT for graduates to those of college students who completed more than 90% of the required coursework but did not receive a degree. Our IV estimations indicate that attending college increases earnings by 46.7% compared to those who did not attend college. From the ATT estimations, we found that attending the higher education premium is 40.2%, with premiums of 46.1% for bachelor's degrees, and 52.4% for diplomas. To estimate the sheepskin effect for college graduates specifically, we calculated the difference in returns between graduates (143.5%) and those with more than 90% of the classes complete but without a degree (51.1%), which averages to 92.4% for the post-college period. However, 10 years after beginning college, the returns for graduates and those without a degree are 220.4% and 62.2%, respectively, resulting in a sheepskin effect of 158.2%.

The empirical results demonstrate that graduating from higher education, especially on time, leads to increased earnings regardless of an individual's socioeconomic background. Graduates had higher earnings than both college dropouts and individuals who did not attend college. For women and low-income households, enrolling in higher education led to improved income compared to their non-college-attending peers. Women and low-income students who enrolled in college had a higher wage premium than women and low-income secondary graduates who did not enroll in college, regardless of their final status in college. Furthermore, although the gender wage gap for all secondary school graduates between 2002 and 2012 was 6.8% in favor of men, the wages of female college graduates were higher and grew faster than those of male college graduates, indicating a gradual reduction of the gender wage gap. Enrolling in higher education also benefited socioeconomically disadvantaged students, who earned a higher income (5 to 10 years after high school graduation) than their non-college-attending peers. High-income students who attended college but failed to graduate and remained active in the higher education system for extended periods earned about the same as their non-college-attending peers.

The Colombian labor market offers a premium of 76.8% to individuals with a college degree compared to workers with five years of work experience but no higher education. Interestingly, individuals who graduate late earn similar wages, five years after their secondary graduation, to those who have five years of work experience but no higher education. These findings support the conclusions drawn by Jaeger and Page (1996) that the labor market places significant value on academic preparation, particularly in the early stages of a professional career. Furthermore, our study highlights the existence of significant premiums for individuals with high cognitive skills or high income. However, we note a stagnation of self-employed individuals, although the study does not establish the underlying cause for the reduced or null premiums observed for this group. The returns to higher education in the medium to long run are positive and not statistically different across various education levels, including apprenticeship, professional, or associate programs, when compared to individuals who did not attend college. This result differs from previous studies conducted by González-Velosa et al. (2015) and Busso et al. (2020), where associate degrees were reported to have negative higher education premiums. However, we attribute this difference to the control group used in our study, which consisted of peers who had similar characteristics to those who attended community college but did not pursue higher education.

In the past two decades, Colombia has seen an increase in higher education enrollment rates and a decline in dropout rates. However, graduation rates have not kept up with these trends (Herrera-Prada, 2013; Ferreyra et al., 2017; Ministerio de Educación Nacional, 2017). This study aims to highlight the costs incurred by students who drop out after completing over 90% of their program, including wasted time and resources. Our results demonstrate the positive impact of higher education on career outcomes, as evidenced by the "Stairway to Heaven" effect of attending and graduating from college. Graduates earn considerably more than non-college attendees, and even low-income students who do not complete their degrees still earn a premium over their non-college peers. However, there is a "Highway to Hell" effect for high-income individuals who do not graduate and remain active in the education system. Dropping out, regardless of when it occurs, represents a significant opportunity cost of not graduating. Nonetheless, dropping out is still a better option than not attending college at all.

The following section presents the literature review. Section 3 describes the data and variables, and Section 4 discusses the conceptual framework and the models. Section 5 presents the results, and section 6 provides conclusions and discussion.

2 Literature Review

In this section, we examine closely related literature to identify existing gaps in research, specifically papers that examine the Colombian case. Several papers from the 1990s have explored workers' performance in the Colombian labor market and its relationship to their educational level. These papers analyzed the demand side of the labor market over different time frames from 1976 to 2000s, examining various characteristics such as returns to investment in education (Tenjo Galarza, 1993; Arias and Chávez, 2002), wages (Núñez and Sánchez, 1998), income distribution (Cárdenas and Bernal, 1999), level of education (Arango et al., 2005; Zárate, 2005) and trade openness in Colombia (Mesa and Gutiérrez, 1996; Santamaría, 2001) . Other papers studied the labor supply side of the Colombian labor market and estimated the returns to education (Rodríguez, 1981; Psacharopoulos, 1985, 1994; Prada, 2006; Mora and Muro, 2008), including studies that examined income distribution (Núñez and Sánchez, 1998; Posso, 2008) . The main findings of these studies suggested a break in the returns to education in the 1980s and 1990s due to changes in the level of education. The share of non-qualified workers in the total workforce decreased, and their salary growth was slower than that of qualified workers since the mid-1990s, which was the opposite of what was observed in the 1970s and 1980s. However, they could not determine what level of qualification triggered better salaries. A common problem faced in the above-mentioned papers was reliance on household survey data that was inaccurate and incomplete (Farné, 2006); as household surveys in Colombia did not contain enough detailed information to examine the question to which they were responding.

The Colombian Ministry of Education (MEN) created the Observatory for Educated Labor (OLE) to gather more detailed data that would allow for accurate responses to questions about the relevance of tertiary education and returns on education(Orozco Silva et al., 2011). The OLE is a follow-up survey that collects information on positions and salaries for college graduates at the individual level. It played an essential role in the growth of literature on returns to education, with Forero and Ramírez (2008) being one of the first papers to use the OLE. However, the sample size of the OLE survey was small and biased towards institutions with digital capacity before 2010. Over the years, the availability of new data allowed for significant improvements in the analysis of returns to higher education. Hernández (2010) studied the returns to education using the first release of the OLE database, which was the best approach at the time but also highlighted several structural problems with the data. Herrera-Prada and Caballero (2013) studied the returns to education to better understand how to finance the higher education system using aggregated data for tuition and expenses merged with the OLE database to estimate the expected time to recover the investment in college. Finally, González-Velosa et al. (2015); MacLeod et al. (2017); Busso et al. (2020); Ferreyra et al. (2020)and de Roux and Riehl (2022) used an advanced version of the OLE to estimate the returns to education for those who graduated from college in Colombia.

Despite the progress made in the literature, there are still gaps in research that this paper aims to address:

1. All the returns to education after college have only been estimated with college graduates, with no tracking of secondary school graduates to evaluate if attending college made any difference, despite a low attendance rate of just 52%. We follow up with all secondary graduates.

2. This paper includes self-employed workers, which are not included in the current college graduates' income database but represent about 53.1% of the Colombian formal labor force.[2]

3. This paper estimates the sheepskin effect, which previous approaches developed with household surveys did not examine in detail. This document uses detailed data to evaluate the results of students who finished more than 90% of their program but did not receive the degree compared to graduates in their jobs after college.

3 Stylized Facts, Data and Variables

This section presents some stylized facts of the higher education system in Colombia. We will then describe the five databases we use, the criteria to adjust them, and the variables we created and used from each. Finally, we will describe the data management and how we performed the data matching.

3.1 The Colombian Education System

The Colombian education system before college has about 10 million students, of which about 80% are in the public sector. It has no unified curriculum and is divided into primary (5 years) and

[2]According to (OECD 2021), the self-employed constitute 53.1% of the total population in Colombia. However, both the household survey and the social security data register only 23%.

secondary (6 years). Secondary education is divided into lower secondary for years 6 to 9 and upper secondary for years 10 and 11. The last two years differ depending on the type of secondary education (academic, military, or teacher's formation; almost all the students pursue academic degrees). During the second half of the last year (11th year), all students are screened in a test required to graduate from secondary and access college (SABER 11 exam, explained later in this paper). The college system divides its programs into two levels: associate degrees (2 or 3-year programs) and bachelor's programs (4 or 5-year programs). Note that bachelor's degree programs usually last 4-5 years and associate programs 2-3 years (depending on the curriculum that MEN approved for that HEI). To advance from associate to bachelor's, the student must graduate from the associate program before beginning the extra coursework to become a bachelor. For those students who have an associate degree and are pursuing further coursework, they are considered in an associate program until they obtain a bachelor's degree. To pursue graduate programs, a bachelor's degree is required. Graduate programs are divided into diplomas (an intermediate degree between college and master's, usually between 6 and 18 months), masters (2-year programs), and PhDs (about 5 years).

In 2002, the Colombian Ministry of Education launched the "Educational Revolution" program to improve higher education outcomes (Orozco Silva et al., 2011). As part of this initiative, in late 2004, the Ministry created the National Information System for Higher Education (SNIES), a software program that collects individual data on higher education students. The SNIES draws information from four sub-components: The SNIES database (the same name as the overarching program) contains information on the higher education system. The System for the Prevention and Analysis of Dropout in Higher Education Institutions (SPADIES) database contains student information on higher education status from the anti-dropout program. The Observatory for Educated Labor (OLE) database contains labor-related information on graduates entering the workforce. The National System of Quality Certification in Higher Education (SACES) contains information from every program on system quality improvement. The Ministry used the same names for the programs as for their underlying databases; so for example, the SPADIES database contains all the records collected by the SPADIES program.

In 2018, the Colombian higher education system had 2.4 million students, with 93% enrolled in undergraduate programs, 4% in diploma programs, 3% in master's degree programs, and less than 0.2% in PhD programs. Of these students, 51% were enrolled in public institutions and 53% were women. The system consists of 298 higher education institutions and 10,990 programs, and has an overall attendance rate of 52%. Additionally, 52 institutions have received high-quality certification. However, the system faces challenges such as a shortage of doctoral-level faculty; only 8.5% of the 162,209 professors employed throughout the system hold doctoral degrees Observatorio de la universidad colombiana, 2020. In March 2020, prior to the COVID-19 pandemic, Colombia's labor market included 20.5 million workers, with an employment rate of 51.7% and an unemployment rate of 12.6%(DANE, 2023).

3.2 Databases Description

In this sub-section 5 databases are described, from which several variables are employed in the empirical analysis.

3.2.1 ICFES

The Colombian Institute for the Evaluation of Education (ICFES) administers the Saber 11 exam required to graduate from secondary school in Colombia. Since 1968, all Colombian students finishing secondary school have taken this exam. For the empirical analysis, data from the ICFES database was used, which includes information for all students who took the Saber 11 exam between 2002 and 2012. The database contains individual-level information for each student, including the student's school, gender, household income, and exam score details, among other characteristics. The ICFES database was merged with data from the 2016 Census on Schools collected by the Ministry of Education, using the ICFES's school ID code to merge the data for a total of about 15,000 schools with secondary graduates. In merging the data, the complete profile of the school was kept, including the schedules, ownership, and administration (i.e., public or private). In Colombia, multiple schools can operate in the same building, so a secondary school is identified according to the education level. If the government does not operate the secondary school, it is considered a private school. However, a public-school building can be occupied by a public school in the morning and then by a private school in the afternoon or evening. A school administered by a private entity under a contract with the government is considered a private school. In determining the school zone, any school zone not explicitly defined as urban in the census is considered rural.

Several adjustments were necessary to standardize the exam score information in this study because the Saber 11 scoring method changed multiple times during the period under analysis. To standardize the scores, we adopted the general standardization procedure used by the Ministry of Education; sum all the subcomponents scores and with the total assigning to each student the percentile of their performance on the exam relative to the scores of the other students who took the test at the same time. As in Ministerio de Educación Nacional (2017), we created an additional dummy variable that takes a value of 1 if a student scores above 90 and 0 otherwise to differentiate students with high and low scores. The explanatory variable used in the empirical analysis is each student's decile in the total score.

The students report their gender and household income during the Saber 11 exam. However, household income was not consistently collected during all periods, as in (Ministerio de Educación Nacional, 2008) and (Ministerio de Educación Nacional, 2010). Therefore, when available, we imputed the household income using the mode for household income in the same school in comparable periods. If there were two or more modes, we used the highest value. The ICFES standardizes the income level into nine ascending categories[3]. The average and median income for the full test takers are between 2 and 3 minimum wages (Category 2). We defined a dummy variable that takes a value of one if a student has a high income (above Category 2) and zero otherwise.

In summary, we used the reported gender, Saber 11 standardized test score, household income, school location, and school sector (as defined above) from the ICFES database (See Table 3).

3.2.2 SPADIES

The SPADIES database provided information about all the secondary graduates who attended higher education between 1998 and 2017. The information includes the institution where the students enrolled, their identification information, their academic performance, the area of study, level of the program, timeframe of studies, and their status in the system (e.g., dropout, graduated, or active). Some adjustments were made concerning the status categories created for SPADIES so that the final dataset used in the empirical analysis would be more specific (see statuses descriptive statistics in Table 2):

[3] 0 "[0-1) minimum wages" 1 "[1-2) minimum wages" 2 "[2-3) minimum wages" 3 "[3-5) minimum wages" 4 "[5-7) minimum wages" 5 "[7-9) minimum wages" 6 "[9-11) minimum wages" 7 "[11-13) minimum wages" 8 "[13-15) minimum wages" 9 "[15-∞) minimum wages"

1. Graduated: SPADIES defines this status as someone who finished the coursework and received a degree from a higher education program. We divided this category into two groups:

   (a) Graduated on time: those who graduated within 1 year of the expected graduation date.

   (b) Graduated late: those who graduated more than 1 year after their expected graduation date.

   Graduated late and Graduated on time encompass all individuals that graduated from higher education. So, $Graduated = Graduated\, on\, time + Graduated\, late$. The expected graduation year is assumed as five years for bachelor programs and three years for associate programs. In Colombia, the bachelor program is also called a professional program, and the associated program is a technical program.

2. Dropout: SPADIES defines dropout as a student that has not been enrolled in the system for two or more consecutive semesters. This definition has been adjusted to be able to estimate the sheepskin effect by dividing the dropouts into two groups:

   (a) Candidates: students who finished more than 90 percent of their coursework but are reported as dropouts by SPADIES because they did not graduate after two or more consecutive semesters of not being enrolled. They are called "Egresado no graduado" in the Colombian technical language.

   (b) Incomplete: students who completed less than 90 percent of their coursework and have not been enrolled for two or more consecutive semesters. Also, to explore if the time when students leave their programs affects their income, we disaggregate the sample of incomplete students into two groups: "Early Incomplete" for those who left within their first year of college, and "Late Incomplete" for those who left after the first year and did not complete at least 90% of their program before leaving.

   $Dropouts = Candidates + Incomplete = Candidates + Early\, Incomplete + Late\, Incomplete.$

3. Active: using SPADIES definition for Active, that is, any student still active and enrolled in the system as of 2017.

Students enrolled in natural science, engineering, and math schools are coded as STEM using a dummy that takes the value of one if the students belong to these programs and zero otherwise. The program level is university (bachelor) level if the student is pursuing a bachelor's degree or associate level if pursuing an associate degree. Some colleges offer associate programs that, with extra coursework, can become a bachelor's degree.

Only data for students whose expected year of graduation from college is between 2001 and 2014 are used in the empirical analysis. This means the first freshmen cohort started in 1998, the last cohort in 2009 for bachelor's degree programs, and 2011 for associate degree programs. An active status in 2017 means the student was enrolled for at least 6 years (being a freshman in 2011 or earlier), and the lag of its graduation is at least 3 years (depending on program level).

Finally, to account for variations in the timing of college enrollment relative to secondary graduation (which may not be consecutive), we created a variable to distinguish between early enrollees (within 3 semesters after secondary graduation) and late enrollees (those who took more than 3 semesters to enroll).

Summarizing, the SPADIES database is the source for the variables of students' status in the system, the program they were doing, and the higher education institution where the student was enrolled.

3.2.3 SNIES

The SNIES database provides detailed information on higher education institutions, including their programs, levels, and locations. Data from 2017 on each program's term, level, and area of knowledge are merged with the SPADIES database using a program identification code. The Colombian Ministry of Education rates institutional quality through a process called "High-Quality Accreditation," granting a certificate of quality to institutions that meet all requirements. Institutions with this certification through 2017 are designated as high-quality institutions in the database. We create a dummy variable that takes the value of one if a student is enrolled in a high-quality institution. We also use the information on programs, institutions, and educational levels from the SNIES database. Institution information includes their certification status as a high-quality institution by the Ministry of Education.

3.2.4 OLE

Using the OLE database, we complemented SPADIES by adding the latest educational level reached by each student $Graduated = Bachelor + Diploma + Master + PhD$. It means that we could know if a student finished a higher level of education as diploma, masters, or Ph.D. Remember that SPADIES only contains data for undergraduates. The matching between the ICFES and OLE data is done using the same algorithm used in the case of the ICFES-SPADIES discussed in the subsection 3.3. With this process, we organized the graduates into the following categories:

1. Bachelor's degree. This is the group of students who graduated from higher education but did not continue their studies at the postgraduate level.

2. Diploma is a post-undergraduate degree, but less than a Master's degree; it is more flexible and has a shorter duration. It has no international recognition. At the local level, it is known as a specialization.

3. Master's degree. Regarding this category, it is essential to clarify that very few students succeed in reaching this level of education and receive salaries in the period under analysis. Reaching the master's level means at least 5 years in undergraduate studies and another 2 years in the magister program.

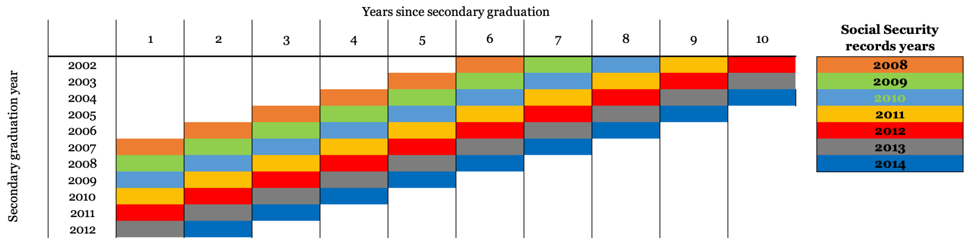
4. Ph.D. Regarding this category, it is essential to clarify that only a few students succeed in reaching this level of education and receive salaries in the period under analysis. Reaching the Ph.D. level implies at least 5 years in undergraduate studies and another 5 years in the doctoral program. Consequently, the students with a Ph.D. level that we find in the database have low salaries, and it is assumed that these salaries correspond to a partial job during their academic career. This helps to explain the results of the regression for this category. Time is not the only barrier to reaching the level; the brain drain is high. A substantial number of students who are able to attain postgraduate levels of education continue their studies outside the country. In the case of master's degrees, many students return, and this situation is somewhat different for doctoral students.

Summarizing, from the OLE database, we use the additional information about the new levels reached by the students.

3.2.5 PILA

The PILA database contains the records of all payments to Social Security for every person in the formal sector, the annual days worked, type of employment (including self-employed), and type of employer. This information is collected for the contribution of all formal Colombian workers to pension and health funds. The raw dataset is the complete list of payments in a given year attached to a national identification number. The contributions are reported per day, and people can have multiple jobs; it is possible to have people with more than 360 working days per year. Each contribution payment reported includes information on current salary, labor time reported for that payment, location of the payment, organization or individual making the payment, and economic activity of the worker. The data used for this paper includes the annual sum of all the contributions payments from all the Colombian workers that contributed to the health system from 2008 to 2014. The only identifier this database has is the national ID number, the variable used by the Ministry of Labor for the merge. Since there is no information on the tenure for cohorts pre-2006, this paper uses the difference between the year of reporting on the PILA and the year of graduation from secondary school, indicated by the date of the Saber 11 test, as a proxy for experience. As such, each ICFES cohort contributes in different periods to the normalized experience variable (See Chart 1).

Chart 1 Cohorts Distribution



Each column shows which cohorts are considered in the timing variable depending on the year of Social Security records.

The social security records identify self-employed workers using their contributor type (See Table 3). The students whose statuses in the PILA are self-employed workers, self-employed workers in an association, self-employed workers without regulations to contribute and other codes for the transition from employee to self-employed (code 42 and code 49) were marked as self-employed. Public employees were identified by code 3 of the variable for the type of employer (See Table 3). It is essential to take into account that the State is a large employer, but not all its workers are registered in the PILA [4] (Descriptive statistics of PILA in Table 1).

From the PILA database, we use the annual income, the annual labor supply, and the information related to the public servants or the self-employed workers.

[Table 1 about here.]

3.2.6 SENA

The National Apprenticeship System (SENA) was established by the government in 1957 to provide job training to workers at no cost to employers and to mitigate the risk of investing in employee training. Although SENA training programs were not initially considered formal education, they became highly specialized and were eventually recognized as valuable qualifications for employment. As a result, SENA training programs effectively became professional programs, with many students using their training as a source of income. In 2003, SENA began offering formal professional programs based on their existing courses and became partially integrated into the higher education

[4] Teachers and military are examples of groups not reported in the social security system.

system. Before 2013, SENA was not included in the SNIES database and has only reported data to the higher education system in a limited capacity since then. Students enrolled in SENA's programs were referred to as apprentices and, as a final part of their training, were required to complete internships with participating companies. The programs were divided into lectures and practical components, with a total duration ranging from 12 to 36 months. These companies were required to pay the apprentices between 50% and 75% of a minimum monthly salary and report them to the pension and health funds (PILA) as SENA apprentices. We created a variable for apprentices using this data. It is essential to make two observations:

1. The students who are marked as SENA's apprentices are the few who arrive in the formal labor market and are the comparable ones with the higher education students (see Table 2). At the same time, it is essential to remark that their job is likely to be part of the informal sector because the kind of work they perform has been traditionally informal in Colombia. These students may be the most qualified labor force in the informal labor market

2. We can only track secondary school graduates who went to SENA in the final stages of their training that are similar to the associate programs time line (when they are working as interns), and we do not have any information about any other apprentice's status.

[Table 2 about here.]

[Table 3 about here.]

3.3 The Administrative Data Matching Process

The search algorithm[5] used by the Ministry of Education of Colombia is employed to merge the above-mentioned databases. The full name and the date of birth are the two-parameter used to match the information for single individuals. The ICFES dataset (5,425,850 students) is merged with

[5]The algorithm takes two key variables, namely the full name and the date of birth, from the databases. Firstly, the algorithm removes the spaces, converts all alphabetic characters to uppercase, and then decomposes the strings into all possible combinations of the characters. For instance, the name "Tom" is transformed into TOM, MOT, OTM, OMT, TMO, MTO. Next, the algorithm compares each discomposed key variable for every observation in each database to all possible observation matches between the databases. If the comparison reaches a certain "trigger" level, the algorithm identifies the observation as a match. The level of match is the percentage of similarity between the discomposed variables. The algorithm is cautious, meaning that if there is more than one potential matching option, it will not execute the matching. In this paper, the trigger value used is 98%, the same as the value used by the Ministry of Education in the SPADIES-ICFES match.

the information from the SPADIES (7.2 million students), obtaining 2,764,503 matched observations and with the PILA database (16.8 million workers) using the individual's ID number (Table 4).

The final database holds information for 418,699 students in an unbalanced panel of students and years of appearance in the PILA (2008-2014), accounting for a total of 2,074,267 observations. The variables from PILA are time variants, including income, labor supply in days and years after secondary, a dummy variable for self-employment, and another for public servant status. The years after secondary are proxied by the difference of the cohort of presentation of the Saber 11 test and the year of entry in the PILA database. The time-invariant variables are from the ICFES, SNIES, and SPADIES databases, including status, gender, Saber 11 test score, location and sector of the school, and household income when the Saber 11 exam was taken.

[Table 4 about here.]

4 Conceptual Framework and Model

This section first presents the theories that conceptualize the returns to education and the sheepskin effect. Next, the empirical model specification is discussed. A modified Mincer equation, which is the base of the theoretical framework used to derive the empirical model, is used to estimate the returns to education and the sheepskin effect.

4.1 Conceptual Framework

An extensive body of literature examines the relationship between human capital formation and the benefits of education. Human capital refers to intangible assets, such as education, which improve earnings, habits, and health. Investment in these assets is critical, as they must be kept from their owners (Becker, 1962). Becker's theory suggests that schooling increases human capital, which, in turn, increases earnings up to a certain point. In other words, individuals earn more as they acquire more education until the opportunity cost of obtaining it outweighs the marginal utility of an additional year of education. This process is illustrated in Mincer's seminal contribution (1974). However, subsequent literature has challenged the relationship between years of schooling and labor market earnings. Researchers argue that education is a signal of qualification or even a filter, and employers often face asymmetric information that makes it challenging to select the best employee (Phelps, 1972; Arrow, 1973; Spence, 1973).

Phelps (1972); Arrow (1973); Spence (1973) propose that tertiary education signals quality or skill, creating a shortcut in hiring. In this process, employees use their degrees to signal their preparation and skills, while employers signal to the market that they value such degrees and would prefer to hire individuals with them over those without. This helps reduce the level of asymmetric information about the true fit or ability of the employee for a given employer. Interestingly, while some potential employees may already possess the necessary skills, they still pursue education to signal their qualifications. Wood (2009) suggests that high salaries are paid to reward those who obtain an education, but this also increases the opportunity cost of continuing education, particularly for those who already have high skills. Thus, highly skilled workers can prepare and move up the education ladder faster and farther than less skilled workers. In contrast, less skilled workers may soon realize that they cannot compete with skilled workers while studying and that the opportunity cost of remaining in the system is too high, leading them to drop out.

According to Collins (1979), the pursuit of social mobility drove students to obtain degrees and credentials to secure better jobs, increasing the number of available graduates. This rise in the supply of qualified workers caused employers to increase job requirements, such as demanding certain degrees, grade point averages, or coursework, to select candidates. The increased number of available graduates also had unintended consequences, such as school grade inflation and rising education costs. For instance, schools started charging extra fees for degrees. These second and third-order effects help explain why the broader push for more education failed to increase social mobility. Only those who could afford multiple degrees, tutors, and additional school fees reaped the benefits. The privilege associated with these credentials was significant enough to prevent some individuals from competing in the job market. Those with the right credentials obtained better jobs and higher salaries, while others were left behind.

The value of a credential has become more important than the actual knowledge and skills acquired, as under the human capital framework, two individuals with the same education should earn the same salary, regardless of whether one obtained the degree or not. However, research shows that salaries rise faster for individuals with more education and a degree. This phenomenon was termed the sheepskin effect by Hungerford and Solon (1987), who estimated a Mincer Equation with a discontinuity in the years of higher education. Their findings reveal significantly higher gains compared to the previous year (8, 12, and 16 years of schooling). However, data limitations explain why it took some time to recognize the true phenomenon in related literature. From the late 1960s until the early 1980s, researchers only had access to basic information on individuals' backgrounds, years of education, and earnings. In the 1980s and 1990s, information on degree completion became available, and after 2000, more sophisticated information on individuals' cognitive skills, earnings, and degrees became accessible.

With the availability of more sophisticated data, researchers were able to directly test the credentials theory and the sheepskin effect. This allowed authors to use these theories as a framework and conduct case studies in various countries. For example, Shabbir (1991) was able to differentiate the impact of master's and basic education on resource allocation. Belman and Heywood (1991, 1997) and Jaeger and Page (1996) analyzed how the sheepskin effect affected women and men in minority groups and found that the signal fades over time as workers gain experience. Bilkic et al. (2012) analyzed at what point the opportunity cost of the credential becomes relevant for workers to continue studying or enter the labor market. Other case studies include Gibson (2000) for New Zealand, Ferrer and Riddell (2002) for Canada, Mora (2003) , García-Suaza et al. (2014) and Bacolod et al. (2021) for Colombia, Schady (2003) and Olfindo (2018) for the Philippines, Bauer et al. (2005) for Japan, Calonico and Nopo (2007) for Peru, Crespo and Reis (2009) for Brazil, Son (2013) for Indonesia, and Yunus (2017)for Malaysia. The main findings of this new strand of literature are twofold. First, the labor market considers years of schooling as relevant work experience. Second, the returns for each additional year of schooling are small compared to the difference in returns for degree holders versus non-degree holders.

4.2 Model Specification

4.2.1 The Basic Model

To estimate the returns to education and the sheepskin effect, a modified Mincer equation is used as main framework for the empirical analysis. Based on Hungerford and Solon (1987), the model is specified as:

$$Eq:1 \qquad y_{it} = \alpha + \sum_{j\in J} \gamma_j 1 Status_i = j + \beta X_{it} + \vartheta_i + \epsilon_{it}$$

$J = \{Graduated, Candidate, Incomplete, Active\}$,

where y_{it} is the annual income expressed in logarithm for the student i in the year t. The variable of interest is γ_j. The statuses were defined in the data section and are a set of dummies from the set

$J = \{Graduated,\ Candidate,\ Incomplete,\ Active\}$, where y_{it} is the annual income expressed in logarithm for the student i in the year t. The variable of interest is γ_j. The statuses were defined in the data section and are a set of dummies from the set

$J = \{Graduated,\ Candidate,\ Incomplete,\ Active\}$ take the value of one if the student has a certain status j, and the base level is given by those students that did not attend higher education. The vector of controls X_{it} includes time variant variables. Among them, dummy variables for apprenticeship, self-employment and public servant status, and the years elapsed after completing secondary school. This vector also includes some continuous time invariant variables, such as the standardized test score in the Saber 11 exam and dummy variables for urban area, public school and whether household income when the Saber 11 test has taken is higher than 3 monthly minimum wages, the year of graduation from secondary school, the department[6], the secondary school id code, and the higher education institution code (if attended college).

We use an extension of this model to test if the sheepskin effect results are due to skills and experience as in the empirical model of Farber and Gibbons (1996). To do so, the dependent variable y_{it} is expressed in levels not as logarithm[7] and the two new interactions included in the vector X_{it} should be positive and significant. The interactions included to the vector are: 1) the interaction of the score and the dummy of graduate; 2) the interaction of the years of experience and the dummy of graduate.

We also extend the results of Equation 1 by expanding the Graduated in the set J to graduates on time and graduates late, or bachelor, diplomas, master and PhD as: $Graduated = Graduated\,on\,time + Graduated\,late = Bachelor + Diploma + Master + PhD$

[6] Colombia is a unitary republic conformed by 1,123 municipalities in 32 departments. Located between the nation and the municipality, the departments are autonomous administrating the resources granted by the State.

[7] Farber and Gibbons (1996) use levels and not logarithm because their "*theoretical model provides clear implications for how education and other variables are related to the wage level.*".

4.2.2 Advanced Models

The basic model (Equation 1) provides an opportunity to compare the results with the existing literature, gain insights from the data, and move closer to answering the questions posed in this paper. This model falls short of inferring causality despite its consistency with previous research and relevance to our research questions. More crucially, it is important to note that one of the significant challenges in models exploring the link between education and earnings is the non-random assignment of education. Individuals consciously decide to pursue a certain level of education based on their assessment of the opportunity cost involved (Becker, 1962; Card, 1993; Wood, 2009). We must use more sophisticated estimation techniques that account for omitted variable bias or simultaneity concerns to address potential econometric issues like sample selection and endogeneity. To do so, we use the Callaway and Sant'Anna (2021) framework to estimate the Average Treatment effect on the Treated (ATT) and the Instrumental Variables approach to estimate the Local Average Treatment Effect (LATE) of attending college.

Heterogeneous Difference in Differences (DiD) Approach The canonical Difference-in-Differences (DiD) approach employs a 2X2 model that considers two periods and two groups. In the first period (T=0), the two groups are identical in terms of treatment exposure, as neither group receives the treatment. In the second period (T=1), a portion of individuals receive the treatment, creating a "treated" group (Went to College=D=1), while those who do not receive the treatment are deemed "controls" (Went to College=D=0). Assuming that the treated group would follow a predetermined path absent the treatment, any deviation from this path can be attributed to the causal effect of the treatment on the group. This deviation or difference is the ATT (Equation 2).

$Eq : 2 \qquad ATT = E(\tau_i|D_i) = E(Y_{i,1}|D_i = 1) - E(Y_{i,1}(0)|D_i = 1)$

However, given that the path the treated group would have followed in the absence of treatment is unknown, the assumption is made that this path would be parallel to the path followed by the control group, a concept known as the Parallel Trend Assumption (PTA). Nonetheless, this assumption can be difficult to fulfill in practice, as the treated and control groups may not possess similar characteristics. As such, Callaway and Sant'Anna (2021) have suggested generalizing the canonical approach by including additional groups and fixed effects in the specification.

Moreover, DiD designs often feature more than two periods or more than two treated groups, which can further complicate the PTA assumption. To address this issue, Sant'Anna and Zhao (2020) propose using the PTA for groups with identical pre-treatment characteristics as the vector of controls X from Equation 1, thereby reducing the risk of bias due to differences between treated and control groups (Equation 3). Where $\theta(X)$ is the ΔY_i if there was no treatment conditional to X. With this new assumption, the new DiD estimator becomes $\widehat{ATT}_*$ (Equation 4).

$Eq:3 \qquad E(Y_{i,1}(0) - Y_{i,0}|D_i = 1, X) = E(Y_{i,1} - Y_{i,0}|D_i = 0, X) = \theta(X)$

$Eq:4 \qquad \widehat{ATT}_* = E(Y_{i,1}|D_i = 1) - \left[E(Y_{i,0}|D_i = 1) + \widehat{E}(\theta(X)|D_i = 1)\right]$

Various approaches have been proposed in the literature to estimate the component $\widehat{E}(\theta(X)|D_i = 1)$ from Equation 4. In this paper, we adopt the Improved Doubly Robust (IMP) estimator proposed by Sant'Anna and Zhao (2020). The IMP estimator employs a two-step procedure. In the first step, the estimator obtains an estimate of $E(\theta(X)|D_i = 1)$ using only control data and without any weighting. In the second step, the estimator adds a correction term Λ, which captures the difference between the predicted and the observed outcome in the control group, weighted by the inverse probability of treatment weights. Specifically, this correction term is calculated as the difference between the predicted and observed outcomes in the control group, weighted by the inverse probability of receiving treatment among the treated individuals. The detailed formulation of the IMP estimator is provided in Equation 5.

$Eq:5 \qquad \widehat{ATT_{IMP}} = E(\Delta Y_i|D_i = 1) - E(\widehat{\theta}(x_i)|D_i = 1) - \Lambda Where \Lambda = E(\omega(x_i)\Delta Y_i|D_i = 0)/E(\omega(x_i)|D_i = 0) - E(\omega(x_i)\widehat{\theta}(x_i)|D_i = 0)/E(\omega(x_i)|D_i = 0)$

Finally, college enrollment occurs in different cohorts, meaning the treatment is given at different times and to different groups. To address this issue, the framework proposed by Callaway and Sant'Anna (2021) fixes a group g (which represents the cohort of enrollment in college) and allows variation in t, in order to understand how the proposed Average Treatment Effect on the Treated (ATT) evolves over time for a specific group (Equation 6). In our estimation, we use Rios-Avila et al. (2021) methodology which disaggregates the combinations of groups and times into multiple 2X2 models that are then aggregated per the fixed group g.

$Eq:6 \qquad ATT(g,t) = E[Y_t(g) - Y_t(0)|G_g = 1]$

After this process, an ATT and weights are calculated for each period group, allowing us to consolidate the ATT by time (similar to an event analysis as we report in the results section), and by group to analyze impacts per group and make comparisons. As mentioned earlier, the groups can have different times, so under this framework, the previously divided population into two groups (treatment and control) is now sorted into three sets: treated, not yet treated, and control.

As students may delay their enrollment in higher education and may need to work prior to enrollment, our control group contains those who have not yet been treated. This framework allows us to estimate the causal effect of college enrollment for each cohort and analyze how this effect changes over time.

Instrumental Variables Approach Instrumental variables are a suitable approach when proper instruments for potential endogenous variables are available. This study uses the non-linear two-stage least squares (NL2SLS) estimator to address these concerns. Specifically, we use the distance from secondary school to college as an instrument for education attainment. Since finding instruments for the entire set of variables is difficult, we instrument the "Attend" status, which indicates those who attended college. This approach allows us to better understand the relationship between education and earnings while addressing potential sources of bias in the data.

Using now “Attend” as main target variable, the model specification is given by:

$Eq: 7 \qquad y_{it} = \alpha + \gamma Attend_{it} + \beta X_{it} + \vartheta_i + \epsilon_{it}$

Since “Attend” is a dummy variable, a probit model is estimated to predict the probability of attending college using the instrument (distance school to college) as an independent variable. The vector of controls X_{it} is the same as before. The first step equation is specified as:

$Eq: 8 \qquad Attend_{it} = \eta + \lambda Distance + \mu X_{it} + \vartheta_i + \epsilon_{it}$

The distance to college has long been utilized in the literature for similar purposes since it was first proposed by Card (1993). The intuition is that proximity to a college can significantly impact a student’s decision to pursue higher education. The distance to college may influence the likelihood of attending college, but it is not necessarily related to one’s ability or wealth (Card, 1993; Frenette, 2004).

Unlike in other countries, Colombia does not have university cities that are dependent on college campuses. Therefore, high schools and colleges are located randomly in the cities, and families of varying income levels can be found within any radius from secondary schools or colleges. Finally, distance to college is not related to wages, as individuals with different income levels can be found at any point within the same distance radius.

In Equation 9, we use the estimated probability of attending college as an instrument for "Attend." As the distance from the secondary school to college satisfies the "exclusion restriction," the exogenous variation provided by the instrument in the instrumental variables (IV) approach gives a precise local average treatment effect (LATE). Therefore, the results in Equation 9 can be interpreted as the causal effect of attending college on future salaries.

$$Eq:9 \qquad y_{it} = \beta + \gamma \widehat{Attend}_{it} + \beta X_{it} + \vartheta_i + \epsilon_{it}$$

5 Results

In the first part of this section, we analyze the results for Equation 1 and Equation 9. The results from the IV approach (LATE estimation) are presented and discussed. These results are compared with the results obtained from the pooled OLS model, given by Equation (1). Finally, the heterogeneous difference in differences results (ATT estimation) are presented graphically in aggregate and considering different groups of individuals.

5.1 Main Results

The estimated coefficient for "Went to college" in Equation 1 is 0.05, suggesting that individuals who attend college earn 5% more than those who do not. However, this estimate may be biased due to omitted variable bias or sample selection issues. To address this, we estimated Equation 9 using distance from school to college as an instrumental variable. The estimated coefficient (LATE) for college attendance using this approach indicates that attending college is associated with a 46.7% increase in income compared to those who did not attend college after completing secondary school.

[Table 5 about here.]

5.2 Disaggregated Results

In this section, we report the results for Equation 1 for different disaggregations. Estimations are not causal, as findings from Table 5, but their results provide interesting descriptives to understand how students who attend college reach an income 46.7% higher than those who never attended. Results from the pooled analysis can be interpreted as a conservator lower bound.

In column (1) and column (2) in Table 6, we show the results on returns to education of students who went to college compared to students who graduated from secondary school but did not go to college using Equation 1. Differences in columns are the fixed effects used; column (1) has department fixed effects, and column (2) has the secondary school fixed effects. Columns (1) to (6) also have the secondary graduation year as a fixed effect. Thus, we find that the returns to education for students who graduated from college reach 18.4% if the fixed effects are by the department and 15.6% if the fixed effects are by the secondary school. The return to education for candidates is 1.9% with the department fixed effects and 3% with secondary school fixed effects. Students with incomplete status do not significantly differ from their peers who did not enroll in higher education. They earn only 1.6% more than those who did not attend college when using the secondary school fixed effect specification. Column (3) shows the results of Equation 1, but this time with a higher education institution's fixed effect. We use this equation to make a more accurate measure of the sheepskin effect. We find that the returns for those who go to higher education and graduate are 20.4% higher than the base group, which are the students with status "Active". Candidates, on the other hand, have a salary only 3.2% higher than the base group. The difference between these coefficients means that the average sheepskin effect is 17.2%.

In Table 7, we analyze the results for postgraduates. We find that the returns for students who only hold a bachelor's degree are 17.9% when the model includes the department's fixed effects and 15.1% when the model includes fixed effects by the secondary school. Wages for students that earn a Diploma are 39.5% higher in the specification with department fixed effects and 33.6% using the secondary school fixed effects' specification. On the other hand, those who attain a Master's degree report a salary 64.8% and 53.2% higher than those who did not attend higher education in the department and college fixed effects specifications, respectively. Although we included the Ph.D. graduates in the regression, we do not consider PhD.s in the analysis as the secondary school graduates that we track and reach this level are small in number. Their results are not shown (Table 7).

Our objective in columns (4) to (8) is to exploit the potential of the data and to corroborate the results of Farber and Gibbons (1996) regarding the correlation between experience and wages and skills unknown to the market and wages. The regression we present is specified in levels, and it corresponds to the equivalent one conducted by the authors. In our analysis, we find that experience (measured as years since secondary school) continues to be significant. However, unlike the results presented by Farber and Gibbons (1996), we find that education and experience have a positive and significant relationship in Colombia. In columns (4), (5), and (6) of Table 5 and Table 6, we test whether cognitive skills measured by the Saber 11 test's score and the experience are positively related to income. In this specification, we added two variables: i) the interaction of test score and having obtained the degree (Column 4) and ii) the interaction of being a graduated with years after high school; this is the replica for the results from Farber and Gibbons (1996) (Column 5). A third specification has the two new variables at the same time in the Column (6). With the positive and significant results in all three cases, we have evidence to think that graduates' experience, education, and skills have a high relationship with wages. These results mean that the students' years of experience and skills will compensate for the imbalances that could exist with the entry salaries, as the more skilled will have a steeper earning slope than the less skilled as they accumulate experience.

Columns (7) and (8) of Table 5 and Table 6 show a robustness test using only the secondary graduates' cohort of 2002. We use the same specification presented in column (6), but controlling with the higher education institution's fixed effects where the students attended and the secondary school. The results we found on the correlation of skills and experience with the salary reported in column (6) remain the same.

[Table 6 about here.]

[Table 7 about here.]

5.3 Heterogeneous Difference in Differences (DiD) Results Analysis

In this subsection, we aim to analyze the yearly dynamics of the higher education premium since high school graduation. To accomplish this, we will employ the heterogenous difference-in-differences (DiD) results from Equation 6 using the CSDID command by Rios-Avila et al. (2021).Our primary focus will be on the values of $T > 5$, located on the right-hand side of the value 5 in the X-axis. We approximate $T = 5$ as the completion of the academic program. Notably, the values reported during college are comparatively lower than those observed in the control group due to the fact that the students do not work full-time, as illustrated in Figure 1. Additionally, the delay in getting the degree can also be attributed to the students studying and working simultaneously. It is also possible to estimate a benchmark for comparison, as the starting salary for those without formal sector experience (represented by the zero line) should be equivalent to the minimum wage, while the earnings of working students would likely come from part-time or hourly work.

The first part will present the general results, and the second part will report the results by splitting the population into three types of covariates. The first group includes those collected during the Saber 11 test, such as gender, score, household income, school area, and school sector. The second group includes those collected during reporting to social security. The third group includes those collected during tertiary education, such as program level, higher education institutions' quality, and program area. For the first and second groups, the comparison is against all the same groups not enrolled in higher education. For the third group, the comparison includes all remaining students, those who went to college and those who did not. Thus, for example, women are compared to women not enrolled in higher education, but STEM is compared to all non-STEMs, including those who never enrolled in higher education.

5.3.1 Main Results

In the ideal scenario, college students would not be engaged in work activities before and during their academic program. However, the economic context in Colombia and the need to obtain financial resources to pay for their studies may explain the prevalence of employment during these periods. A clear pattern emerges from the graph during the college years, indicating that individuals who work less or those who are less compelled to do so are more likely to successfully complete their academic program (Figure 1). Additionally, we found evidence of six relevant facts:

1. The sheepskin effect, which measures the difference in earnings between graduates and candidates, is found to consistently increase after graduation from college, according to Figure 2. On average, the sheepskin effect is 92.4% for all graduates, as calculated by subtracting the post-college average earnings of non-graduates from the post-treatment average earnings of graduates in Figure 3. Additionally, the sheepskin effect is higher for females, those with higher income levels, those with higher academic skills, those who attended secondary schools outside Bogota, and those who attended private schools. Moreover, the sheepskin effect in-

creases over time, reaching an average of 158.2% at year 10 after secondary school graduation, as shown by the subtraction of the average earnings of graduates from the average earnings of non-graduates at 10 years after treatment in Figure 3.

2. In the Colombian labor market, there is an average premium of 158.2% during the 5 to 10 years after freshman year for graduates who completed their education on time, compared to workers with the same level of work experience but without higher education (Figure 8). This indicates a significant advantage in terms of wages and career progression for those who have obtained a college degree within the expected timeframe. However, for those who graduate late, their earnings in the fifth year after secondary graduation are similar to those of workers with five years of experience but no higher education. These findings confirm Jaeger and Page (1996) previous research, which suggests that the labor market values academic preparation as much, if not more, than experience in the early stages of the professional career (Figure 4).

3. The benefits of graduate education surpass those of students who discontinue their studies after completing a college degree. The results are particularly apparent in the case of specializations, given the limited number of students who complete masters and doctoral studies within the timeframe of this study. However, despite the wide range of results for master's and doctoral degrees, the benefits of such degrees are remarkable, with returns of 307.8% and 413.8% after 6 and 10 years, respectively (Figure 5 and Figure 6).

4. Regardless of the timing of dropout, there is no difference in returns for those who drop out of college. First-year dropouts, late dropouts, and candidates have statistically similar salaries after year 6. However, students who drop out early receive a premium for attending college and earn more than their non-college peers after year 4. The mere fact of having attended college provides a premium in future returns, regardless of whether the students dropped out or when they did so. The initial increase in returns for dropouts, which equals those of graduates in their graduation years, can be explained by a positive signaling of their quality compared to their non-college peers. However, when graduates begin to enter the labor market, this signaling fades, leading to a higher sheepskin effect (Figure 7).

5. The returns to higher education in the mid-long run are not statistically different across fields

(STEM or non-STEM) or levels (Apprenticeship, Professional, or Associate Programs) when compared to individuals who did not attend college. However, there are some small differences in income for the years leading up to year 7 after completing the program. Associate programs have a higher premium during this period due to their shorter duration (finishing as early as year 3) and the combination of experience and degree. Notably, students in Apprenticeships exhibit a rapid recovery, since they receive a fraction of the legal salary while simultaneously studying and working during their programs (up until year 3). However, after year 7, all levels have statistically the same premium compared to their non-college-educated peers (see Figure 9, Figure 10, and Figure 11). The quality of higher education institutions is a crucial factor in determining the economic benefits of attending college. Empirical evidence suggests that students who attended certified institutions, reported higher earnings premiums compared to those who attended non-certified institutions (see Figure 12).

6. The analysis reveals that the starting point of the premium after college is the same for all students, regardless of their time of enrollment. However, students who enrolled soon after graduating from secondary school exhibit a higher premium (92.4% compared to their peers) than those who enrolled late (54.9% compared to their peers) until year 8, when both premiums become statistically similar again. Early enrollment may signal high skills that students convey to the labor market. However, similar to the boost received by on-time graduation, this boost diminishes over time, and by year 10, their premium becomes lower (100.9% compared to their peers) than that of students with late enrollment (143% compared to their peers) (Figure 13 and Figure 14).

5.3.2 Results by Covariates

The findings from Table 5 and Table 6 suggest that there exists a gender wage gap, with men generally earning more than women. However, the premium for higher education is substantially higher for women compared to men, particularly in the early stages of their professional careers. Interestingly, the long-term returns for males are similar to the early premiums for females, with females experiencing a significant increase in their premium by year 10 (reaching 301.7%) while males experience a slower growth (reaching 134.9%). In addition, the study reveals that the sheepskin effect, which describes the increase in earnings associated with completing a college degree, is more pronounced for women than for men. This can be attributed to the fact that even women who did not complete their college degree tend to have a higher premium than men who did complete their degree. Conversely, men who attended college but did not graduate tended to earn the same as their peers who did not attend college (Figure 15 and Figure 21).

Regarding Figure 16, which investigates the correlation between Saber 11 test scores and earnings, the findings indicate that graduates with high scores experience a premium of 240.5% a decade after completing their enrollment in higher education. Conversely, those graduates with low scores receive a premium of 176.5%. Furthermore, the sheepskin effect is more pronounced for high-skilled students, with a sheepskin effect of 104.9% for high-skilled individuals and 60% for low-skilled individuals in the years post-college. It is worth noting that all students with high skills attain the highest returns among all categories if they are not active (Figure 16 and Figure 21).

Based on the findings presented in Figure 17, it can be inferred that household income has a more significant impact on future premiums compared to skills income. The data indicates that household income accounts for larger variations among different categories, but all categories still receive premiums compared to peers who did not attend college. However, the results also highlight a distinct redistributive characteristic in the market. Despite the presence of the sheepskin effect, which remains at approximately 107.9% for students from high-income households and 64.2% for students from low-income households in the post-college period, students with low incomes but no college degree still earn higher income than their non-college attending peers. Students with high household incomes who attended college but did not obtain a degree consistently earn more than their peers who did not attend college, and their premium is higher than those with low incomes. This phenomenon can be attributed to social influence, where unsuccessful students with high household incomes are able to access preferred job markets (Figure 17 and Figure 21).

Analyzing Figure 18, it is observed that private secondary school students who enroll in college but fail to obtain a degree within five years of completing their secondary education have earnings that are statistically indistinguishable from their peers who did not attend college. However, having a college degree yields returns of approximately 234.7% by the year 10 and a sheepskin effect of 108.5% for the post-college period. In contrast, obtaining a degree from a public secondary school results in similar returns for those who did not complete their degree and lower returns compared to students from private secondary schools. These findings suggest that social connections or other external factors may be playing a role in the long-term outcomes of students from private secondary schools (Figure 18 and Figure 21).

Examining the results presented in Figure 19, we observe that there is no significant difference in the premium among students who did not complete their college program, regardless of the region, until year 8. However, the graduates from Bogota experience a faster increase in their premium compared to those outside of the region, although their premium becomes lower compared to the return for graduates outside Bogota, 10 years after beginning college. While active students in Bogota report some positive premium after 5 years of college, outside Bogota, their income is similar to those who did not attend college. After 10 years, the income for candidates from outside Bogota is comparable to the income for students who did not attend college. Consequently, the sheepskin effect is 74.8% for students from Bogota and 89.5% for students from the rest of the country. It is noteworthy that in Bogota, the mere fact of attending college confers a boost to students' future salary, irrespective of their final status (Figure 19 and Figure 21).

The findings presented in Figure 20 indicate that self-employed college graduates are the only group that reports any sort of premium in comparison to non-college educated workers. However, the premium is quite low at only 20%. For all other categories, the income reported is similar to that of non-college workers. It is important to note that the results for self-employed individuals should be interpreted with caution, as this category may also include students with informal, part-time jobs that may have underreported income in the formal sector.

Finally, Figure 21 summarizes the main results for γ_i when $\tau = 10$ from Equation 6 on graduates and candidates outcomes and the sheepskin effect showed in Figure 15 to Figure 20.

6 Discussion and Conclusions

In the first part of this section, we open a discussion and provide some suggestions to the public policy makers based on the results of this document. In the second part of the section, we analyze the most important results from the previous section.

6.1 Discussion

This paper sheds light on the benefits of pursuing higher education in Colombia, demonstrating the positive impact on future formal sector income for students who attended at least one semester of college compared to their peers who did not attend college. Regardless of household income, academic skill level, program type, or field of study, students who attended college reported higher future income solely by virtue of their attendance. The study also reveals significant premiums for those with high cognitive skills or high income, but notes a stagnation of self-employed individuals. However, the study does not identify the cause of the reduced or null premiums for self-employed workers.

These findings raise several questions, such as whether self-employed workers underreport their actual wages and only report the minimum required for health and pension systems, or whether the labor market for contractual workers is undervaluing the added value of a college education. Furthermore, it is unclear whether self-employed workers in the formal market are considered part of the informal sector or whether they are informal workers in a formal market. This issue warrants further research.

Overall, this paper contributes to our understanding of the advantages of higher education in Colombia and highlights the need for continued research to fully understand the complexities of the labor market and how they intersect with education and income.

The findings of the study underscore the importance of obtaining a higher education degree, particularly when done on time, in increasing income levels regardless of socioeconomic status. Pursuing higher education alone represents an improvement over peers in income for socioeconomically disadvantaged groups. Colombia must urgently address the quality and articulation of community college and apprenticeship programs with the labor market. The study's results are interesting, particularly as the premium in the medium to long run is very similar to the professional programs. However, the incentive to pursue professional programs over community college or apprenticeship programs may arise due to the higher salaries offered by professional programs.

Moreover, Colombia must increase the college graduation rate, especially for on-time graduation, to enhance income opportunities for individuals, irrespective of their socioeconomic background. Increasing college graduation rates will also open up opportunities for further academic achievement at the master's and Ph.D. levels, which have demonstrated important improvements in the future income of students, albeit with limited data available to track their progress.

6.2 Conclusion

Based on the results presented in this document, attending higher education can be a life-changing experience in terms of income, especially for those who obtain their degree on time. Furthermore, it represents a significant improvement for female secondary graduates, or graduates with low household incomes or who attended public schools. Even students who only attended college for some time earn more than 50% than those who never went to college, and those who complete their degree report up to over 300% higher income. These returns are similar to those reported by students with advanced degrees, as individuals with a master's degree earn 307.8% (after 6 years) or a PhD earn 413.8% (after 10 years) more income than those who never attended college.

The study reveals a consistent increase in the sheepskin effect after graduation from college, with higher effects observed for females, individuals with higher income levels, those with better academic skills, those who attended secondary schools outside Bogota, and those who attended private schools. The sheepskin effect reaches an average of 158.2% at year 10 after secondary school graduation, highlighting the long-term benefits of pursuing higher education in Colombia.

The Colombian labor market rewards individuals with a college degree with a premium of 76.8% on graduates on time over workers who possess five years of experience but lack higher education. However, individuals who graduate late experience similar earnings in the fifth year after secondary graduation as those who have five years of work experience but no higher education. These findings support the research of Jaeger and Page (1996), suggesting that the labor market values academic preparation as much, if not more, than experience in the early stages of a professional career.

The benefits of graduate education in Colombia are found to surpass those of students who discontinue their studies after completing a college degree, particularly in the case of specializations. The limited number of students who complete masters and doctoral studies within the timeframe of this study implies that the benefits of these degrees may be even greater. Interestingly, regardless of the timing of dropout, there is no difference in returns for those who drop out of college. First-year dropouts, late dropouts, and candidates have statistically similar salaries after year 6. However, students who drop out early receive a premium for attending college and earn more than their non-college peers after year 4.

The results show that a college can bring small but redistributive income improvements for socially disadvantaged individuals, and for women who graduate from college, it can reduce the gender gap. Students from high-income households may report similar income to their peers who did not attend college if they remain active for more than 5 years. However, low-income students or students from public secondary schools who attend college experience positive returns just by attending.

Regarding gender, females experience a higher premium than their female peers who did not enroll in higher education. Female college graduates earn higher income premiums and earn them faster than male college graduates, suggesting slow but steady improvements in reducing the gender gap.

Overall, the data suggests that higher education provides significant advantages in terms of future income, particularly for those who graduate on time and pursue graduate education. The sheepskin effect remains consistent and significant over time, with greater premiums observed for females, those with higher academic skills, and those who attended private schools or secondary schools outside Bogota. Additionally, there is evidence of a redistributive effect in the labor market, as even low-income students who attend college report higher future income than their non-college peers.

However, household income also plays a significant role in determining future premiums, with students from higher-income households consistently reporting higher premiums compared to their peers than students from low-income households. This suggests that there may be social and economic barriers that prevent low-income students from fully realizing the potential benefits of higher education. Further research is needed to understand the underlying factors driving these differences and to identify strategies to address these disparities.

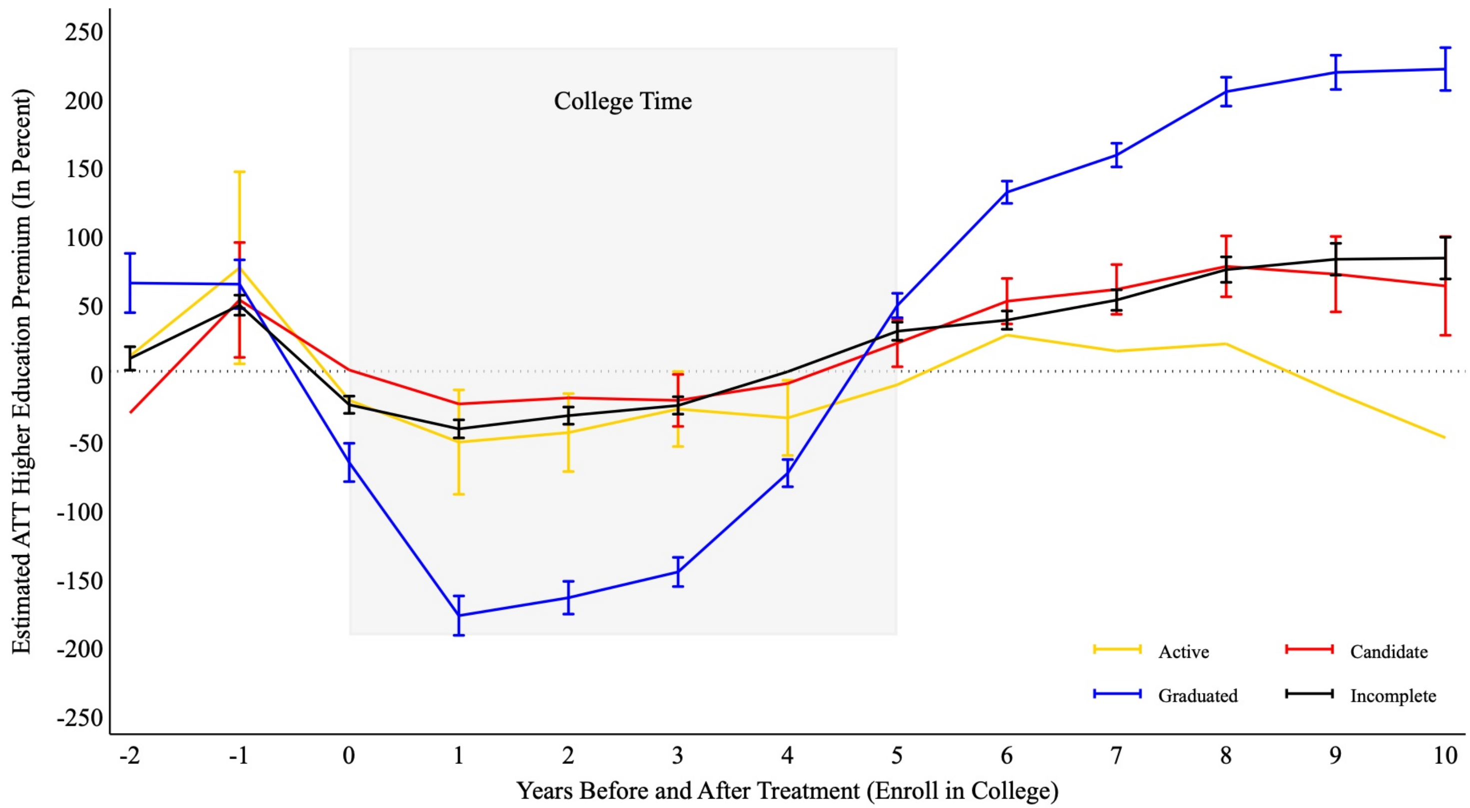
The returns to higher education in the mid to long run are positive and not statistically different across levels, including apprenticeship, professional, or associate programs, when compared to individuals who did not attend college. This result differs from previous studies by González-Velosa et al. (2015)and Busso et al. (2020), where associate degrees were reported to have negative higher education premiums. However, this difference can be explained by the control group used in our study, which consisted of peers who had graduated from secondary school and attended a community college. While the income of individuals who pursued associate programs or apprenticeships may be better compared to their peers, their improvement may not be as substantial as those who pursued professional programs, which constitute 70% of the system. Thus, while students who pursued associate programs or apprenticeships performed better than their peers, their income improvements may not be as substantial as those who pursued professional programs.

Our findings emphasize the positive impact of higher education on career outcomes, as demonstrated by the “Stairway to Heaven” effect of college attendance and graduation. Graduates experience a considerable increase in earnings compared to those who did not attend college, and even low-income students who attended college but did not complete their degrees enjoy a premium over their non-college peers. However, a significant “Highway to Hell” effect exists for high-income individuals who do not complete their degrees and remain active in the system. Dropping out, regardless of when it occurs, represents a significant loss due the opportunity cost of not graduating. Nonetheless, dropping out remains a better option than not attending college at all.

In light of these results, policymakers should focus on designing policies that encourage higher education and support timely graduation to improve career outcomes for graduates. By prioritizing educational attainment and reducing barriers to graduation, policymakers can help to ensure that all students, regardless of their income or background, have the opportunity to succeed in the labor market.

In conclusion, the findings presented in this paper have important implications for Colombia in terms of reducing barriers to graduation and promoting the benefits of higher education. This can lead to a more highly trained and skilled workforce, which in turn can promote economic growth, reduce the gender gap, and encourage social mobility. Policymakers should therefore focus on designing policies that encourage higher education and support timely graduation to improve career outcomes beyond the success reported by Ferreyra et al. (2017) and Ministerio de Educación Nacional (2017). In addition, higher education institutions should evaluate how their degree requirements affect their students and take action to reduce barriers to graduation. Obtaining a degree not only sends signals to the labor market and increases students’ knowledge and training, but also opens the door to new levels of education that can further empower students and contribute to a more prosperous and equitable society.

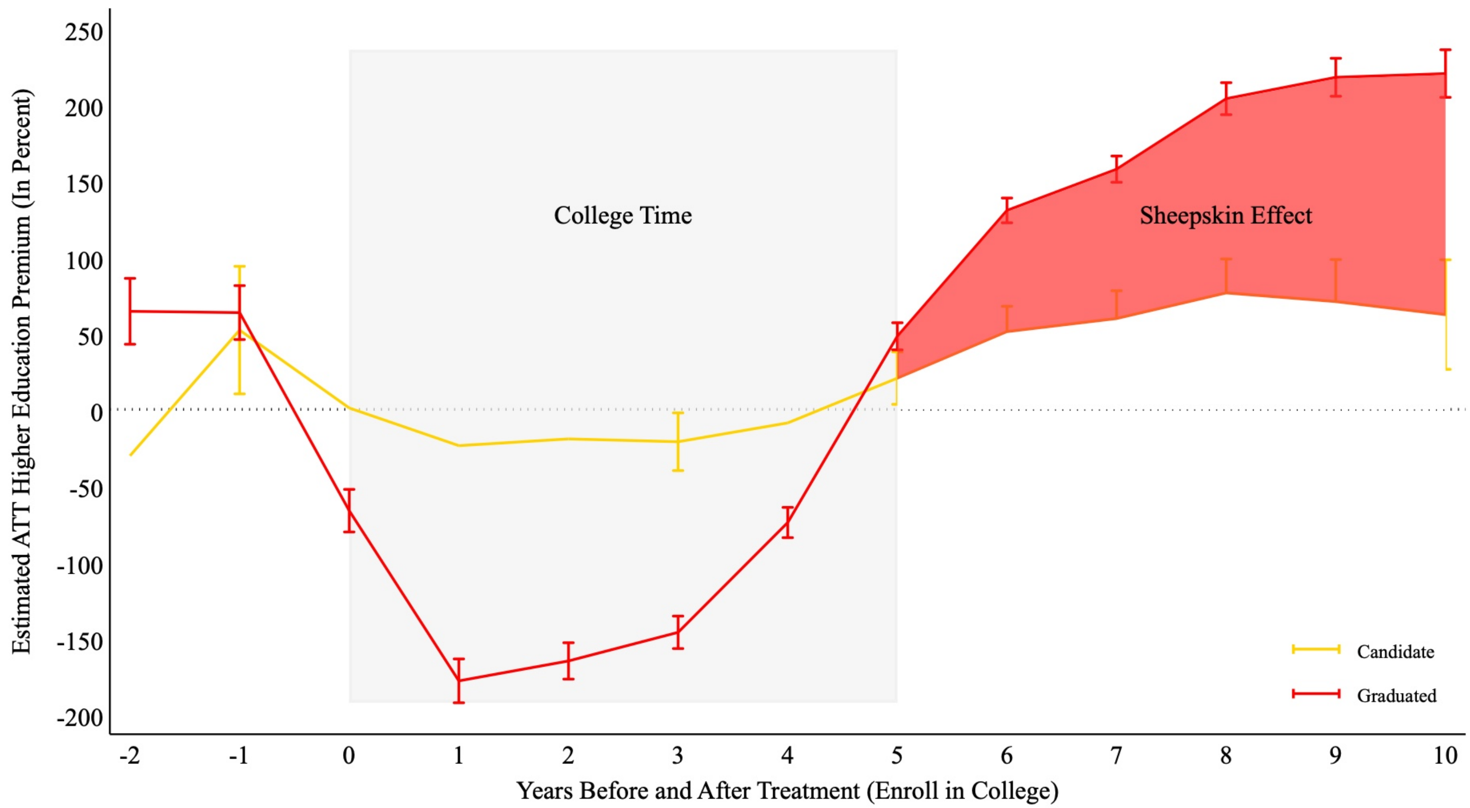
Figure 1: Higher Education Premium in Colombia



Improved Doubly Robust Estimator IMP - Sant'Anna & Zhao (2020)
Confidence interval was removed if Pvalue>0.05

Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further divided into four subgroups: graduates, candidates, incomplete, and active students. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis represents the years before and after enrollment in higher education, while the whiskers depict the 95 percent confidence intervals. In Colombia, students who discontinue their studies are categorized as incomplete or candidates. The shadowed area indicates the expected duration for completing a professional degree, which is five years after commencing as freshmen, whereas an associate program typically lasts three years. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

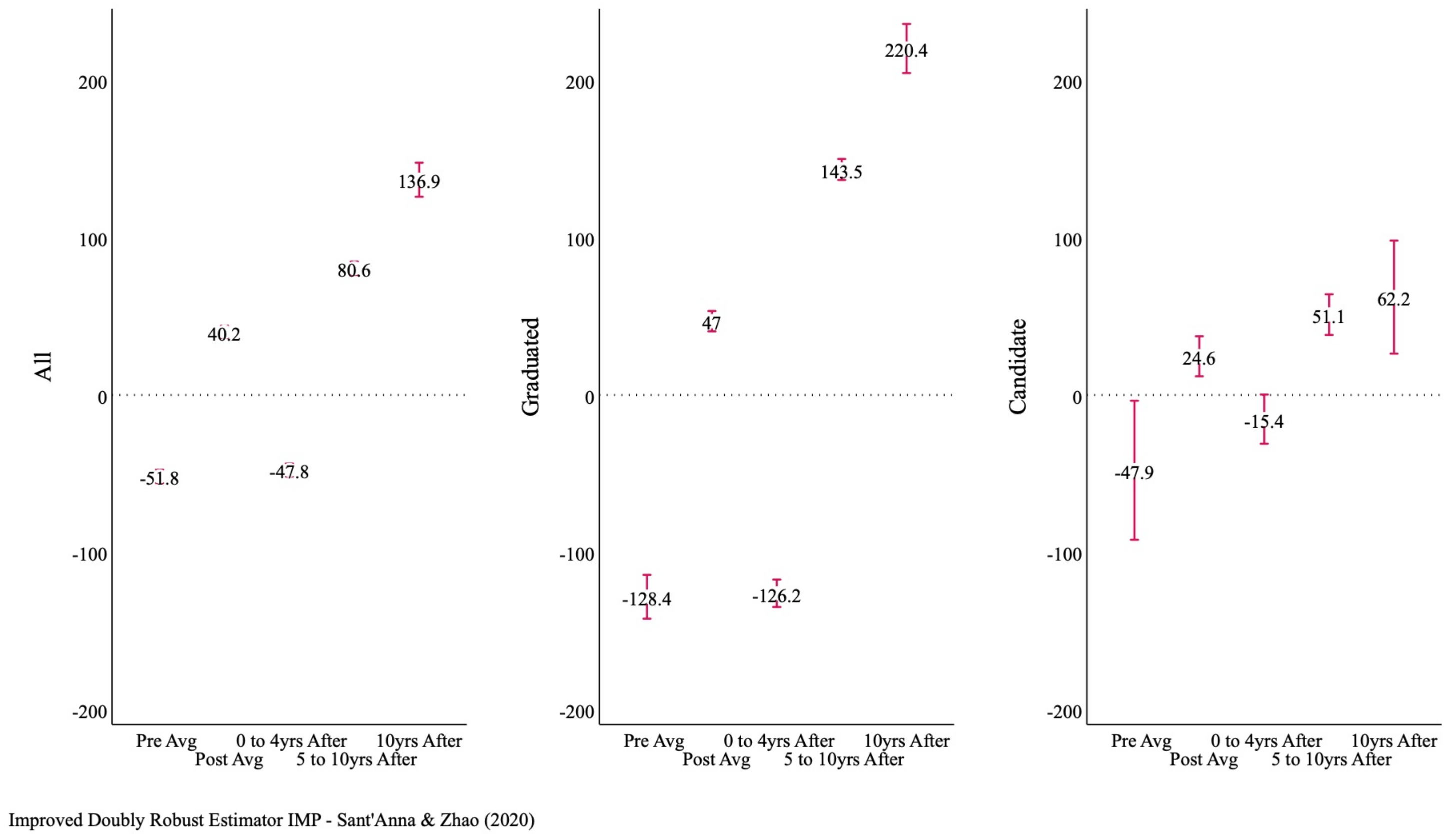
Figure 2: Higher Education Degree Sheepskin Effect



Improved Doubly Robust Estimator IMP - Sant'Anna & Zhao (2020)
Confidence interval was removed if Pvalue>0.05

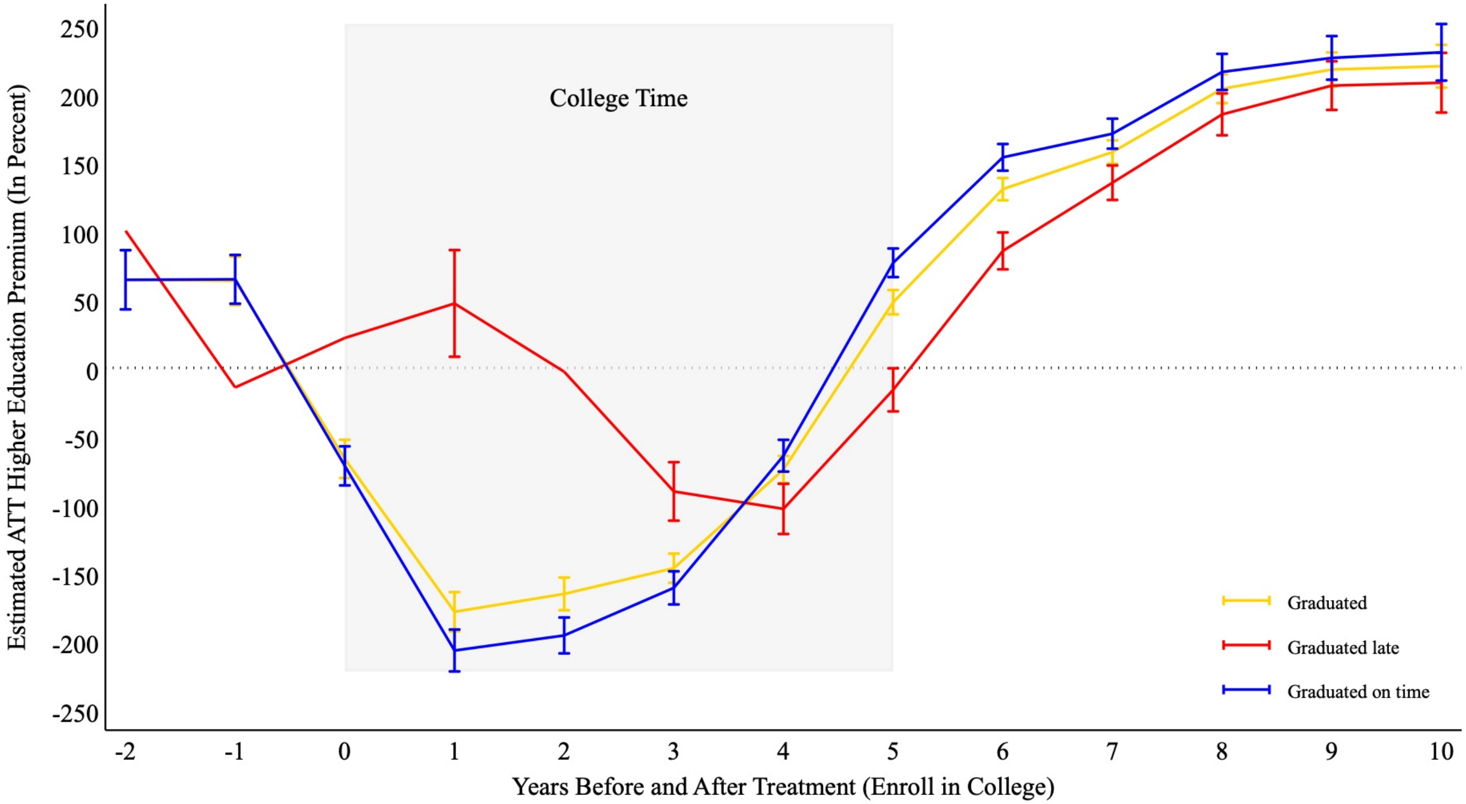
Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further divided into two subgroups to get the sheepskin effect: graduates and candidates. Sheepskin effect as the area created by the difference between the Graduates and the Candidates. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis represents the years before and after enrollment in higher education, while the whiskers depict the 95 percent confidence intervals. In Colombia, students who discontinue their studies are categorized as incomplete or candidates. The shadowed area indicates the expected duration for completing a professional degree, which is five years after commencing as freshmen, whereas an associate program typically lasts three years. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 3: Estimated Aggregated ATT in Higher Education Comparing All and Sheepskin Effect



Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further divided into three subgroups: all, graduates, and candidates. Sheepskin effect would be the difference between the coefficient for Graduates and Candidates after treatment. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the post treatment average, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

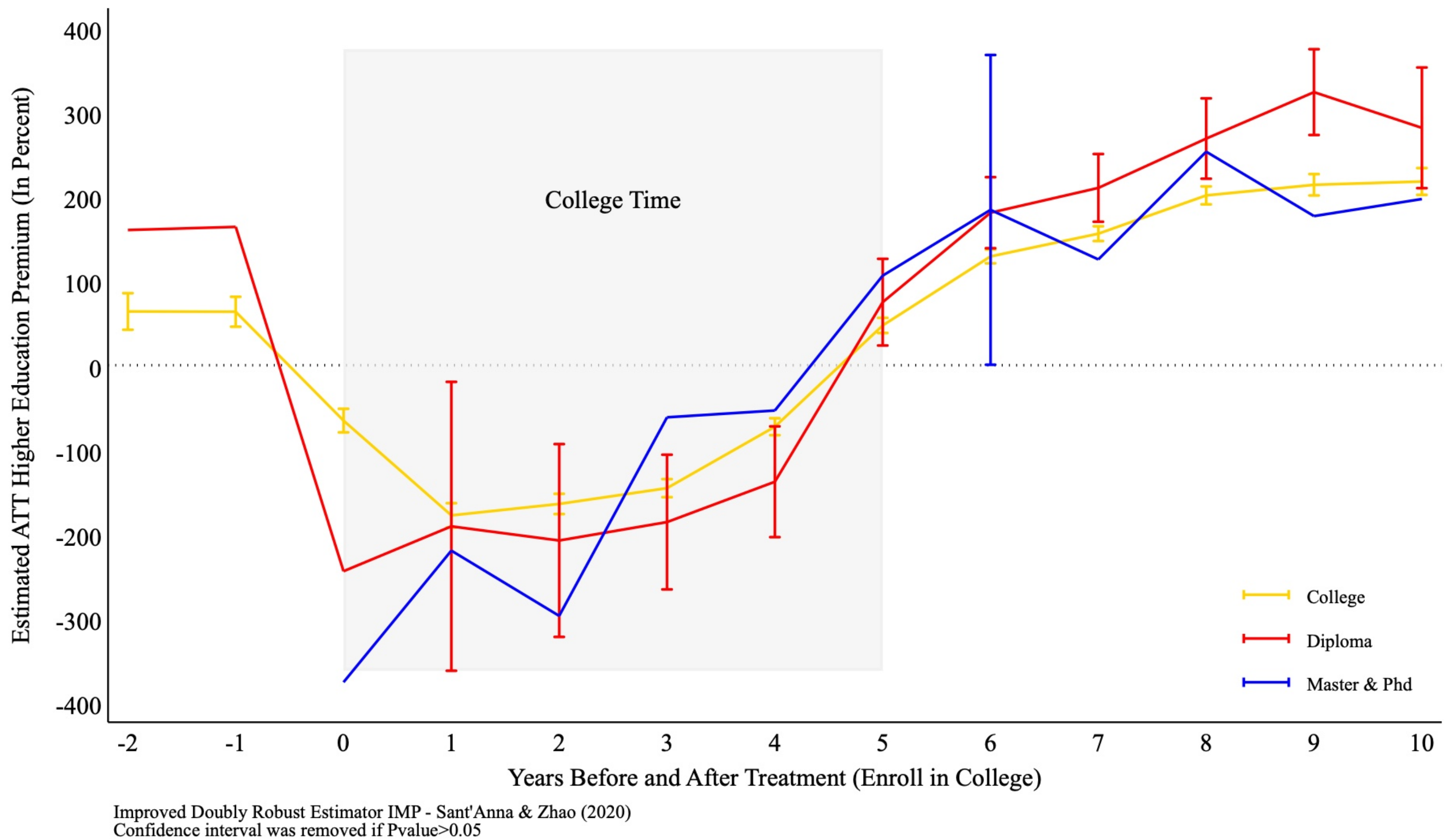
Figure 4: Higher Education Premium for Graduates



Improved Doubly Robust Estimator IMP - Sant'Anna & Zhao (2020)
Confidence interval was removed if Pvalue>0.05

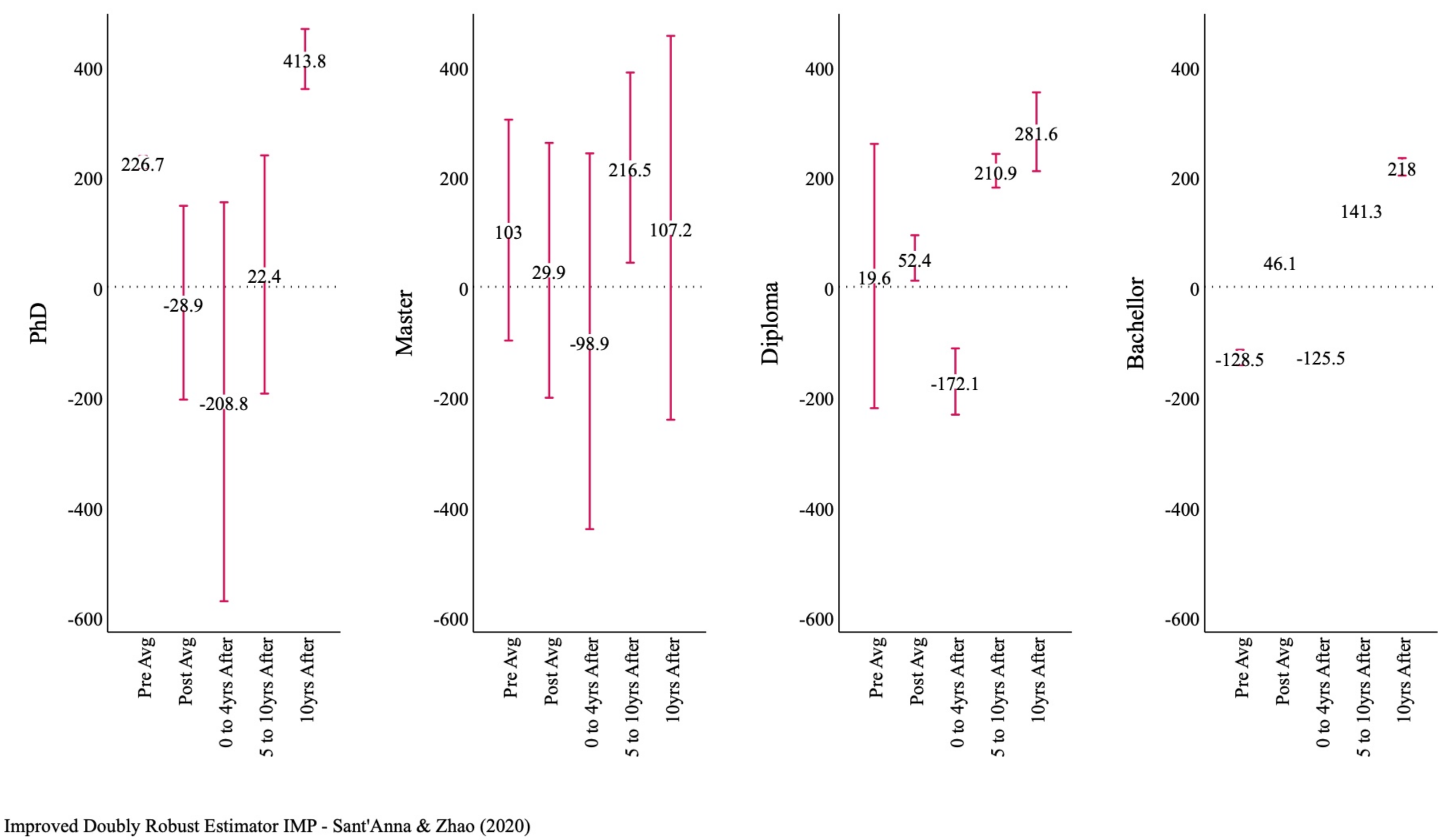
Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further divided into three subgroups: graduates, graduates on time, and graduates late. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis represents the years before and after enrollment in higher education, while the whiskers depict the 95 percent confidence intervals. In Colombia, students who discontinue their studies are categorized as incomplete or candidates. The shadowed area indicates the expected duration for completing a professional degree, which is five years after commencing as freshmen, whereas an associate program typically lasts three years. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 5: Higher Education Premium for Post-Grades



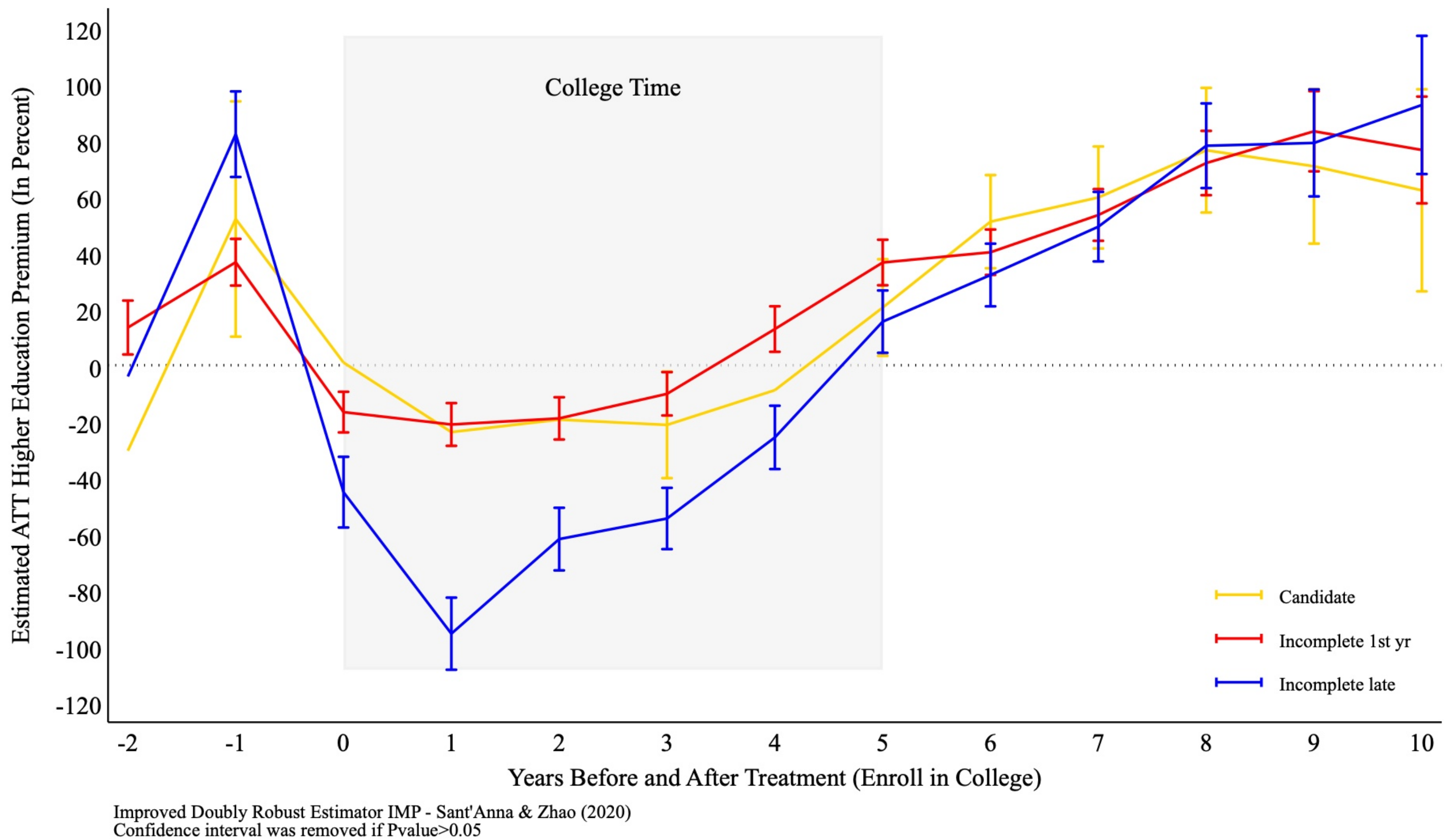
Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further the graduates are divided into three subgroups: those with only a bachelor degree, those with a diploma, and those with Master or PhD. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis represents the years before and after enrollment in higher education, while the whiskers depict the 95 percent confidence intervals. In Colombia, students who discontinue their studies are categorized as incomplete or candidates. The shadowed area indicates the expected duration for completing a professional degree, which is five years after commencing as freshmen, whereas an associate program typically lasts three years. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 6: Estimated Aggregated ATT in Higher Education by Graduate Level



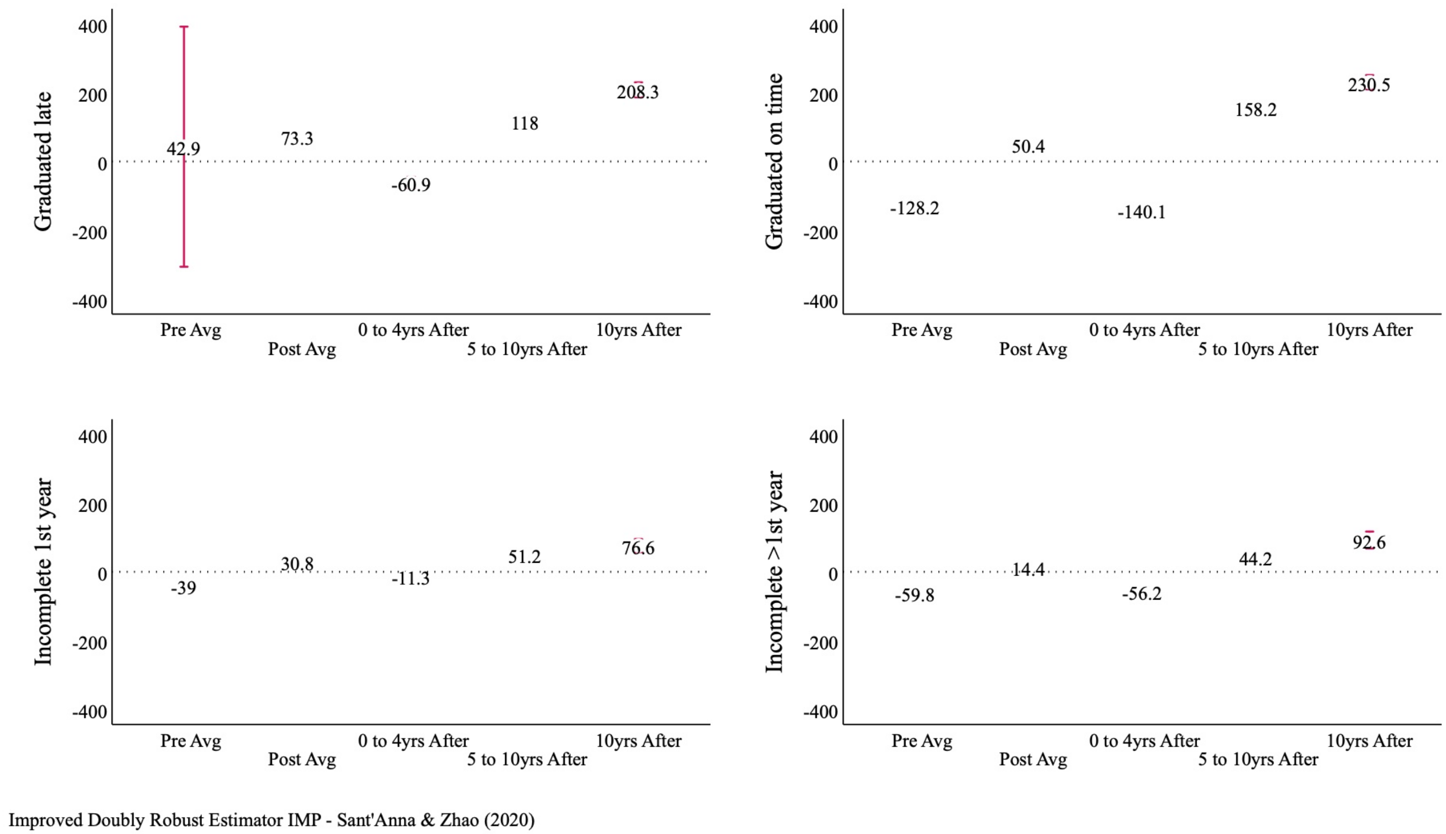
Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further divided into four subgroups: PhD, Masters, Diplomas, and Bachelors. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the average after 6 years, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 7: Higher Education Premium for Drop-outs



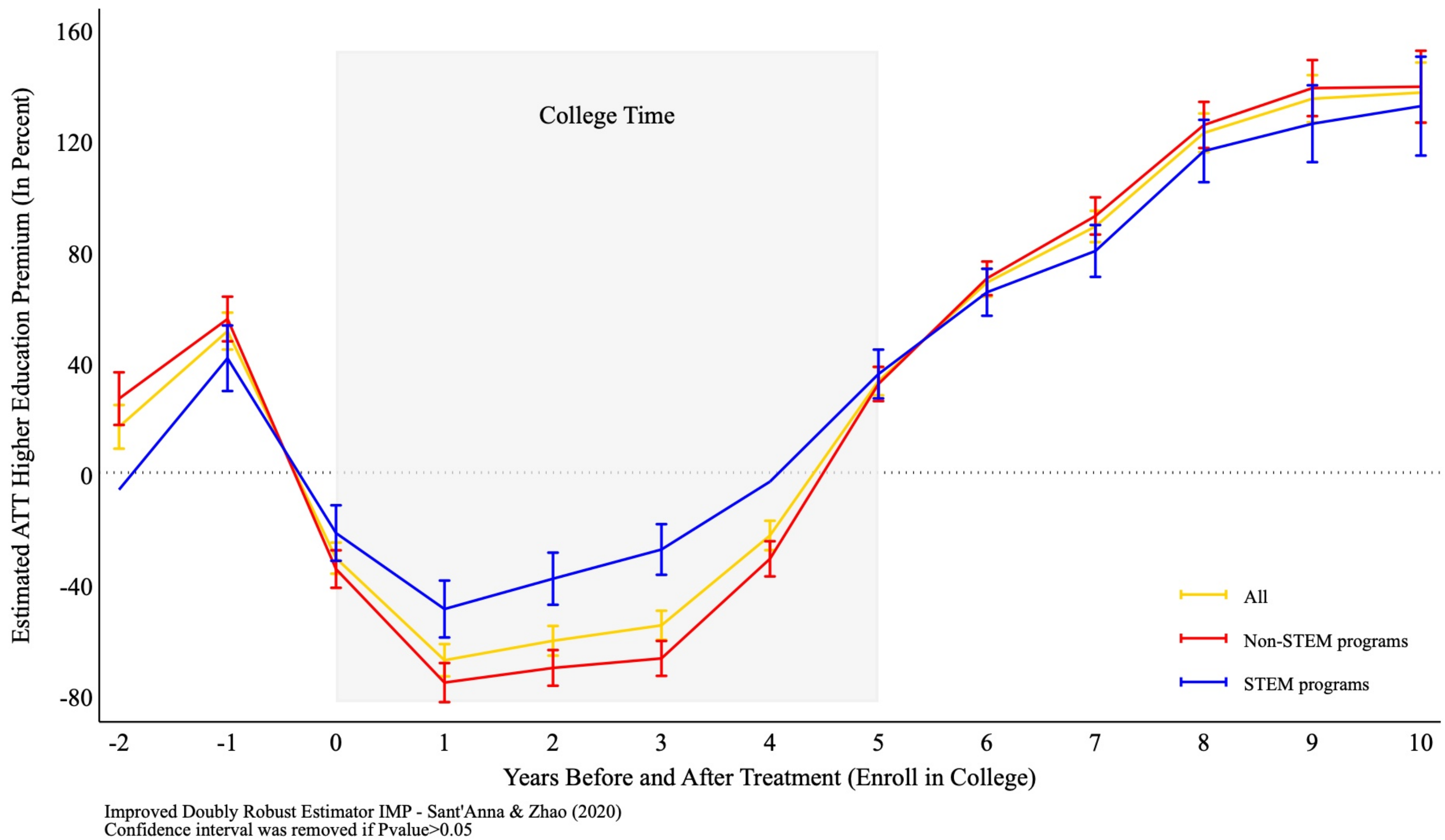
Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further the drop-outs are divided into three subgroups: Candidates, Incompletes in the first year, and Incompletes after the first year. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis represents the years before and after enrollment in higher education, while the whiskers depict the 95 percent confidence intervals. In Colombia, students who discontinue their studies are categorized as incomplete or candidates. The shadowed area indicates the expected duration for completing a professional degree, which is five years after commencing as freshmen, whereas an associate program typically lasts three years. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 8: Estimated Aggregated ATT in Higher Education by Types of Graduates and Incomplete



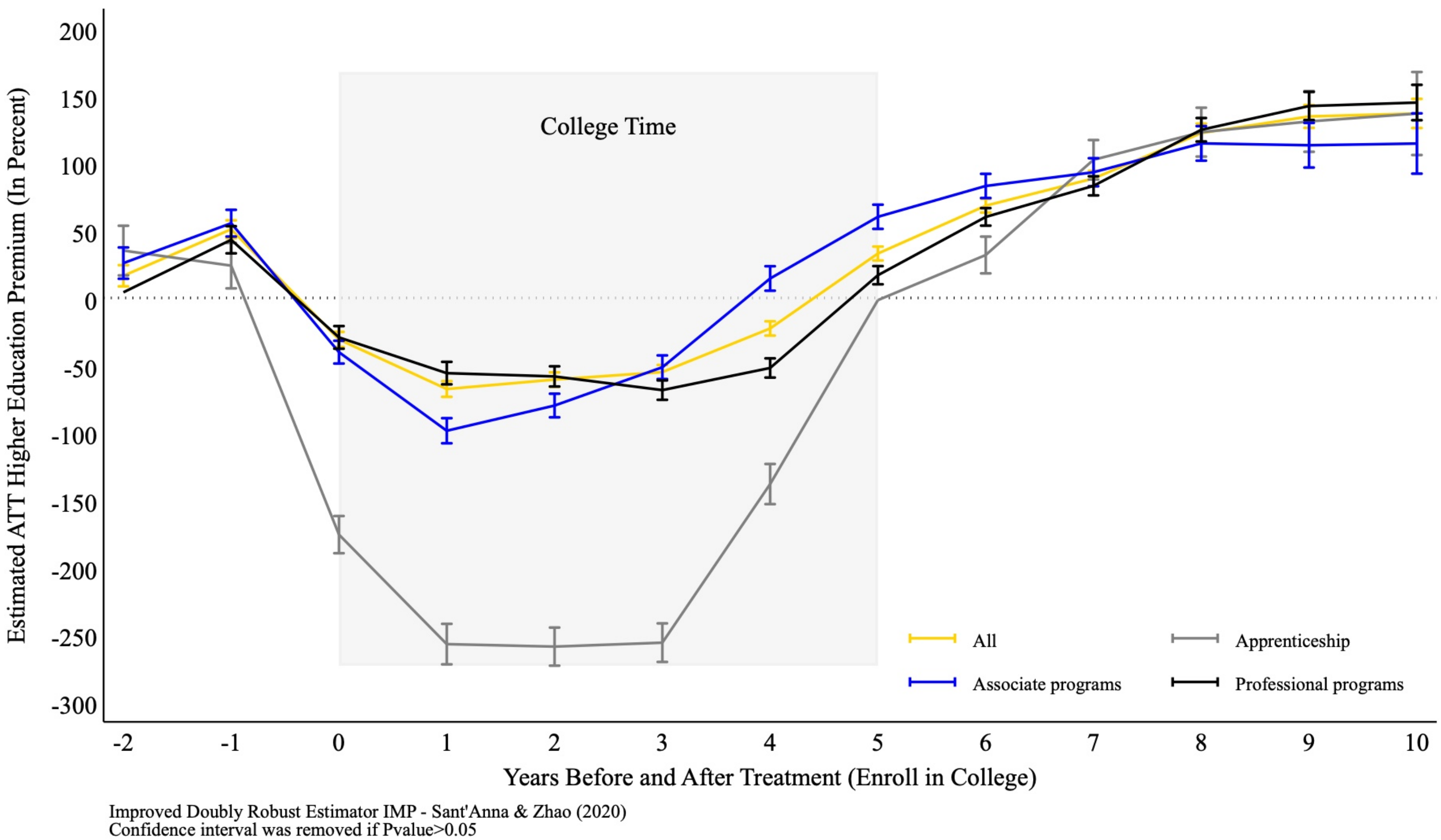
Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further divided into three subgroups: Professional, Associate, and Apprenticeship programs. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the average after 6 years, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 9: Higher Education Premium by Field of the Program



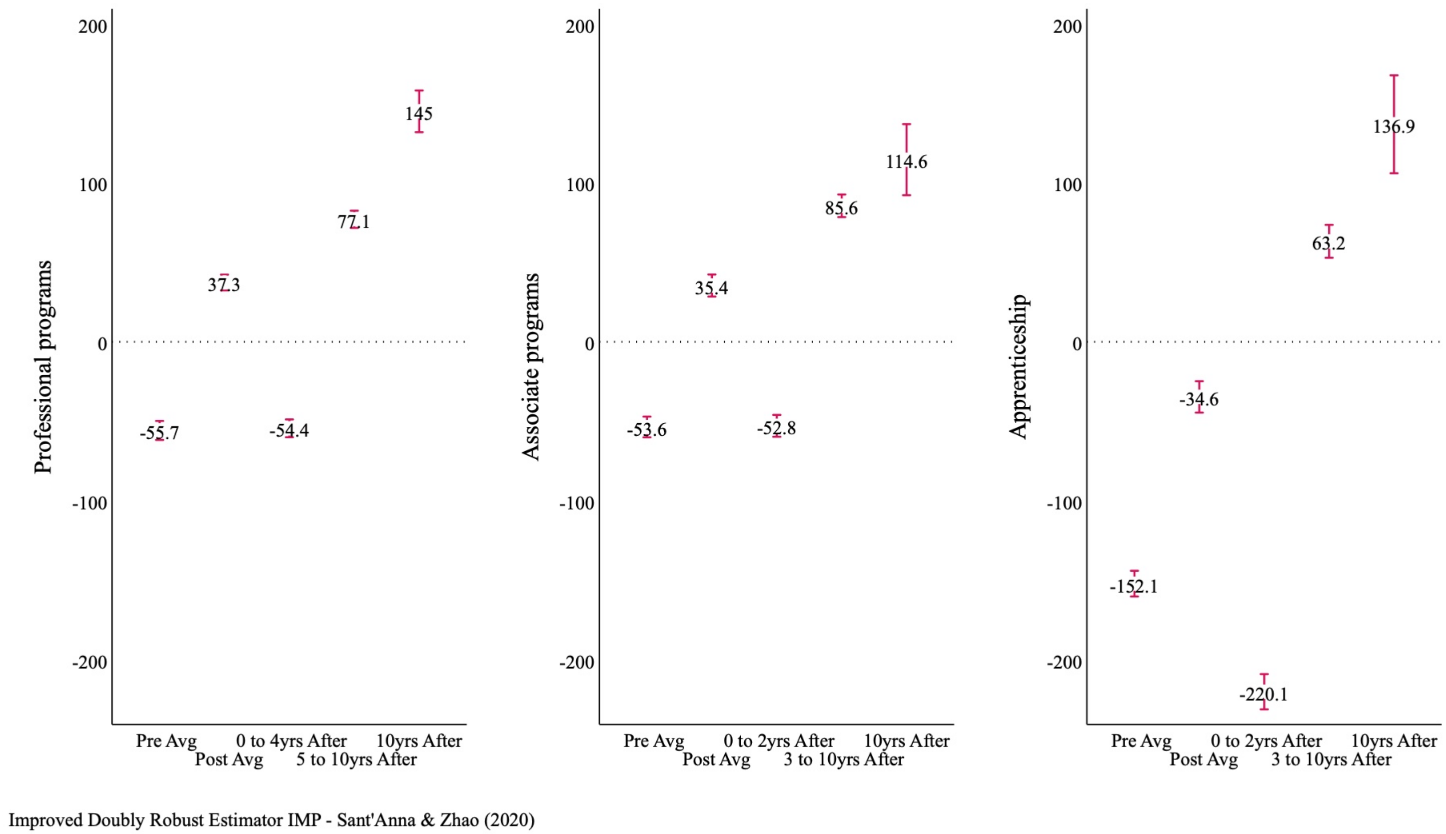
Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further divided into four subgroups: all, those who went to STEM programs, and those who went to non-STEM programs. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the post treatment average, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 10: Higher Education Premium by Level of the Program



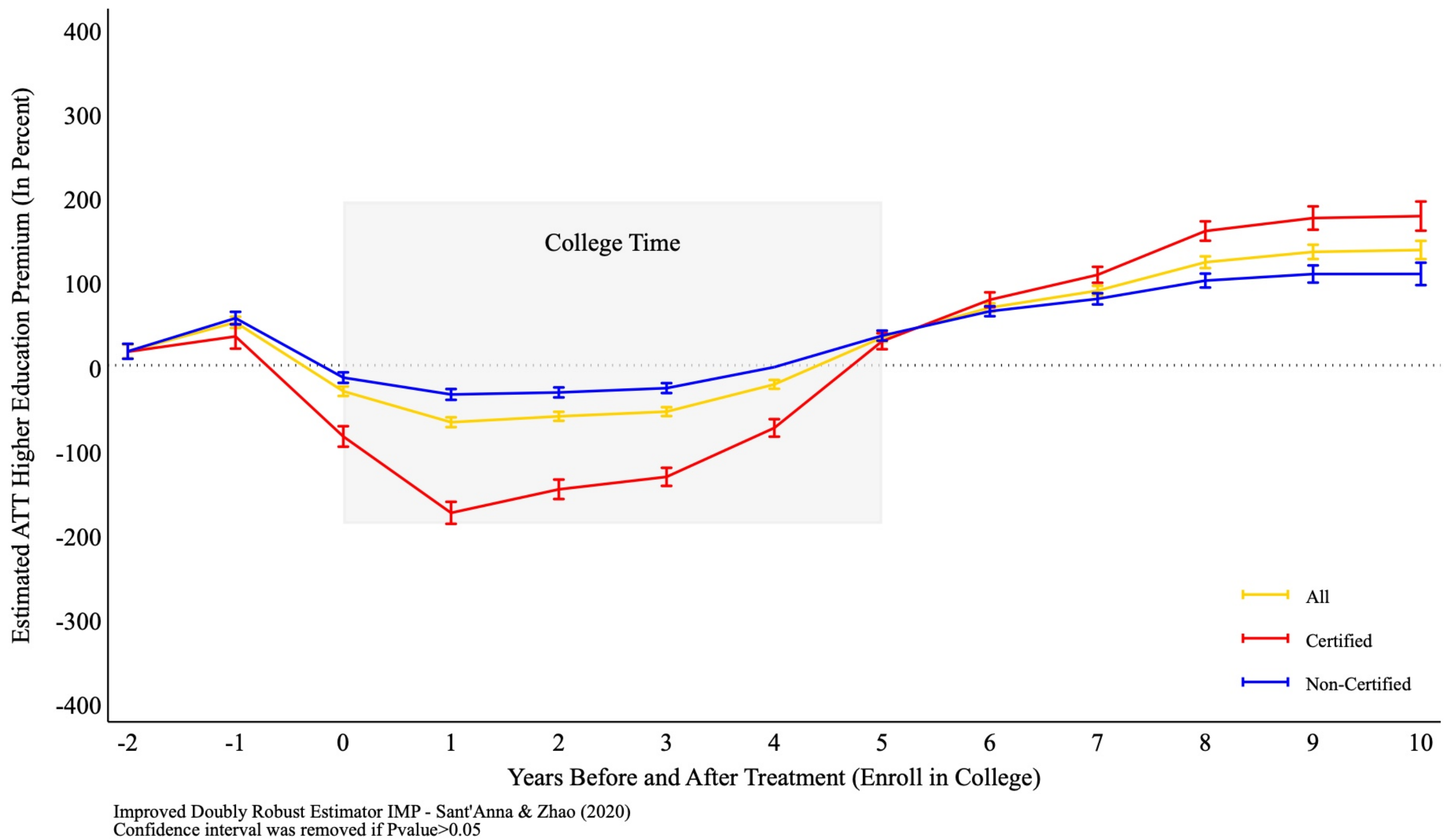
Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further divided into four subgroups: all, those who went to Associate programs, Tose who went to Profesional programs, and those who went to Apprenticeships. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the post treatment average, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 11: Estimated Aggregated ATT in Higher Education by Program Level



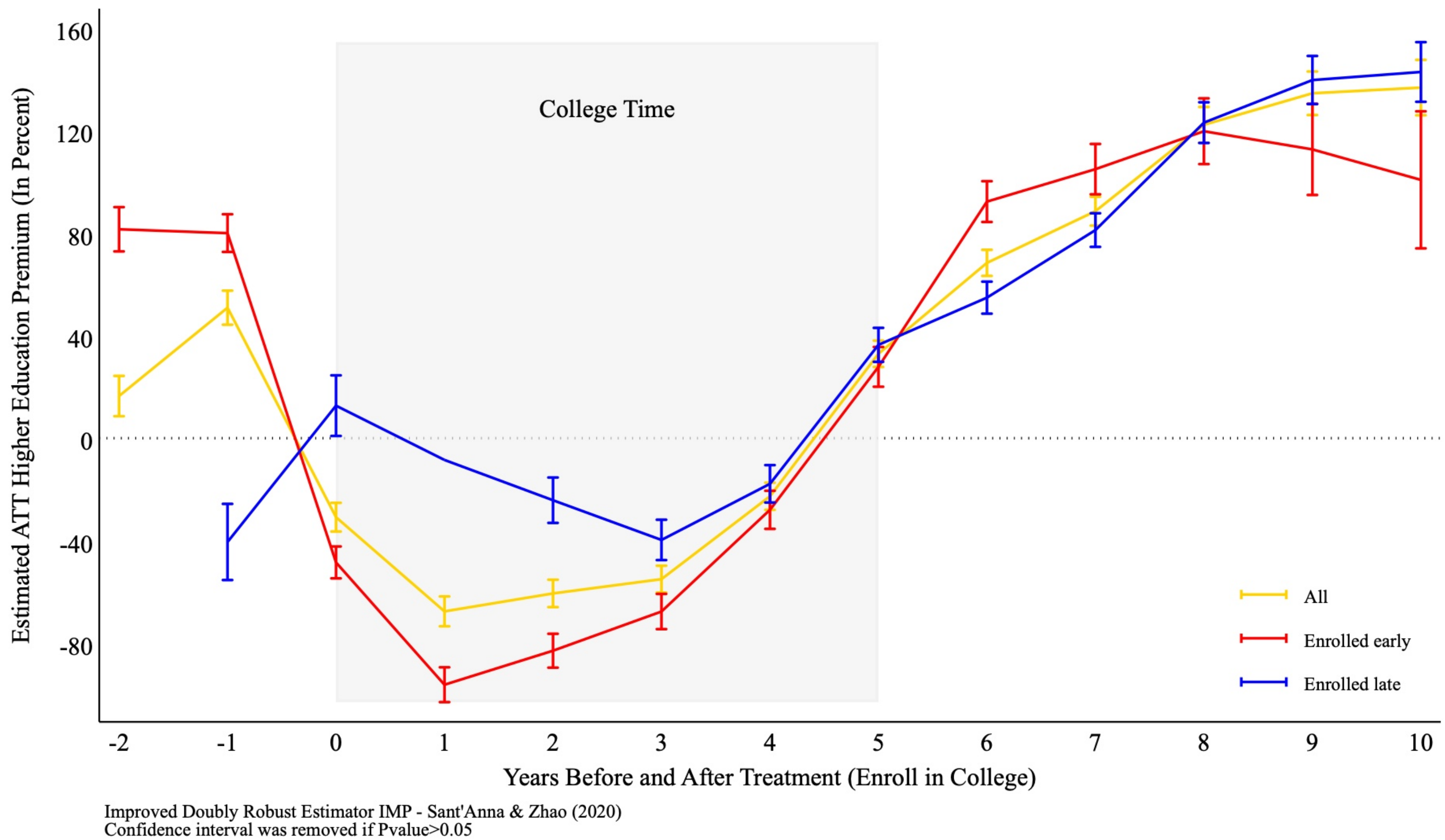
Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further divided into three subgroups: Professional, Associate, and Apprenticeship programs. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the average after 6 years, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 12: Higher Education Premium by HEI Quality



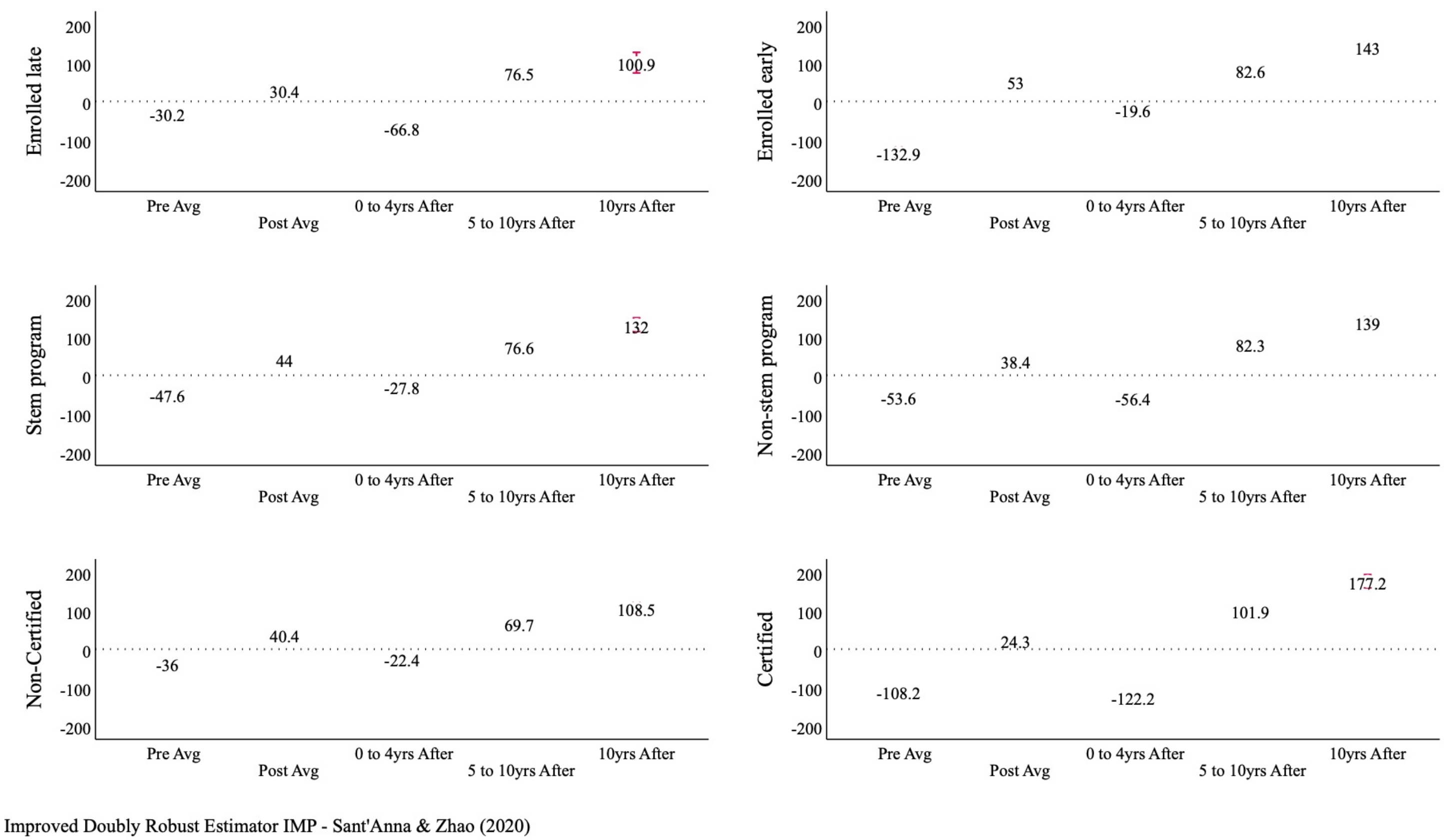
Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further divided into three subgroups: all, those who attended a certified HEI, and those who attended a non-certified HEI. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the post treatment average, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 13: Higher Education Premium by Time of Enrollment



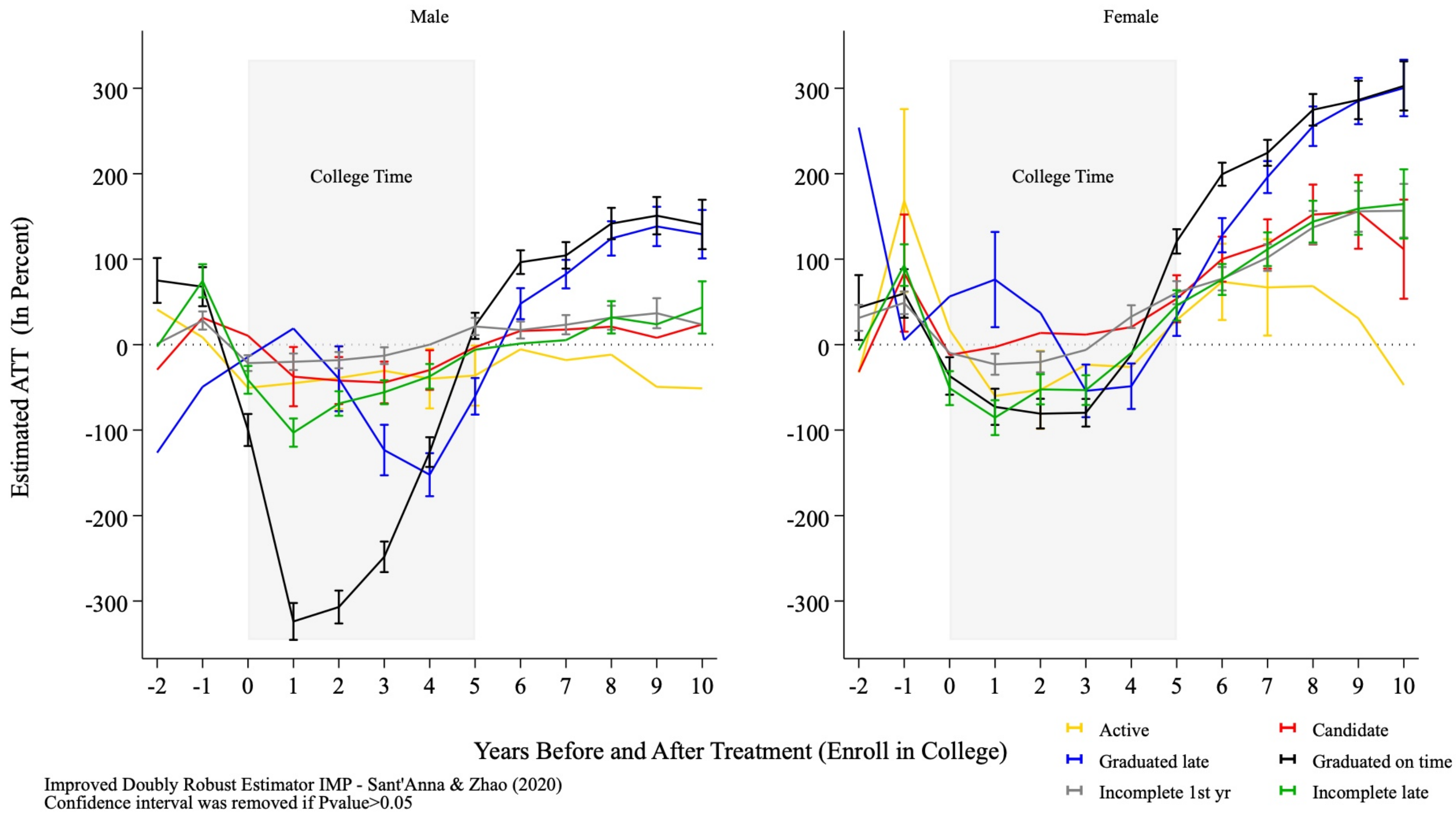
The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further divided into three subgroups: all, those who enrolled early, and those who enrolled late. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the post treatment average, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 14: Estimated Aggregated ATT in Higher Education by Other Characteristics



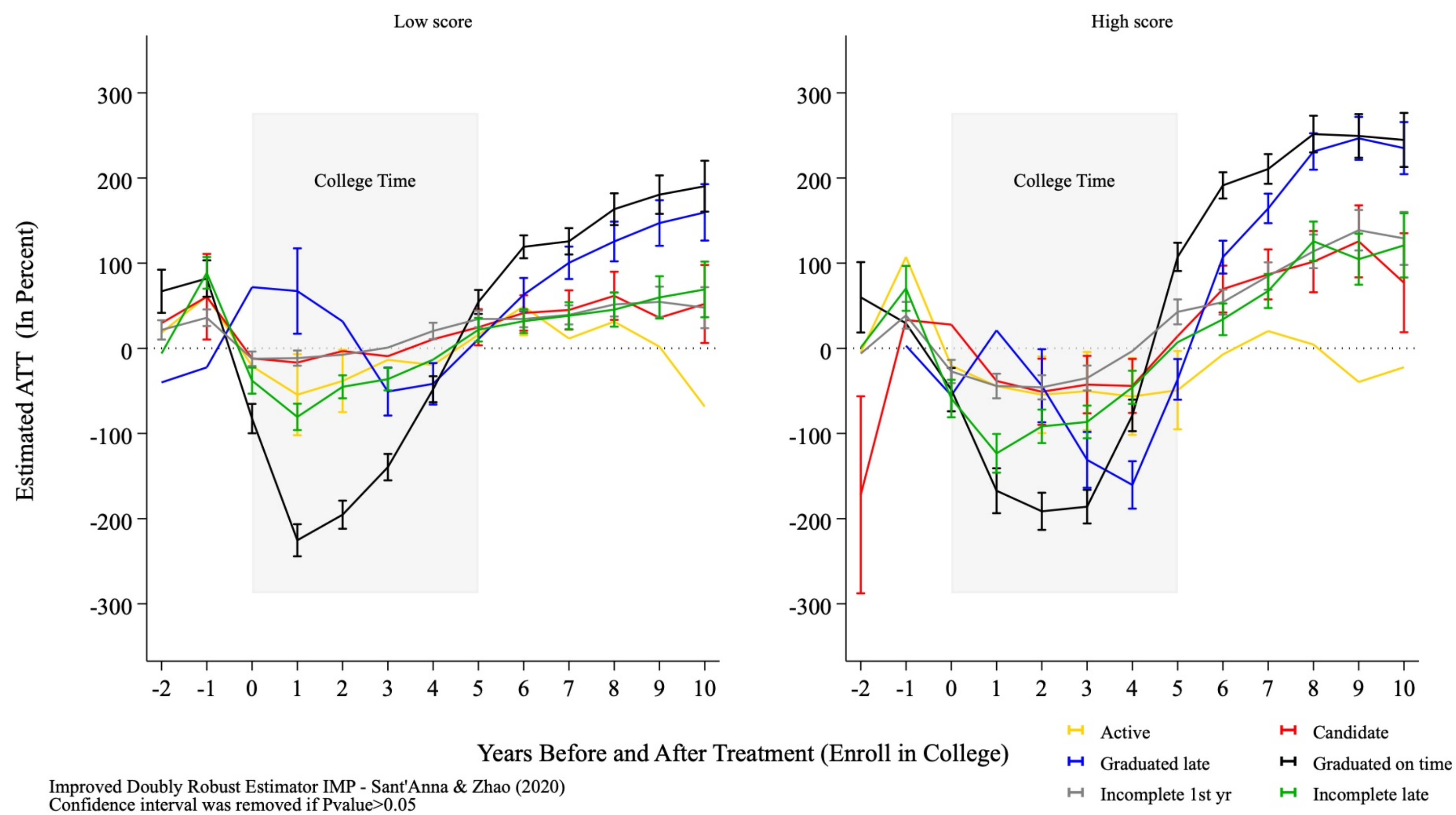
Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education. The treated group consists of individuals who attended college and is further divided into subgroups for time of enrollment, field of the program and quality of the HEI. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the average after 6 years, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 15: Higher Education Premium in Colombia by Gender



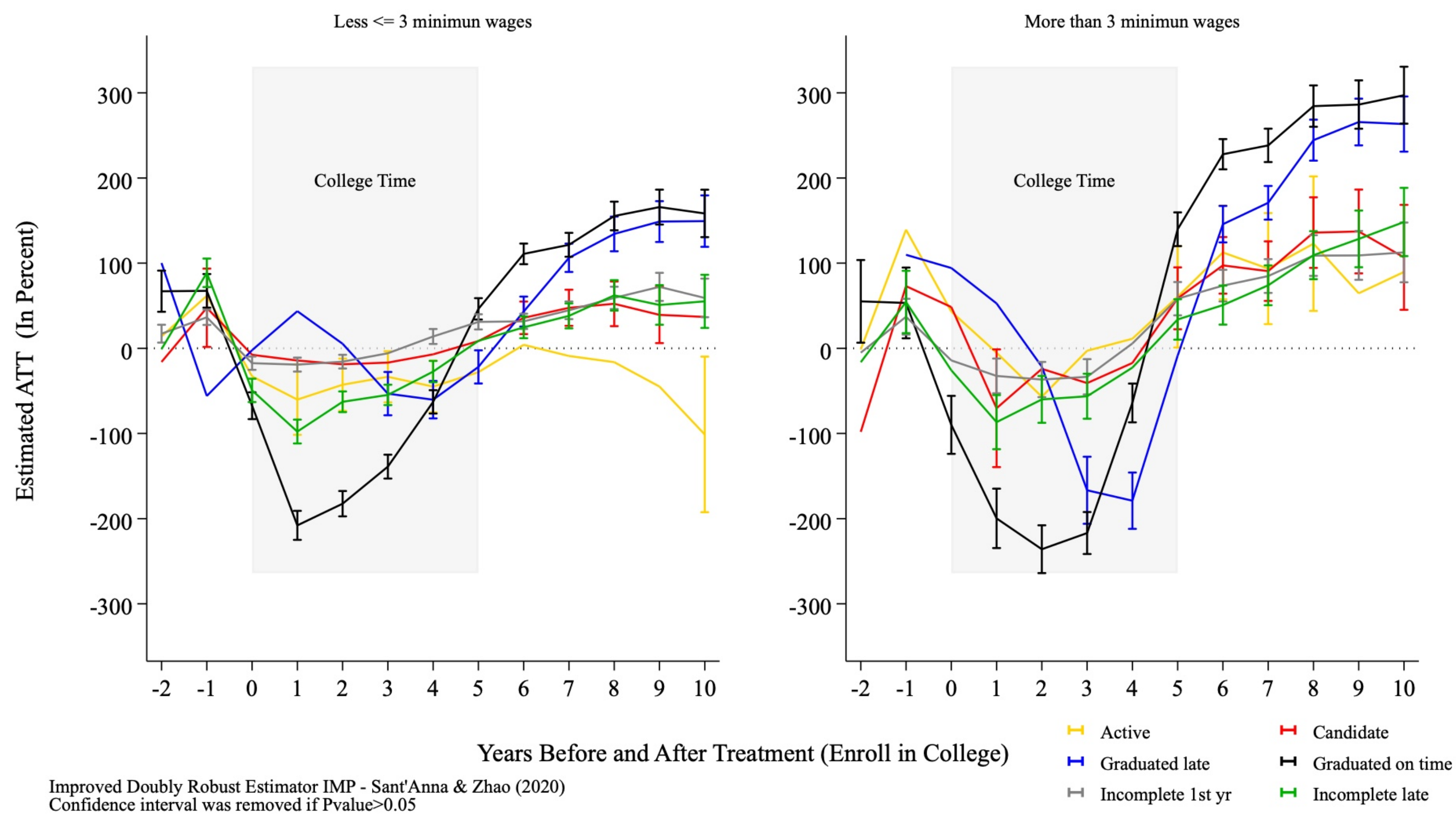
Note: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education for Males and Females. The treated group consists of individuals who attended college and is further divided into six subgroups: Active, Candidates, Incompletes before the first year, Incompletes after the first year, Graduates late, and Graduates on time. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the post treatment average, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 16: Higher Education Premium in Colombia by Saber 11 Score



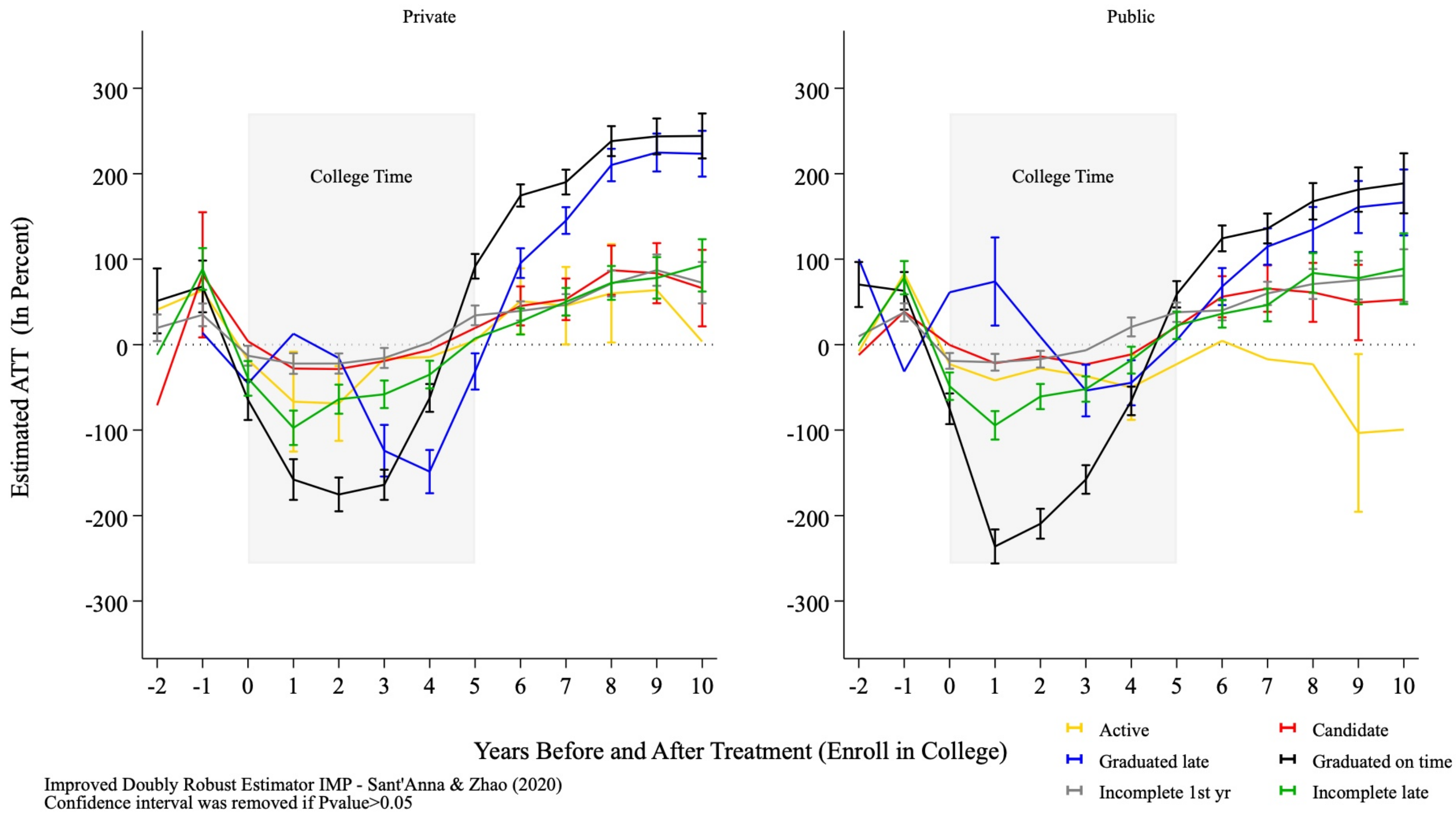
Note: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education for students with Low and High scores in the Saber 11 test. The treated group consists of individuals who attended college and is further divided into six subgroups: Active, Candidates, Incompletes before the first year, Incompletes after the first year, Graduates late, and Graduates on time. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the post treatment average, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 17: Higher Education Premium in Colombia by Household Income



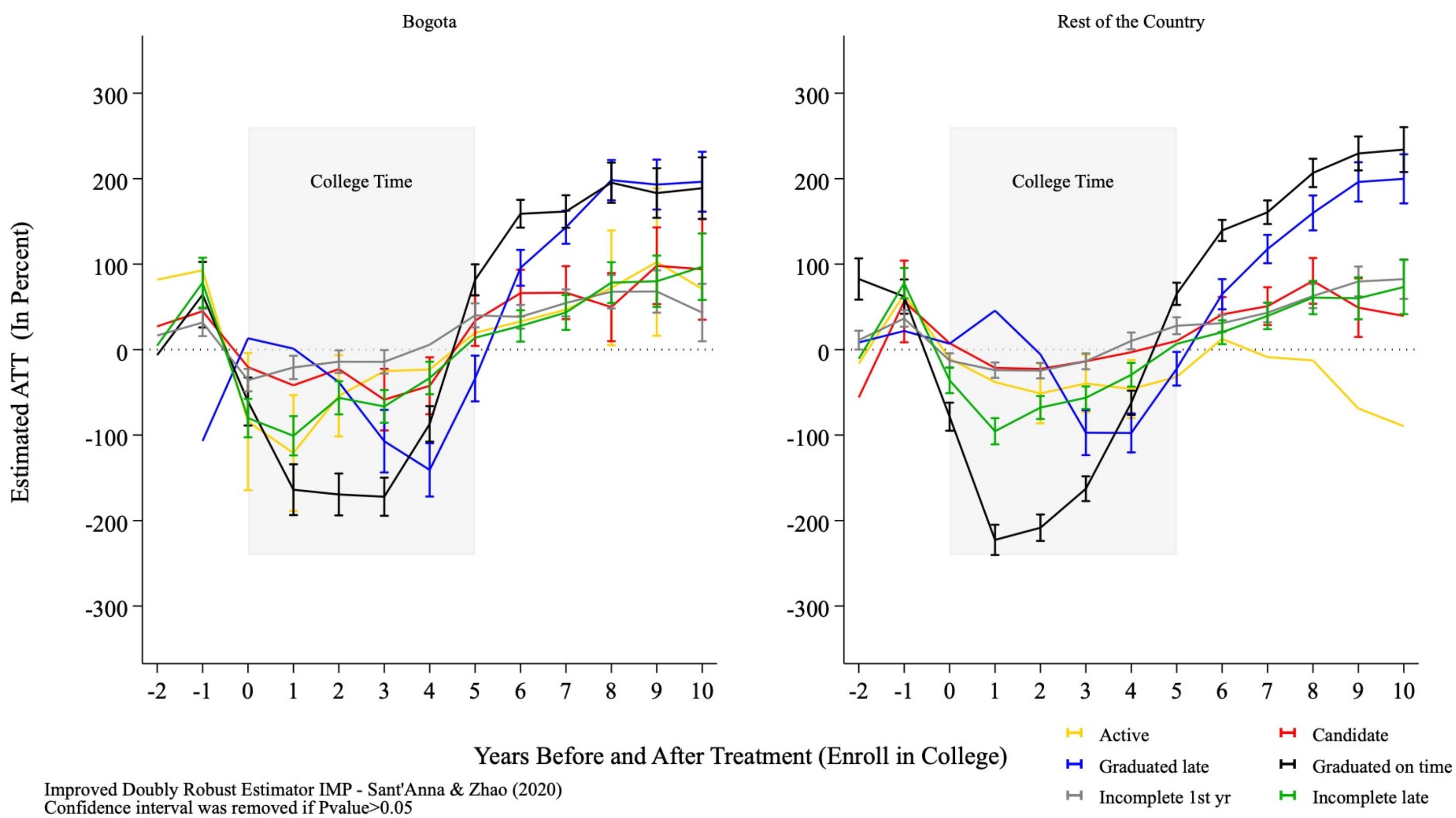
Note: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education for students with Low and High household income. The treated group consists of individuals who attended college and is further divided into six subgroups: Active, Candidates, Incompletes before the first year, Incompletes after the first year, Graduates late, and Graduates on time. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the post treatment average, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 18: Higher Education Premium in Colombia by School Sector



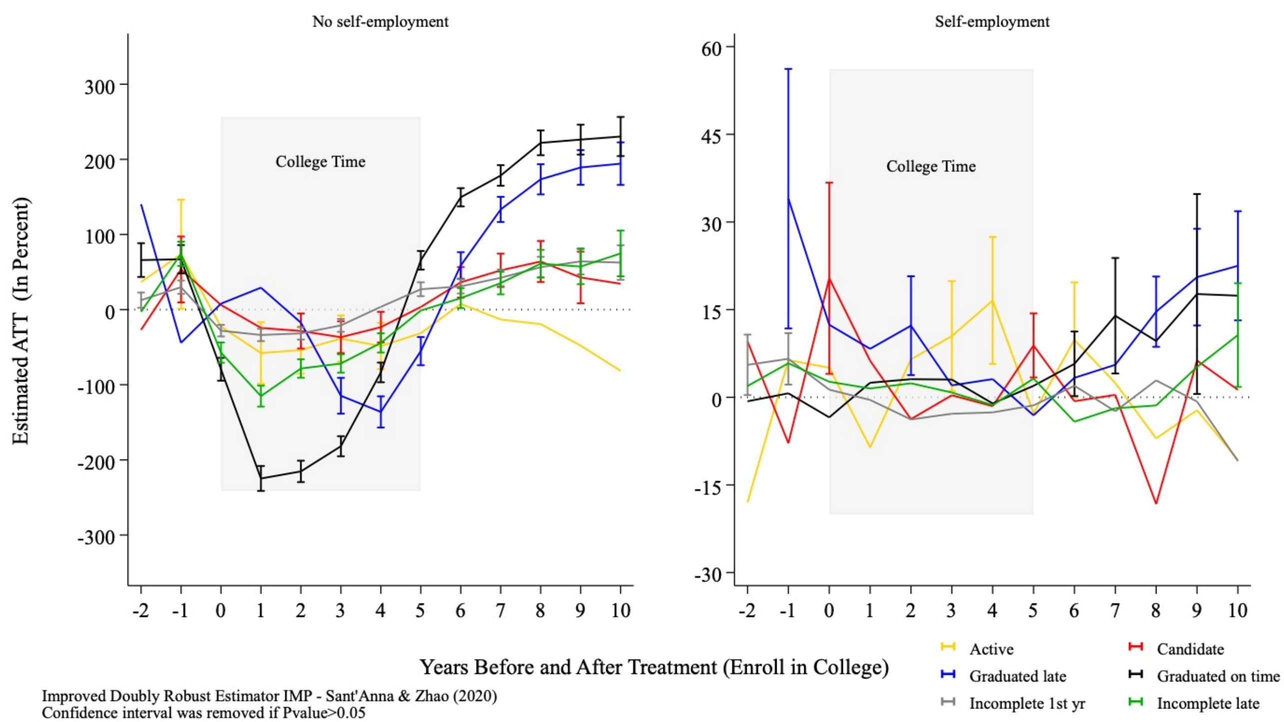
Note: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education for students from Public or Private secondary schools. The treated group consists of individuals who attended college and is further divided into six subgroups: Active, Candidates, Incompletes before the first year, Incompletes after the first year, Graduates late, and Graduates on time. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the post treatment average, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 19: Higher Education Premium in Colombia by Secondary School Location



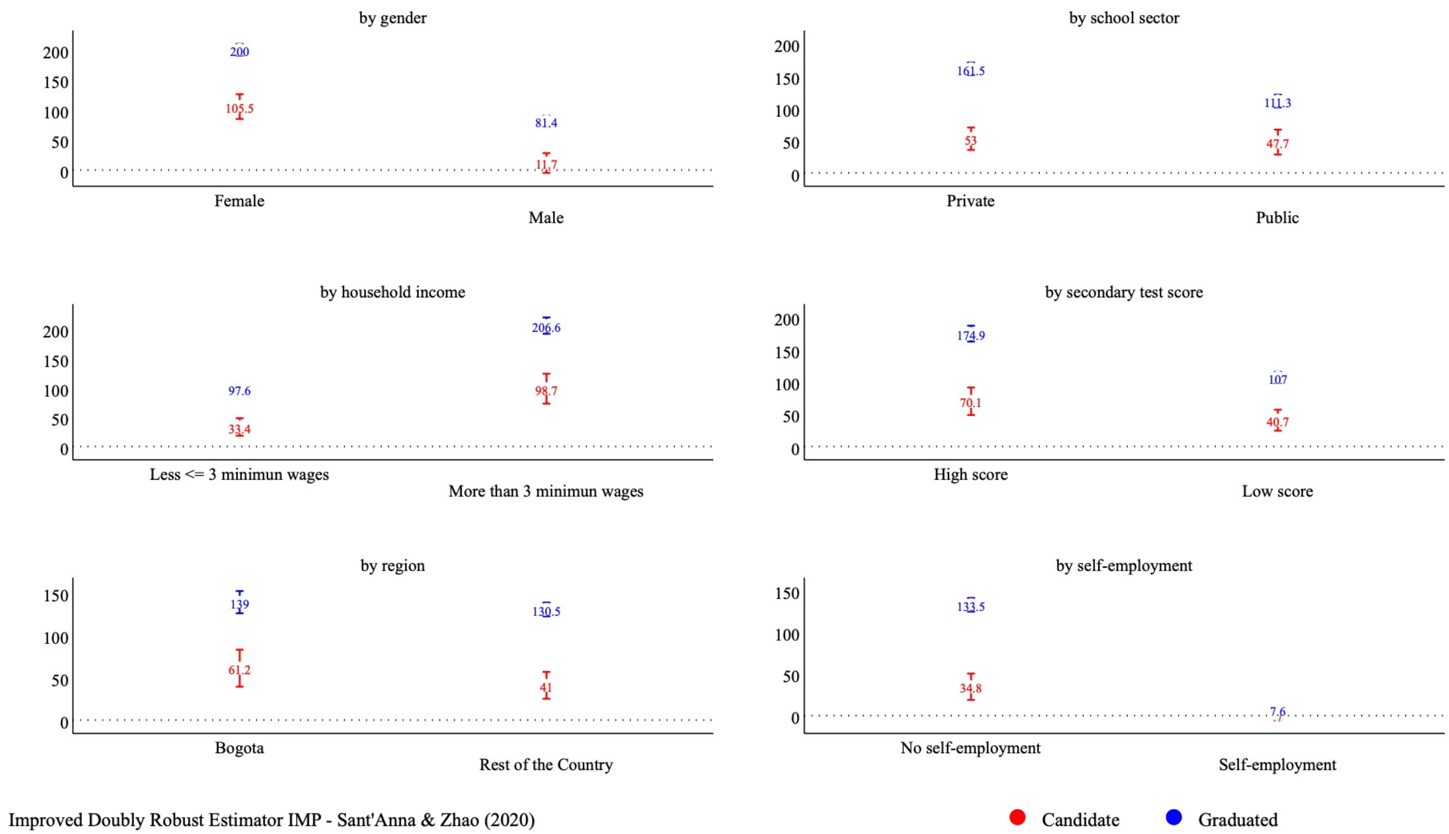
Note: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education for students from Bogota or outside Bogota. The treated group consists of individuals who attended college and is further divided into six subgroups: Active, Candidates, Incompletes before the first year, Incompletes after the first year, Graduates late, and Graduates on time. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the post treatment average, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 20: Higher Education Premium in Colombia by Type of Employment



Note: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education for students working as Self-Employes or non Self-Employes. The treated group consists of individuals who attended college and is further divided into six subgroups: Active, Candidates, Incompletes before the first year, Incompletes after the first year, Graduates late, and Graduates on time. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents the pre treatment average, the post treatment average, and the average 10 years after treatment. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

Figure 21: Summary of Returns and Sheepskin Effect by Characteristic



Notes: The figure shows the estimated ATT coefficients obtained through the modified Mincer regression (Equation 1) that calculates the returns to education for the year 10 after treatment. The treated group consists of individuals who attended college and is further divided into two subgroups: graduates, and candidates. Sheepskin effect would be the difference between the coefficient for Graduates and Candidates coefficient. The control group comprises secondary school students who have not yet enrolled in higher education (dotted horizontal line in Y=0). The X-axis presents different categories. The whiskers depict the 95 percent confidence intervals. The premiums presented in the figure are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 6) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) IMP estimator (Equation 5). The model controls for academic skills, household income, gender, program level, program type, and quality of higher education institutions. For further details on the variables used, readers can refer to Table 1.

References

Arango, Luis Eduardo, Carlos Esteban Posada, and José Darío Uribe, "Cambios en la estructura de los salarios urbanos en Colombia, 1984-2000," *Lecturas de Economia*, 2005, *63* (63), 7–39.

Arias, Helmut and Álvaro Chávez, "Cálculo de la tasa interna de retorno de la educación en Colombia," 2002.

Arrow, Kenneth, "Higher Education as filter," *Journal of Public Economics*, 1973, *2*, 193–216.

Bacolod, Marigee, Jorge De la Roca, and María Marta Ferreyra, "In search of better opportunities: Sorting and agglomeration effects among young college graduates in Colombia," *Regional Science and Urban Economics*, 2021, *87.*

Bauer, Thomas K., Patrick J. Dross, and John P. Haisken-DeNew, "Sheepskin effects in Japan," *International Journal of Manpower*, 2005.

Becker, Gary, "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy*, 1962, *70* (S5), 9.

Belman, Dale and John S. Heywood, "Sheepskin Effects in the Returns to Education: An Examination of Women and Minorities," *The Review of Economics and Statistics*, 1991.

_ **and** _ , "Sheepskin effects by cohort: Implications of job matching in a signaling model," *Oxford Economic Papers*, 1997.

Bilkic, N., T. Gries, and M. Pilichowski, "Stay in school or start working? - The human capital investment decision under uncertainty and irreversibility," *Labour Economics*, 2012.

Binelli, Chiara, "Returns to Education and Increasing Wage Inequality in Latin America," 2008, pp. 1–54.

Busso, Matías, Juan Sebastián Muñoz, and Sebastián Montaño, "Unbundling Returns to Skills: Evidence from Postsecondary Education in Colombia," 2020.

Callaway, Brantly and Pedro H.C. Sant'Anna, "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Calonico, S and H Nopo, "Where did you go to school? Private-Public differences in schooling trajectories and their role on earnings," *Well-Being and Social Policy*, 2007, *3* (2002), 25–46.

Card, David, "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," 1993.

_ , "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, sep 2001, *69* (5), 1127–1160.

Cárdenas, Mauricio and Raquel Bernal, "Changes in the distribution of income and the new economic model in Colombia," *Cepal Naciones Unidas*, 1999, (002116).

Collins, Randall, *The Credential Society: An Historical Sociology of Education and Stratification.*, New York: Academic Press, 1979.

Crespo, Anna and Maurício Cortez Reis, "Sheepskin effects and the relationship between earnings and education: analyzing their evolution over time in Brazil," *Revista Brasileira de Economia*, 2009, *63*, 209–231.

DANE, "DANE," 2023.

de la universidad colombiana, El Observatorio, "El observatorio de la universidad colombiana," 2020.

de Roux, Nicolás and Evan Riehl, "Disrupted academic careers: The returns to time off after high school," *Journal of Development Economics*, may 2022, *156*, 102824.

Duflo, Esther, "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 2001, *91* (4), 795–813.

Epple, Dennis, Richard Romano, and Holger Sieg, "Admission, tuition, and financial aid policies in the market for higher education," *Econometrica*, 2006, *74* (4), 885–928.

Farber, Henry S and Robert Gibbons, "Learning and the wage dynamics," *Quaterly Journal of Economics*, 1996, (November).

Farné, Stefano, "El Mercado de Trabajo de los Profesionales Colombianos," *Boletín del Observatorio del Mercado de Trabajo y la Seguridad Social*, 2006, *9.*

Ferrer, Ana and W. Craig Riddell, "The role of credentials in the Canadian labour market," *Canadian Journal of Economics*, 2002.

Ferreyra, María Marta, Andrea Franco Hernández, Tatiana Melguizo, and Angélica María Sánchez Díaz, "Estimating the Contribution of Short-Cycle Programs to Student Outcomes in Colombia," 2020.

_ **, Ciro Avitabile, Javier Botero Álvarez, Francisco Haimovich Paz, and Sergio Urzúa**, *Momento decisivo: La educación superior en América Latina y el Caribe* 2017.

Forero, Nohora and Manuel Ramírez, "Determinantes de los ingresos laborales de los graduados universitarios durante el período 2001-2004," *Documentos De Trabajo*, 2008, (31).

Frenette, Marc, "Access to college and university: Does distance to school matter?," *Canadian Public Policy*, 2004, *30* (4), 427–442.

García-Suaza, Andrés Felipe, Juan Carlos Guataquí, José Alberto Guerra, and Darío Maldonado, "Beyond the Mincer equation: the internal rate of return to higher education in Colombia," *Education Economics*, 2014.

Gibson, John, "Sheepskin effects and the returns to education in New Zealand: Do they differ by ethnic groups?," *New Zealand Economic Papers*, 2000, *34* (2), 201–220.

González-Velosa, Carolina, Graciana Rucci, Miguel Sarzosa, and Sergio Urzúa, "Returns to higher education in Chile and Colombia," 2015.

Hernández, Gustavo, "Cuán rentable es la educación superior en Colombia?," *Lecturas de Economia*, 2010, *73* (julio-diciembre), 181–214.

Herrera-Prada, Luis Omar, "Determinantes de la tasa de graduación y graduación a tiempo en la educación superior de Colombia 1998-2010," *Coyuntura Económica*, 2013, *XLIII* (1), 143.177.

_ **and Carlos Caballero**, "El financiamiento de la educación superior en Colombia," in "La educación superior:Retos y perspectivas" 2013, chapter 3, pp. 121–175.

Hungerford, Thomas and Gary Solon, "Sheepskin Effects in the Returns to Education," *The Review of Economics and Statistics*, 1987.

Jaeger, David A. and Marianne E. Page, "Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education," *The Review of Economics and Statistics*, 1996.

MacLeod, W. Bentley, Evan Riehl, Juan Saavedra, and Miguel Urquiola, "The big sort: College reputation and labor market outcomes," *American Economic Journal: Applied Economics*, 2017, *9* (3), 223–261.

Manacorda, Marco, Carolina Sánchez-Páramo, and Norbert R. Schady, "Changes in Returns to Education in Latin America : The Role of Demand and Supply of Skills," *Industrial and Labor Relations Review*, 2007, *63* (2).

Melguizo, Tatiana and Jacques Wainer, "Toward a set of measures of student learning outcomes in higher education: evidence from Brazil," *Higher Education*, 2016, *72*, 381–401.

Mesa, Fernando and Javier Gutiérrez, "Efectos de la apertura económica en el mercado laboral industrial," *Revista de Planeación y Desarrollo*, 1996.

Mincer, Jacob, "Schooling, Experience, and Earnings," 1974.

Ministerio de Educación Nacional, "Análisis de determinantes de la deserción en la educación superior colombiana con base en el SPADIES.," Technical Report, Ministerio de Educación Nacional - Universidad de los Andes, Bogotá 2008.

_ , "La Revolución Educativa 2002 - 2010. Informe de gestión.," Technical Report, Bogotá 2010.

_ , "Boletín Educación Superior," Technical Report, Bogotá 2017.

Mora, Jhon James, "Sheepskin effects and screening in Colombia," *Colombian Economic Journal*, 2003.

_ **and Juan Muro**, "Sheepskin effects by cohorts in Colombia," *International Journal of Manpower*, 2008.

Núñez, Jairo and Fabio Sánchez, "Educación y Salarios relativos en Colombia, 1976-1995. Determinantes, Evolución e Implicaciones para la Distribución del Ingreso," 1998.

OECD, "Self-employment rate (indicator)," 2021.

Olfindo, Rosechin, "Diploma as signal? Estimating sheepskin effects in the Philippines," *International Journal of Educational Development*, 2018.

Orozco Silva, Luis Enrique, *La Política de Cobertura: eje de la revolución educativa, 2002-2008.*, Bogotá: Ediciones Uniandes, 2010.

_ **, Alberto Roa Valero, and Luis Carlos Castillo Gómez**, "La Educación Superior en Colombia," 2011, pp. 1–83.

Patrinos, Harry Anthony and George Psacharopoulos, "Returns to Investments in Education: A Further Update. Education Economics," *Education Economics*, 2004, *12* (2).

_ **and** _ , "Returns to education in developing countries," in "The Economics of Education" 2020, chapter 4, pp. 53–64.

Peet, Evan D., Günther Fink, and Wafaie Fawzi, "Returns to education in developing countries: Evidence from the living standards and measurement study surveys," *Economics of Education Review*, 2015, *49*, 69–90.

Phelps, Edmund S, "The Statistical theory of Racism and Sexism," *American Economic Review*, 1972.

Posso, Christian, "Desigualdad salarial en Colombia 1984-2005: cambios en la composición del mercado laboral y retornos a la educación post-secundaria," *Borradores de Economía*, 2008, (529), 1–31.

Prada, Carlos Felipe, "Is the decision of study in Colombia profitable?," 2006, (July 2005), 12–69.

Psacharopoulos, George, "Returns to Education: A Further International Update and Implications," *Journal of Human Resources*, 1985, *20* (4), 583–604.

_ , "Returns to Investment in Education. A Global Update," *World development*, 1994, *22* (9), 1325–1343.

_ **and Harry Anthony Patrinos**, "Returns to Investment in Education A Decennial Review of the Global Literature," *Education Economics*, 2018, *26* (5).

Rios-Avila, Fernando, Brantly Callaway, and Pedro H.C. Sant'Anna, "csdid: Difference-in-Differences with Multiple Time Periods in Stata," 2021.

Rodríguez, Eduardo, "Rentabilidad y crecimiento de la educación superior en Colombia 1971-1978," 1981.

Santamaría, Mauricio, *External Trade, Skill Technology and the Recent Increase of Income Inequality in Colombia* 2001.

Sant'Anna, Pedro H.C. and Jun Zhao, "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, nov 2020, *219* (1), 101–122.

Schady, Norbert R., "Convexity and sheepskin effects in the human capital earnings function: Recent evidence for Filipino men," 2003.

Shabbir, T., "Sheepskin effects in the returns to education in a developing country," *Pakistan Development Review*, 1991.

Son, Hyelim, "Human Capital Investment When Sheepskin Effects Matter : Evidence from Income Shocks in Indonesia," 2013, pp. 1–48.

Spence, Michael, "Job Market Signaling," *The Quarterly Journal of Economics*, 1973.

Tenjo Galarza, Jaime, "Evolución de los retornos a la inversión en educación 1976- 1989." PhD dissertation 1993.

Wood, Tom, "The Sheepskin Effect," *National Association of Scholars*, 2009, (July).

Yunus, Norhanishah Mohamad, "Sheepskin effects in the returns to higher education: New evidence for Malaysia," *Asian Academy of Management Journal*, 2017, *22* (1), 151–182.

Zárate, Héctor Manuel, "Cambios en la estructura salarial: una historia desde la regresión cuanfílica," *Monetaria*, 2005, *XXVIII* (4), 339–364.

Table 1: Descriptive Statistics PILA

	All Students		Males		Females	
	Mean	Sd	Mean	Sd	Mean	Sd
	(1)	(2)	(3)	(4)	(5)	(6)
Annual income in 2008	3.0	5.2	3.2	5.0	2.8	5.5
Days worked in 2008	227	404	228	363	225	460
Annual income in 2009	3.2	5.9	3.4	5.5	3.1	6.4
Days worked in 2009	248	532	245	480	254	600
Annual income in 2010	2.9	3.2	3.1	3.3	2.7	3.0
Days worked in 2010	220	178	220	171	219	186
Annual income in 2011	3.6	3.6	3.8	3.7	3.4	3.4
Days worked in 2011	255	183	253	172	258	199
Annual income in 2012	4.1	4.0	4.3	4.1	3.8	3.9
Days worked in 2012	266	198	264	183	268	219
Annual income in 2013	4.7	4.3	4.9	4.5	4.3	3.9
Days worked in 2013	287	125	286	125	289	125
Annual income in 2014	4.9	4.7	5.2	4.8	4.5	4.3
Days worked in 2014	285	115	285	114	284	117

Source: PILA. Income in thousands of 2013 USD (1 USD = 1869.1 Colombian pesos). Workers can have more than one job, which explains the high values in the report of days worked.

Table 2: Descriptive Statistics of Statuses

	All Students		Males		Females	
	Mean	Sd	Mean	Sd	Mean	Sd
	(1)	(2)	(3)	(4)	(5)	(6)
Apprenticeship	.008	.09	.008	.089	.008	.092
Enrolled in Higher Ed.	.332	.471	.316	.465	.349	.477
Incomplete	.186	.389	.189	.392	.181	.385
Incomplete in the 1st year	.122	.327	.127	.332	.117	.322
Incomplete after the 1st year	.064	.244	.063	.243	.064	.245
Active	.009	.094	.009	.096	.008	.091
Candidate	.018	.134	.018	.131	.019	.136
Graduated	.119	.324	.1	.3	.14	.347
Graduated on time	.084	.277	.069	.253	.1	.3
Graduated late	.035	.185	.031	.174	.04	.197
Bachelor	.117	.321	.098	.297	.137	.344
Postgraduate	.003	.051	.002	.044	.003	.058
Diploma	.003	.052	.002	.043	.004	.06
Master	0	.013	0	.012	0	.014
PhD	0	.009	0	.01	0	.008

Source:SENA, SNIES, and SPADIES.

Table 3: Descriptive Statistics of Using Variables

	All Students		Males		Females	
	Mean	Sd	Mean	Sd	Mean	Sd
	(1)	(2)	(3)	(4)	(5)	(6)
Female	.479	.5	0	0	1	0
Income over 3 mmw	.188	.391	.185	.388	.192	.394
Student from a public secondary school	.665	.472	.67	.47	.66	.474
Student from an urban secondary school	.715	.451	.703	.457	.729	.445
Student from a Bogota school	.12	.325	.101	.301	.142	.349
Student from a outside Bogota school	.88	.325	.899	.301	.858	.349
Student with high test score	.297	.457	.328	.469	.263	.44
Student in a professional program	.21	.407	.186	.389	.235	.424
Student in a STEM program	.111	.314	.142	.349	.077	.266
Student enrolled in a high quality HEIs	.116	.321	.118	.323	.114	.318
Student reported as self-employed ‡	.022	.145	.02	.14	.023	.15
Student reported as public servant ‡	0	.02	0	.019	0	.02
Student with early enrollment †	.216	.412	.201	.401	.232	.422
Observations	4,489,955		2,340,821		2,149,134	

Source: ICFES, SPADIES, PILA. ‡ as percent of all students. † Student enrolled in a HEI $<$ 18 months after secondary

Table 4: Merge Results between SPADIES and PILA

Student in...	PILA		Total
SPADIES	NO	YES	
NO	2,680,704	319,128	2,999,832
	59.7%	7.1%	66.8%
YES	1,390,552	99,571	1,490,123
	31.0%	2.2%	33.2%
Total	4,071,256	418,699	4,489,955
	90.7%	9.3%	100.0%

Source:ICFES, SPADIES and PILA.

Table 5: Main Results

	Pooled Data	IV First Step Probit Mfx	IV First Step OLS	IV Second Step
	(1)	(2)	(3)	(4)
Went to college	0.049***			0.467***
	(0.002)			(0.036)
Apprenticeship	-0.331***	-0.081***	0.137***	-0.461***
	(0.002)	(0.001)	(0.001)	(0.040)
Female	-0.049***	-0.019***	0.026***	-0.064***
	(0.002)	(0.000)	(0.001)	(0.005)
Experience	0.206***	0.000	-0.005***	0.216***
	(0.001)	(0.000)	(0.000)	(0.004)
Experience^2	-0.006***	-0.000	0.000***	-0.007***
	(0.000)	(0.000)	(0.000)	(0.000)
Secondary test score	0.002***	0.005***	-0.002***	-0.005**
	(0.000)	(0.000)	(0.000)	(0.002)
Urban school	0.017***	0.043***	-0.011***	-0.048**
	(0.002)	(0.000)	(0.001)	(0.020)
Public school	-0.060***	-0.033***	0.001	0.009
	(0.002)	(0.000)	(0.001)	(0.021)
More than 3mmw	0.066***	0.115***	0.047***	-0.276***
	(0.002)	(0.000)	(0.002)	(0.105)
Self-employeed	-0.081***	-0.066***	0.136***	-0.240***
	(0.002)	(0.001)	(0.001)	(0.049)
Public servant	0.213***	-0.051***	0.087***	0.130***
	(0.009)	(0.003)	(0.003)	(0.028)
$\widehat{Attend}$			0.922***	
			(0.012)	
Distance to HEI		-0.003***		
		(0.000)		
Persons	410,049	4,483,503	410,049	410,049
Observations	2,074,267	23,965,734	2,074,267	2,074,267
R^2	0.085		0.114	0.063
IV F-stat			5577	
Pseudo R^2		0.166		
χ^2 p-value		0		

Notes: The table shows the coefficients of the regressions corresponding to Equation 1 and Equation 9. Experience measured in years. Secondary test score is the percentile in the Saber 11 test. Urban school, Public school, Household Income over 3 mmw, self-employed, and public servant are dummies for the individual. In columns (1), (2), and (4), the dependent variable is expressed in logarithm. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Disaggregated Results by Status

	Log Income (1)	Log Income (2)	Log Income (3)	Income (4)	Income (5)	Income (6)	Income (7)	Income (8)
Graduated	0.184***	0.156***	0.204***	-0.110***	-1.782***	-2.584***	-2.069***	-1.837***
	(0.003)	(0.003)	(0.010)	(0.026)	(0.032)	(0.038)	(0.061)	(0.062)
Candidate	0.019***	0.030***	0.032***	0.225***	0.226***	0.254***	0.153***	0.156***
	(0.006)	(0.006)	(0.012)	(0.024)	(0.024)	(0.024)	(0.047)	(0.048)
Retired	0.003	0.016***	0.023**	0.135***	0.125***	0.147***	0.033	0.045
	(0.002)	(0.002)	(0.010)	(0.008)	(0.008)	(0.008)	(0.041)	(0.042)
Active	-0.028***	-0.010		0.120***	0.104***	0.133***		
	(0.010)	(0.010)		(0.039)	(0.039)	(0.039)		
Apprenticeship	-0.334***	-0.326***	-0.260***	-0.788***	-0.780***	-0.782***	-0.766***	-0.748***
	(0.002)	(0.002)	(0.004)	(0.010)	(0.010)	(0.010)	(0.018)	(0.019)
Female	-0.053***	-0.053***	-0.014***	-0.425***	-0.429***	-0.428***	-0.256***	-0.314***
	(0.002)	(0.002)	(0.003)	(0.006)	(0.006)	(0.006)	(0.013)	(0.014)
Experience	0.206***	0.205***	0.209***	0.342***	0.342***	0.342***	0.298***	0.290***
	(0.001)	(0.001)	(0.003)	(0.005)	(0.005)	(0.005)	(0.011)	(0.011)
Experience^2	-0.006***	-0.006***	-0.005***	0.004***	0.001***	0.001***	0.013***	0.014***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Secondary test score	0.002***	0.002***	0.003***	0.006***	0.007***	0.006***	0.008***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Urban school	0.018***	0.038**	0.015***	0.100	0.092	0.096	0.042***	-0.678***
	(0.002)	(0.019)	(0.004)	(0.072)	(0.072)	(0.072)	(0.014)	(0.153)
Public school	-0.059***	0.042*	-0.044***	0.237***	0.235***	0.228***	-0.249***	0.873***
	(0.002)	(0.022)	(0.003)	(0.084)	(0.084)	(0.084)	(0.014)	(0.185)
More than 3mmw	0.055***	0.021***	0.038***	0.208***	0.221***	0.219***	0.467***	0.220***
	(0.002)	(0.003)	(0.004)	(0.012)	(0.012)	(0.012)	(0.018)	(0.022)
Self-employeed	-0.083***	-0.087***	-0.181***	-0.962***	-0.977***	-0.976***	-1.732***	-1.705***
	(0.002)	(0.002)	(0.004)	(0.009)	(0.009)	(0.009)	(0.017)	(0.017)
Public servant	0.208***	0.183***	0.155***	-0.444***	-0.488***	-0.486***	-1.405***	-1.261***
	(0.009)	(0.009)	(0.016)	(0.028)	(0.028)	(0.028)	(0.052)	(0.053)
Graduate X Secondary test score				0.019***		0.017***	0.019***	0.013***
				(0.000)		(0.000)	(0.000)	(0.001)
Graduate X Experience					0.397***	0.384***	0.297***	0.298***
					(0.004)	(0.004)	(0.005)	(0.005)
FE Sec. School	No	Yes	No	Yes	Yes	Yes	No	Yes
FE HEI	No	No	Yes	No	No	No	Yes	Yes
FE Department	Yes	No	No	No	No	No	No	No
Persons	410,049	409,188	98,767	421,053	421,053	421,053	103,117	102,747
Observations	2,074,267	2,073,506	522,076	2,097,731	2,097,731	2,097,731	531,835	530,944
R^2	0.087	0.108	0.132	0.113	0.116	0.117	0.168	0.249

Notes: The table shows the coefficients of the regressions corresponding to Equation 1. Graduate takes the value of 1 if the student graduated from higher education; Candidate takes the value of 1 if the student attended more than 90 percent of the program but did not receive the degree; incomplete is the student who is absent for two or more consecutive semesters without registering a degree and has less than 90 percent of the program; Apprentice is the student we found as a SENA intern. Years after secondary is the time elapsed since the time of the secondary school degree and the apparition in the Social Security (a proxy to experience). The score on the ICFES test is the percentile per student. Urban school, Public school, Income of household over 3 mmw, self-employed, and public servant are dummies for the individual; the last two are dynamic in time. Income is in thousands of dollars of 2013 (1 US dollar = 1,869.1 Colombian pesos in 2013). In columns (1), (2), and (3), the dependent variable is expressed in logarithm, the others are measured in levels according with Farber and Gibbons (1996)'s test. Columns (7) and (8) analyze the results for the 2002 cohort. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Disaggregated Results by Status and Maximum Degree Level Obtained

	Log Income (1)	Log Income (2)	Log Income (3)	Income (4)	Income (5)	Income (6)	Income (7)	Income (8)
Bachelor	0.179***	0.151***	0.066	-0.123***	-1.787***	-2.578***	0.385	0.633
	(0.003)	(0.003)	(0.137)	(0.026)	(0.032)	(0.038)	(0.566)	(0.562)
Diploma	0.395***	0.336***	0.228	1.145***	-0.671***	-1.550***	1.425**	1.680***
	(0.019)	(0.019)	(0.139)	(0.077)	(0.080)	(0.083)	(0.571)	(0.567)
Master	0.648***	0.532***	0.483***	2.020***	0.143	-0.758**	2.665***	2.806***
	(0.083)	(0.084)	(0.160)	(0.301)	(0.301)	(0.302)	(0.646)	(0.642)
Phd	0.150	0.074		-0.723	-2.394***	-3.339***		
	(0.140)	(0.141)		(0.542)	(0.541)	(0.542)		
Candidate	0.019***	0.030***	-0.102	0.226***	0.227***	0.255***	2.600***	2.623***
	(0.006)	(0.006)	(0.138)	(0.024)	(0.024)	(0.024)	(0.568)	(0.565)
Retired	0.004*	0.016***	-0.111	0.136***	0.125***	0.147***	2.480***	2.511***
	(0.002)	(0.002)	(0.137)	(0.008)	(0.008)	(0.008)	(0.568)	(0.564)
Active	-0.028***	-0.009	-0.134	0.121***	0.105***	0.134***	2.447***	2.467***
	(0.010)	(0.010)	(0.138)	(0.039)	(0.039)	(0.039)	(0.569)	(0.566)
Apprenticeship	-0.334***	-0.326***	-0.260***	-0.788***	-0.780***	-0.782***	-0.767***	-0.747***
	(0.002)	(0.002)	(0.004)	(0.010)	(0.010)	(0.010)	(0.018)	(0.019)
Female	-0.053***	-0.053***	-0.014***	-0.426***	-0.429***	-0.429***	-0.257***	-0.316***
	(0.002)	(0.002)	(0.003)	(0.006)	(0.006)	(0.006)	(0.013)	(0.014)
Experience	0.206***	0.205***	0.209***	0.342***	0.342***	0.342***	0.298***	0.290***
	(0.001)	(0.001)	(0.003)	(0.005)	(0.005)	(0.005)	(0.011)	(0.011)
Experience^2	-0.006***	-0.006***	-0.005***	0.004***	0.001***	0.001***	0.013***	0.014***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Secondary test score	0.002***	0.002***	0.003***	0.006***	0.007***	0.006***	0.008***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Urban school	0.018***	0.038**	0.015***	0.100	0.092	0.096	0.041***	-0.681***
	(0.002)	(0.019)	(0.004)	(0.072)	(0.072)	(0.072)	(0.014)	(0.153)
Public school	-0.059***	0.042*	-0.044***	0.237***	0.234***	0.228***	-0.248***	0.869***
	(0.002)	(0.022)	(0.003)	(0.084)	(0.084)	(0.084)	(0.014)	(0.185)
More than 3mmw	0.054***	0.021***	0.038***	0.208***	0.221***	0.218***	0.464***	0.219***
	(0.002)	(0.003)	(0.004)	(0.012)	(0.012)	(0.012)	(0.018)	(0.022)
Self-employeed	-0.083***	-0.087***	-0.181***	-0.962***	-0.977***	-0.976***	-1.731***	-1.704***
	(0.002)	(0.002)	(0.004)	(0.009)	(0.009)	(0.009)	(0.017)	(0.017)
Public servant	0.208***	0.183***	0.154***	-0.447***	-0.490***	-0.488***	-1.414***	-1.270***
	(0.009)	(0.009)	(0.016)	(0.028)	(0.028)	(0.028)	(0.052)	(0.053)
Graduate X Secondary test score				0.019***		0.016***	0.018***	0.013***
				(0.000)		(0.000)	(0.000)	(0.001)
Graduate X Experience					0.394***	0.382***	0.295***	0.296***
					(0.004)	(0.004)	(0.005)	(0.005)
FE Sec. School	No	Yes	No	Yes	Yes	Yes	No	Yes
FE HEI	No	No	Yes	No	No	No	Yes	Yes
FE Department	Yes	No	No	No	No	No	No	No
Persons	410,049	409,188	98,767	421,053	421,053	421,053	103,117	102,747
Observations	2,074,267	2,073,506	522,076	2,097,731	2,097,731	2,097,731	531,835	530,944
R^2	0.087	0.108	0.132	0.114	0.116	0.117	0.168	0.250

Notes: The table shows the coefficients of the regressions corresponding to Equation 1 . Bachelor's degree if the student did not continue his or her studies after obtaining the higher education degree; Diploma if the student enrolled in a diploma; Master's degree if the student enrolled in a master's degree program; and PhD if the student enrolled in a doctorate program. Candidate takes the value of 1 if the student attended more than 90 percent of the program but did not receive the degree; incomplete is the student who is absent for two or more consecutive semesters without registering a degree and has less than 90 percent of the program; Apprentice is the student we found as a SENA intern. Years after secondary is the time elapsed since the time of the secondary school degree and the apparition in the Social Security (a proxy to experience). The score on the ICFES test is the percentile per student. Urban school, Public school, Income of household over 3 mmw, self-employed, and public servant are dummies for the individual; the last two are dynamic in time. Income is in thousands of dollars of 2013 (1 US dollar = 1,869.1 Colombian pesos in 2013). In columns (1), (2), and (3), the dependent variable is expressed in logarithm, the others are measured in levels according with Farber and Gibbons (1996)'s test. Columns (7) and (8) analyze the results for the 2002 cohort. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.