LUND UNIVERSITY
School of Economics and Management

Master's Programme in Economic Development and Growth

Trends of intergenerational transmission of education

Evidence from Colombia

by

Ottavia Anna Veroux

oaveroux@hotmail.it

Inequalities among individuals can arise from different amounts of efforts and abilities or from circumstances that go beyond the control of the individual; opposite ethical and economical conclusions arise from these differences but the literature on inequality has often overlooked the issue. This thesis focuses on the concept of inequality of opportunity in its acceptation of intergenerational transmission of education, studying how parents' and children's education are related. The trends of intergenerational transmission of education is analysed in Colombia using data over a period of 40 years, from 1957 to 1996. Through the computation of eight different indices, the paper creates an exhaustive picture of the phenomenon that can distinguish between different trends in absolute and relative educational mobility. The main findings illustrate the high enhancements in terms of absolute mobility, but they also highlight the lack of improvements in the level of relative mobility. The increase in average schooling achievements in the last decades has permitted individuals to be on average more educated than their parents. Nevertheless, this has not changed their position in the distribution of educational achievements, meaning that individuals from disadvantaged families still face lower opportunities.

Keywords*: Intergenerational Mobility; Inequality of Opportunity; Inequality; Education; Family Background; Colombia*

EKHS42
Master's Thesis (15 credits ECTS)
June 2020
Supervisor: Luciana Quaranta
Examiner: Tobias Karlsson
Word Count: 15,484

Table of contents

List of tables

List of figures

1.Introduction

Inequality is a widely debated topic in economics, with several studies within and across countries about both its causes and consequences. Nevertheless, the literature on the effects that inequality has on economic growth and development is inconclusive. Some authors find a certain degree of inequality to be necessary to enhance economic growth through its positive effects on the level of savings, investments, and effort (Marrero & Rodríguez, 2013). On the other hand, in presence of credit markets imperfections, inequality generates inefficiency moving away the endowments of an economy from their most profitable employments, with detrimental effects on growth: indeed, investments in both human and physical capital often require a minimum initial amount of resources, lacking which poor individuals are not able to increase their educational level or to start economic activities even if that would be profitable (Galor & Zeira, 1993; Banejeree & Newman, 1993). The opposite effects on growth would eventually have an impact also on the poorest individuals in the society, since enhancing economic growth has been found to be the main driver for poverty reduction (The World Bank, 2006).

These conflicting results arise from the fact that the literature has usually focused on inequality of output, without specifying the causes and channels behind it. Indeed, inequality derived from different amounts of effort produces consequences that differ with respect to disparities originated by familiar background, discrimination, or other elements that the individual cannot control. The different causes behind output disparities produce opposite effects on economic development and growth (for instance the ones cited above), as well as generating different ethical judgments about inequality. Within this issue, Amartya Sen (1980) makes an important contribution differentiating *functionings* and *capabilities* and putting particular emphasis on the latter: the output itself (i.e. functioning) does not give the right information about the wellbeing of an individual because it also depends on effort and preferences; it is rather more useful to examine the range of possibilities "*to do and to be*" (i.e. capabilities, or set of possible functionings) that each individual has. It is from this differentiation that the concept of inequality of opportunity arises, defined as the inequality in the outcome that originates from *circumstances* that are outside the control of each individual (e.g. from family background, gender, ethnic origins) (Roemer, 1998). It is precisely inequality of opportunities that strongly relates to the negative effects on economic growth cited above since it prevents individuals from disadvantaged groups to increase their education or to invest in projects regardless of their abilities or efforts.

Inequality of opportunities arises from different factors and affects several aspects of people's wellbeing. This paper focuses on the intergenerational transmission of education, i.e. on how parental educational background affects the decisions of human capital investments of their children. In this framework, inequality of opportunity translates into intergenerational immobility: the educational level is *inherited* from parents to children across generations, producing fewer opportunities for individuals coming from disadvantaged families and creating a poverty trap that in the long run generates dynasties

of non-educated and poor families (Galor & Zeira, 1993). Parental educational background affects children's opportunities through its relation with the amount of initial endowments, personal preferences, and the transmission of abilities in the first years of life (Roemer, 2004).

This thesis aims to analyse the trend that intergenerational transmission of education had in Colombia in the last decades, in order to answer the research question on whether the relationship between parents' and children's education has changed across different cohorts (i.e. for individual born between 1957 and 1996). To do so, I compute several indices of intergenerational mobility using microdata retrieved in the *Encuesta Nacional de Calidad de Vida* in Colombia in 2018. The computation of different indices permits to thoroughly explore the topic analysing the various facets that mobility can have, going beyond the mere regression coefficient which is usually employed in this type of analysis. Indeed, I find that the regression coefficient has significantly decreased for the youngest cohorts, showing that an additional year of parental education translates today to a smaller increase in children's education respect to the past. Nevertheless, when the changes in distribution are taken into account, other indices suggest that the improvements in intergenerational mobility in the country have been mostly structural, i.e. related to an average improvement in educational achievement rather than to better opportunities for the disadvantaged groups.

Similarly to the rest of Latin America, Colombia stands out for its levels of both income and opportunity inequality (Hertz, et al., 2007; Deininger & Squire, 1998). Despite that, while the literature on income inequality in Latin America has been historically important, only a few and recent studies tried to analyse trends, causes, and consequences of inequality of opportunity in Latin American countries (Roemer & Trannoy, 2015; Binder & Woodruff, 2002). In its report about social mobility[1], the OECD (2018) found Colombia to be the country with the highest level of inequality of opportunity among the ones considered in the study[2] using the elasticity of incomes between generations at the bottom of the distribution: 300 years (i.e. ten generations at least) are needed for a Colombian individual from the bottom decile of the income distribution to reach the average income level of the country. Moreover, parental education appears to be an important circumstance that influences children's opportunities: almost the totality of individuals at the bottom of the income distribution comes from families with low educational level (Paes de Barros, et al., 2008). Average educational achievements have improved in the country, nevertheless, primary school is today the only level that is close to reach an enrolment rate above 90%, school quality is further below the OECD average, and still many individuals drop out before the age of 15, i.e. the age of mandatory schooling. These difficulties of the educational system affect the entire population, but they are particularly harsh for individuals from more disadvantaged groups, i.e. from ethnic minorities, from rural places or with low socioeconomic status, lowering the

[1] OECD, 2018. *A broken social elevator? How to promote social mobility,* Paris: OECD Publishing
[2] i.e. the OECD countries plus Argentina, Indonesia, China, Brazil, South Africa, and Colombia.

probability for those individuals to reach a sufficient educational level (Government of Colombia, 2014).

The main concepts described in this thesis are strongly related to the goals that the same Government of Colombia set for the country in its National Plan for Development (*Plan Nacional de Desarrollo*) for the period 2014-2018 and today for the next years from 2018 until 2022. Indeed, the three pillars on which the country aims to establish the future economic growth of its society are peace, equality (also seen in its acceptation of equality of opportunity), and education (Government of Colombia, 2014; Government of Colombia, 2019):

"*Education, precisely the third pillar of this Plan, is conceived as the most powerful tool for social equality, since it not only equalizes opportunities of the individuals, but it opens the doors of the progress and it enhances the quality of the democracy*."[3]
(Government of Colombia, 2014).

The remainder of the thesis is organized as follows. Chapter 2 reviews the main theoretical and empirical literature on the inequality of opportunity and intergenerational mobility of education, as well as describing the main features and difficulties of the educational system in Colombia. Chapter 3 introduces the *Encuesta Nacional de Calidad de Vida* and the variables employed in the analysis. Chapter 4 presents the indices computed in the study, whose results (and robustness checks) are presented in chapters 5 and 6. Finally, chapter 7 concludes and states possible future research on the topic and policy implications.

[3] Original text, translated by the author: *"La educación, precisamente el tercer pilar de este Plan, se concibe como el más poderoso instrumento de igualdad social, pues no solo nivela las oportunidades de la personas, sino que abre puertas de progreso y mejora la calidad de la democracia."*

2.Literature review

In this chapter, I present the most important literature review on the topic. I first introduce the concepts of inequality of opportunity and intergenerational transmission of education, explaining how both are related to economic growth and individual wellbeing. I then present the empirical evidence on the topic, reporting the main findings on the relationship between parents' and children's education in several countries using different indices. Finally, the chapter concludes with information about the levels of inequality of opportunity and intergenerational mobility in Colombia and with a description of the main features and issues of the educational system of the country.

2.1 Theoretical approach

2.1.1 Inequality of opportunity

When talking about inequality it is fundamental to start with specifying "*inequality of what*". Previous theoretical and empirical literature that looks at the effects of inequality on economic growth has not reached a conclusive answer because inequality affects the economy through different and opposite channels thus the final outcome depends on which channel is predominant (Marrero & Rodríguez, 2013). Such ambiguous relationship between inequality and growth could be related to the fact that the concept of inequality generally used comprehends different types of inequality without differentiating them. Indeed, different types of inequality have diverse consequences both ethically and economically depending on what drives them (Ferreira & Gignoux, 2011). Economically, it has been proved that a certain degree of income inequality can be optimal for growth because it enhances investments and consumptions; moreover, inequality could arise from different amounts of effort and abilities, factors that should be rewarded in order to create higher incentives for people to invest in human capital and to put more effort in their actions (Marrero & Rodríguez, 2013). Nevertheless, some inequalities do not depend on personal effort, but that are generated by characteristics that are exogenous to the individual and that could be detrimental for growth; therefore, it becomes substantial to solve the issue that opened this paragraph and to examine whether disparities are a product of different amounts of effort or whether they depend on characteristics that are exogenous to the individual.

Equality of opportunities, as defined by Roemer (1998), requires the results to be independent of any circumstance that is not in control of the individual, thus there is inequality of opportunity when personal characteristics that the individual cannot affect have an impact on some outcome (e.g. income, education, health). A natural point of start when talking about inequality of opportunities is the capability approach of Amartya Sen (1980) in which he defines the difference between *capabilities* and *functionings*: the focus of policymakers should not be on the final output itself (i.e. the functionings) but on capabilities, defined as "*what people are able to do and be*" (Sen, 1980). In this way, the

author highlights the importance of equal opportunities but still leaving space to individual aspiration and effort (Robeyns, 2005).

The previous definitions do not solve the initial problem of "*inequality of what*", in particular when we are concerned about measurements and empirical analysis. Authors such as Rawls and Dworkins included in the "*initial circumstances*" every variable not in control of the individual, therefore comprehending also characteristics such as genes and innate talents (Roemer & Trannoy, 2015). More in general, inequality of opportunities can arise from three different factors: personal characteristics (e.g. talent and motivation), discrimination, and distribution of "basic opportunities" both during childhood and adulthood (e.g. nutrition, education, job opportunities, etc) (Paes de Barros, et al., 2008; Blanden, 2013). Nevertheless, when moving the discussion from a philosophical point of view[4] to a more economical one, it becomes important to study those variables that not only are exogenous to the individual but that can also be observed and changed through policies or which influence on the final output can at least be weakened. For instance, the policymaker should have the goal to eradicate discriminations due to gender, ethnic group, social classes or to decrease the impact that the family background has on different adult outcomes. The latter is the main ground from which this thesis starts: as it will be explained in the next paragraphs, parents' education and socioeconomic status have a high impact on children's education and thus on future income. Public policies can have an important role in weakening such relationship in a way that personal educational achievements depend more on personal abilities and choices and less on external circumstances.

The focus on inequality of opportunity instead of using the mainstream concept of inequality of output (be it income, consumption, or other) is not just trivial: indeed, the two concepts are not different ways to measure the same thing but they rather capture different phenomena that affect the distribution in a society (Behrman, et al., 2000; Ferreira & Gignoux, 2011). The evidence of such difference is given by the fact that cross-countries rakings of the two measurements provide close but different results (Ferreira & Gignoux, 2011). Indeed, the traditional measures are unable to capture the differentiation between inequalities produced by external circumstances and the ones produced by choices and effort. Leaving behind ethical arguments on which type of inequality is more acceptable, the economic effects of the two measures are different too. Indeed, inequality of opportunity more than inequality itself is related to economic inefficiency since it entails that investments in human and physical capitals do not depend on abilities and effort but on external characteristics, diverting resources from the most efficient and productive investments (Paes de Barros, et al., 2008; Blanden, 2013; Ferreira & Gignoux, 2011; Roemer & Trannoy, 2015; The World Bank, 2006; Heckman, 2012). Moreover, the importance of inequality of opportunities could rely on the fact that it is more likely to be seen as unfair by most of the people: this produces higher social conflicts when its level is substantial and, moreover, it makes easier to implement public policies because

[4] The philosophical discussion behind those definitions is beyond the scope of this paper, I refer to Roemer (2015) for more information on the topic

it changes the attitudes toward redistribution (Paes de Barros, et al., 2008; Ferreira & Gignoux, 2011). Nevertheless, it is important to take into consideration that a measure of inequality of opportunity is more complicated to obtain than just inequality because it is a more multidimensional concept and it requires a lot of information, also about variables that can difficultly be observed or that refer to older generations or the childhood of the individual interviewed (Roemer & Trannoy, 2015; Marrero & Rodríguez, 2013).

To sum up, to focus on inequality itself could be easier because it only takes into account the differences in terms of the chosen variable (e.g. income); nevertheless, in many situations such as the implementation of social policies the final goal is often to give to everyone the same opportunities and the same set of choices rather than the same output. From this point of view, the literature needs to focus more on the concept of inequality of opportunity and this is the framework in which this paper aims to contribute.

2.1.2 Intergenerational transmission of education

As stated above, inequality of opportunity could arise from different factors. When talking about Latin America, variables such as gender, ethnic group, place of residence, socioeconomic status, and parents' education are the most important, with different weights depending on the specific country (Paes de Barros, et al., 2008; Ferreira & Gignoux, 2011). Intergenerational immobility represents an important factor of inequality of opportunity, with the two measures having a strong correlation equal to 0.6 when measured in a cross-country analysis (Brunori, et al., 2013). Intergenerational immobility is defined as the transmission of outcomes across generations, with parental characteristics affecting children outcomes; in this case, the "*circumstances*" of interest could be several, among which parent's earnings or education. When there is intergenerational transmission of education the level (and quality) of each individual's education depends on his own parents' education; using the vocabulary of the inequality of opportunity seen above, parents' education would represent the "*external circumstance*", while the output is the final education of the child (Roemer, 2004). Following Roemer (2004), the family background affects opportunities throughout four different channels, namely the initial endowments (of both physical resources and social connections), the formation of beliefs and skills through family culture and investments, the genetic transmission of abilities, and the development of preferences and aspirations. Most of the previous channels – in particular the firsts two - are directly or indirectly related to education and they could be the object of public policies that aim to reach more equal opportunities (Roemer, 2004; Roemer & Trannoy, 2015).

Another way to classify the different channels through which parents' education could affect child investments in human capital is to differentiate between direct and indirect effects. The direct relation is straightforward but at the same time highly difficult to measure or modify: first of all, more educated parents are more aware of returns of schooling, secondly they are more able to transmit hard and soft skills to their children through the family environment (Marrero & Rodríguez, 2013). The indirect relationship between parents' and children's level of human capital works throughout the

socioeconomic status of the family, which is affected by parental education and which, on the other side, affects investments in children education and nutritional status during childhood (Paes de Barros, et al., 2008). Indeed, several empirical papers found a correlation between proxies for childhood living standards (e.g. height and birth weight) and educational achievement (Behrman, 1996; Currie, 2009; Currie & Vogl, 2013; Bouillon & Tejerina, 2007). The theoretical explanation for this correlation is found by the technophysio evolution theory developed in the last years: it has been proved that if the net nutritional status - dependent by the total effect of nutrients and diseases - during childhood is critically low, the body changes its size and structure in order to need fewer resources (Fogel, 2004). These changes affect primarily future health and mortality, but also brain development and cognitive skills (Fogel, 2004; Currie, 2009), with several negative effects on human capital accumulation.

The importance of nutrition and health during childhood and the interaction between those factors and educational achievement are the reasons why many policies fostering human capital try to alleviate the three issues at the same time, for instance providing together grants for nutritional supplements, health checks, and grants for school enrolment (Legovini & Regalia, 2001). Moreover, the theory cited above gives more directions about which policies should be preferred if the aim is to reduce the intergenerational transmission of education. Indeed, since the negative effects of inadequate nutrients or lacks of intellectual stimulus tend to accumulate across time (The World Bank, 2006), more consideration must be put on policies that intervene during childhood: late interventions can reduce income inequality and poverty rate but they do not directly increase social mobility across generations, and their effectivity depends on the quality of earlier interventions (Heckman, 2012). Thus, public policies such as job training or adult schooling produce low economic returns, while policies that benefit children during the pre-school period or primary school are more effective and their returns are more easily higher than their opportunity costs (The World Bank, 2006; Carneiro & Heckman, 2003; Heckman, 2012).

The intergenerational transmission of education is clearly negative for individuals from disadvantaged families generating ethical issues, but it is also negative for the entire economy with effects in the long run in terms of distribution, efficiency, and growth. The long-term result is a poverty and inequality trap that creates a persistence of poor and non-educated dynasties: uneducated individuals from non-schooled parents will not have enough resources themselves to invest in human capital, thus their own children will be not educated as well, transmitting the low socioeconomic status to future generations (Galor & Zeira, 1993; Demirguc-Kunt & Levine, 2008; The World Bank, 2006). When the relationship between children's and parents' education is stronger it means that the amount of investments in human capital does not depend on ability or effort, thus the resources of the economy are not directed towards the most productive individuals; that translates to losses in efficiency, productivity and human capital accumulation, thus eventually to low economic growth (Blanden, 2013; Marrero & Rodríguez, 2013). The significance of this relation depends on different characteristics of the economy, being stronger with a lack of public spending in education, inefficiency in the financial markets

or bad quality institutions (Galor & Zeira, 1993; The World Bank, 2006; Blanden, 2013; Behrman, et al., 2000; Hertz, et al., 2007).

The same phenomenon of intergenerational mobility could be captured using other variables rather than education, such as income or social class, which would capture more directly the transmission of socioeconomic status; nevertheless, several reasons make education a better option when analysing inequality of opportunity. First of all, education is correlated with both earnings and occupation but also with many other measures of well-being, such as early-life conditions, health, and resilience to shocks (The World Bank, 2006). Moreover, using education means to study one of the main channels through which inequality of opportunity works (Hertz, et al., 2007) since an important portion of inequality of opportunity translate into low human capital accumulation (Marrero & Rodríguez, 2013). Finally, there is the more practical but substantial issue of data availability: information about the education of both the individual and the parents are available in most of the household surveys and variables such as enrolment rate or years of education are quite unambiguous concepts and lead to fewer measurements errors (while for example "social class" can be defined in many different ways) (Black & Devereux, 2010). Additionally, education is practically fixed after a certain age, whereas the income changes over the lifecycle, generating the need to compute the permanent income for the individual. On the other hand, the information about education in the dataset lack to capture many aspects that could be particularly relevant in developing countries. For instance, in the poorest regions many people have no access to formal schooling but still have a positive amount of human capital accumulated by other means and, even with formal schooling, the years of education do not capture substantial differences in the quality of schooling (that can be captured using data such as the PISA[5] results) (Hertz, et al., 2007).

2.2 Empirical evidence

The empirical evidence sustains the theory cited above, proving that several outputs and their distribution depend on external circumstances that cannot be controlled by the individual (Brunori, et al., 2013): in particular, the correlation between adult socioeconomic status and parents' education is found to be statistically and economically significant in every society where the analysis has been conducted (Hertz, et al., 2007). Nevertheless, the magnitude and the significance of each circumstance differ depending on the country or time analysed and on the method used: indeed, socioeconomic characteristics such as the quantity and quality of public expenditure, institutions, ethnic divisions, etc. influence the weight of each circumstance on the final output (Brunori, et al., 2013). Likewise, similar instruments used by the policymakers could have different

[5] Programme for International Student Assessment developed by the OECD: throughout standardized tests, PISA measures 15 years olds' ability in reading, mathematics and science, permitting to compare school quality and outputs of the different countries.

results, reducing the importance of circumstances in some environment but producing no effects in others.

Although the literature on the topic is still not particularly expanded, some authors tried to measure the level of inequality of opportunity or intergenerational mobility for different countries studying the trends over time. Hertz et al. (2007) study the trends of intergenerational transmission of education in 42 nations in a timeframe of fifty years[6] in order to find global patterns of the last half-century as well as differences between countries. The authors find that the intergenerational persistence of education decreased in most of the countries in the sample over the fifty years analysed but that the correlation between parents' and children's education did not vary on average. This means that a year increase in parents' education has today less effect on children's education, but the causal connection has not changed (i.e. the intergenerational correlation represented by the R^2 has not decreased). Despite the improvements during the last decades, Latin America is the region with the lowest levels of intergenerational mobility; on the other hand, northern European countries occupy the first positions of the ranking. Hertz et al. (2007) find that the average correlation between parents' and children level of education is 0.60 in the seven Latin American countries analysed while the average regression coefficient is 0.79; the magnitude of the two indices stands out compared to the same values of other regions such as northern Europe (0.36 and 0.55), Asia (0.39 and 0.51), or US (0.46 and 0.46). For Colombia, the authors find that one additional year of parental education is related to an average increase in children education by 0.80 years, but the coefficient decreases over the fifty years studied being around 0.6 in 1975; the correlation coefficient is found to be only slightly decreasing from 0.7 to around 0.5. Other authors find similar results with different sets of countries or using income mobility (Blanden, 2013; Brunori, et al., 2013).

Although the regression coefficient is the most widely index used by the literature and it has the quality of being easily interpretable, it does not always give an exhaustive picture of the trend in intergenerational mobility. Indeed, the coefficient does not differentiate between two different aspects of mobility, capturing both structural and exchange mobility. Structural (absolute) mobility identifies the absolute change between parents' and children's education, capturing the educational expansion in the country; on the other side, exchange mobility focuses solely on changes within the distribution. It could be the case that children with non-educated parents are more educated than their parents but with no changes in exchange mobility (i.e. in its position in the country's distribution of education) because such enhancements originate uniquely from a general improvement in school achievement in the country (Biagi & Stuhler, 2018; Neidhofer, et al., 2018). It becomes therefore useful to report measures of different indices, which derivation will be extensively explained in the following method's chapter.

Besides the most common regression coefficient and intergenerational correlation, Neidhofer et al. (2018) compute different indices for Latin America across time, that together give a more extensive picture of intergenerational mobility in the region.

[6] From 1916 to 1983 with the exact years depending on the country. For Colombia, the data are available from 1928 to 1977.

Similarly to other authors, they find that the regression coefficient has decreased over time from 0.6 to 0.4 respectively for individuals born between 1940 and 1980, whereas the intergenerational correlation keeps almost steady across time around 0.5. The different trend between the two index reveals that the region has experienced mostly improvements in absolute mobility rather than in exchange mobility. Differently to the other authors cited, Neidhofer et al. (2018) include also the computation of less widely used indices, for instance, the absolute mobility (M1), the directional mobility (M2) and indices of persistence at the tails (i.e. the different probability of high education for individuals from non-educated or highly educated parents). The results show that both M1 and M2 present an inverted U-shape respectively between 3.4 and 4.4, and between 2.5 and 3.5 depending on the cohort analysed, while there are slight improvements in terms of persistence at the tails: the upper-class persistence is almost steady across time around 0.7, but the youngest generations are more likely to be enrolled in higher grades even if they come from low educated parents (the probability goes from 0.15 to 0.4). The authors conclude that on average Latin America obtained improvements in terms of intergenerational mobility but that such enhancements are more related to absolute mobility obtained through an average increase in educational achievement rather than to exchange mobility obtained by changes in the distribution (Neidhofer, et al., 2018).

As seen in the previous paragraph, in several studies Latin America stands out for its low levels of equality of opportunity and intergenerational mobility, no matter the measure used. Nevertheless, researchers did not find a final reason that could explain this long-run pattern. One hypothesis is related to historical and institutional explanations: Latin American colonies were characterized by extractive institutions[7] that implied a high level of inequality and that translated to a delay in the expansion of public schooling outside the elite and to many other inequalities in income and in opportunities that characterize the region today (Hertz, et al., 2007; Acemoglu, et al., 2002). The current negative ranking persists despite the substantial efforts and improvements that most South American countries experienced in the last decades. Indeed, social spending has increased significantly in the last years with stronger benefits for the poorest part of the population and thus with a reduction of the poverty level (Bouillon & Tejerina, 2007). The raised social spending made possible the implementation of several public policies that targeted the most vulnerable individuals in the society. When speaking about inequality in educational opportunities, those policies took the form of different strategies with the aim to increase the human capital accumulation increasing both demand and supply of schooling. Some examples are conditional cash transfer programs[8] (*Progresa-Oportunidades* in Mexico, *Bolsa Familia* in Brazil, *Plan Familias* in Argentina, *Familias*

[7] Institutions created by the colonizers in territories that, given their natural endowments, were better places to extract resources rather than to settle in. Extractive institutions have characteristics (e.g. low protection of the property rights) that made it easier for the elite to extract resources from the rest of the population.

[8] Conditional cash transfers (CCT) are programs through which the Government provide a certain amount of money to eligible families conditional to some requirements. When talking about education, CCT require that the child attend a minimum amount of days of school and they are usually integrated with nutritional and health interventions.

en Acción in Colombia), schools construction (the infrastructure expansion in Argentina between 1994 and 2000), educational grants (*Beca Presidente de la Republica* in Chile, *Programa Nacional de Becas Estudiantiles* in Argentina) (Bouillon & Tejerina, 2007).

Intergenerational mobility has been proven to be correlated to several characteristics of a society, which influence the relationship between children's and parents' levels of education. In most of the countries, inequality and mobility are negatively related, meaning that they are affected by a third common factor or that they cause one another: indeed the distribution of resources affects future opportunities but also opportunities – and in this case education – influence future earnings (Blanden, 2013; Brunori, et al., 2013). Moreover, greater governmental expenditure in education has been proved to be on average effective to reduce inequalities in education (Hertz, et al., 2007; Biagi & Stuhler, 2018). Nevertheless, some authors find that an increase in expenditure has significant effects only when directed to primary education (Behrman, et al., 2000). The differences in results can be explained by the specificity of each country and each type of public expenditure. Many other characteristics of the economic environment are correlated with intergenerational mobility, such as financial depth and inflation: indeed, both financial market development and inflation affect the families consumption behaviour and their resilience to economic shocks, with effects on the demand for schooling (Behrman, et al., 2000).

2.3 Colombia

Colombia is the fifth biggest country in Latin America, and the third for the amount of population (almost 50 million in 2018). The primary administrative division is made into 33 departments, while geographically the country is divided into six regions (i.e. *Amazona*, *Andes*, *Caribe*, *Insular*, *Orinoquía,* and *Pacífica*). Most of the population is concentrated in the coastal territories in the north-west and in the urban centers (76% of the population); indeed, a high percentage of the population is settled in few larger cities, that are also the main economic fulcrum of the country (the capital *Bogotá* has more than 7 million inhabitants, followed by *Medellín*, *Cali*, and *Barranquilla*). On the other hand, the rural areas are today scarcely inhabited and suffer from a high level of underdevelopment with several issues in terms of infrastructure, poverty level, illiteracy, and violence (OECD, 2016).

The colonial past of the country (from 1492 until 1810, when it became independent from Spain) meant the disappearance of most of the indigenous population that originally lived in the region and its substitution with individuals with European or African origins. Indeed, today 86% of the population does not recognize in any ethnic minority, 10.62% is part of black, afro-Colombian, *raizales*, or *palenqueras* communities, 3.34% is from indigenous populations and 0.01% is Rom.

In 2018 Colombia was the fourth most unequal country in Latin America, with a Gini index equal to 50.4 (the rest of the countries for which the data are available goes from 38.6 in El Salvador to 53.9 in Brazil) (World Bank Database). From the 1960s Colombia

did not experience significant improvements in terms of inequality, with the Gini Index fluctuating around an average level of 53.1 and decreasing only by 7 points between 1960 and 2018 (from 58.1 to 50.4) (GCIP Database). The position of the country compared to other Latin America countries in term of poverty level is similar: in 2018 Colombia was the fifth most poor country in the region, having a poverty gap ($1.90 poverty line) equal to 1.7 (the extreme for the region were 0 in Uruguay and 7.9 in Haiti); 4.1% of the population lives in extreme poverty (below the $1.90 poverty line) and 27.8% below the $5.50 poverty line. Nevertheless, the magnitude of the improvements in terms of poverty has been higher: in 1960 more than half of the population was living with less than 4$ a day while today the percentage is below 40%; similarly, the level of extreme poverty (below the $1.25 poverty line) went from 8.1% in 1960 to 3,5% in 2015 (GCIP Database)[9].

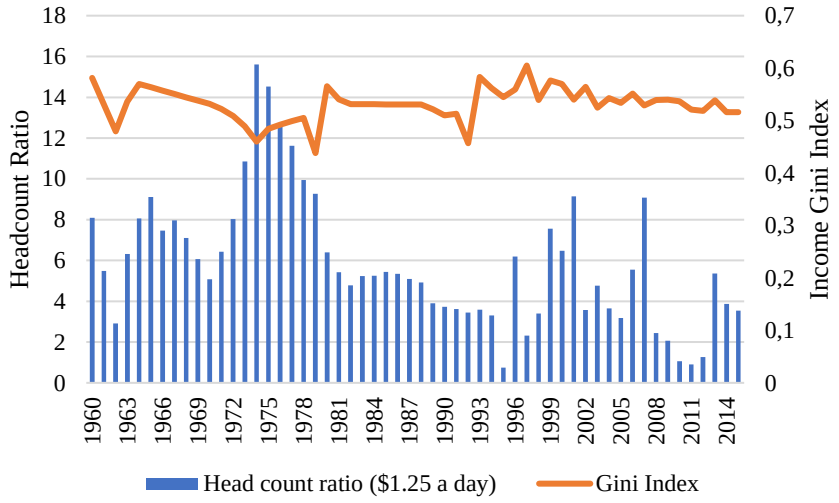

Figure 1: Income Gini Index[1] and headcount ratio[2], Colombia (1960-2014)
Source: Global Consumption and Income Project (GCIP). Notes. 1: Gini Index on the right axes. 2: HCR on the left axes, compute with $1.25 poverty line.

During the same period, Colombia experienced significant economic growth with an average GDP per capita annual growth of 3.2% between 1961 and 2018 (the only year with a negative growth was 1999) and being today in the middle of the income distribution in the region with a GDP per capita of 7,692$ (constant 2010 US$) (World Bank Database). The discrepancy between the enhancements in GDP per capita for one side and the flat trend in inequality or slight improvements in the poverty level for the other could be related with the level of inequality of opportunity of the country: the unequal distribution of possibilities among the individuals would mean the exclusion of part of the population from the economic growth gains and could explain why the country (as well as the entire Latin American region) has a persistent level of inequality which is higher than the level that countries with similar GDP per capita show in the rest of the world (Prados de la Escosura, 2007).

[9] The absolute values slightly differ depending on the source and on the estimation method used, but the trends are constant.

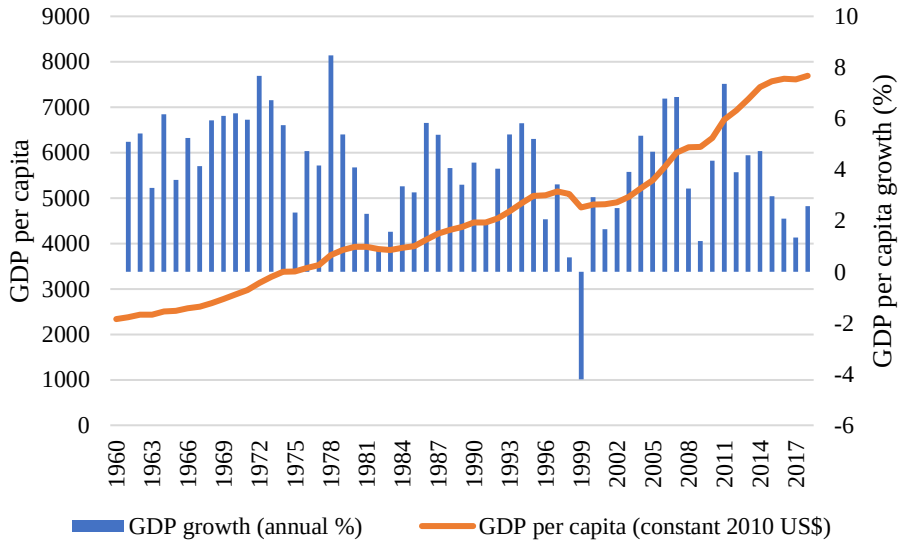

Figure 2: Annual GDP per capita growth (%)[1] and GDP per capita[2], Colombia (1960-2017)
Source: World Bank Database. Notes. 1: GDP per capita growth on the right axes. 2: GDP per capita in constant 2010 US$ on the left axes.

Using the Human Opportunities Index (HOI)[10] developed by Paes de Barros et al. (2008), Colombia ranked 12th among the nineteen Latin American countries analysed, with an index equal to 78%, meaning that 78% of the opportunities available were distributed following the principle of equality of opportunities (the same index goes from 90% for Chile to 51% for Guatemala). Moreover, Colombia is the country with the highest annual change in the index since 1995, when the HOI was 68%. In order to compute the index, the authors take into account different circumstances that are proved to affect children's opportunities, namely place of residence (urban or rural), parents' education, household income per capita, household size, gender, and presence of the parents in the household. They find that in Colombia the most important circumstances for the educational opportunities are the place of residence and parents' education. More in general, they find that among the most economically disadvantaged group almost the totality of the individuals has parents with no education or primary education only, 65% comes from peripheral departments and 33% are from an ethnic minority; on the other side parents with at least secondary education are strictly related to higher incomes of the individuals (Paes de Barros, et al., 2008).

Both the National Plans of Development (*Plan Nacional de Desarrollo*) redacted by the Government of Colombia in 2014 and 2019 put as one of the main objectives the equality of opportunities among different groups and they recognize that the regional differences

[10] The Human Opportunities Index is a synthetic index that measures the level of equality of opportunities in the basic goods and services, i.e. goods and services that are considered fundamental for the human development (in education it is finishing primary school on time and school enrolment for children between 10 and 14 years old). The two components that affect the index are the general availability of the basic opportunities as well as whether the opportunities are distributed equally among the population (i.e. not depending on external circumstances). The authors compute the index separately for education and the household's services plus an average index between the two components, I here refer uniquely to the education index.

and the disadvantages of some ethnical groups are among the primary problems of the country. Although the majority of the population do not recognize in any ethnical minority, the 14% that belongs to indigenous groups or that recognize as Afro-Colombian suffers from lower opportunities related to several factors such as disadvantaged location, discrimination, or poor family background. The disadvantages experienced by the ethnic minorities are related to different dimensions of welfare, among which education is included: the percentage of families that suffer from illiteracy and school absenteeism is double within the ethnic groups respect to families that are not in any ethnic group. For these reasons, different public policies have been developed with a specific target on the indigenous population, in order to improve nutrition level, welfare, and educational attainment (Government of Colombia, 2014; Government of Colombia, 2019).

The disadvantages of the ethnic minorities are strictly related to the second issue cited above because they are usually located in peripheral rural zones. The administrative division in Colombia is characterized by zones with specific characteristics in terms of geography and socio-economic features. For instance, most urban centers are in the highlands of the Andes, with the three main cities accounting for 41% of the total population and 80% of the economic activities. On the other hand, the peripheral regions and the ones close to the sea are the poorest. There is no consensus about whether Colombia has experienced regional convergence or polarization in the last decades depending on the method used or on the period analysed, but all authors highlight the regional differences in terms of GDP per capita (figure 3) and other socioeconomic indexes (e.g. the illiteracy rate goes from a maximum of 22% to a minimum of 2% depending on the region) (Royuela & Garcìa, 2015; Corpoeducación, 2006). The departments have a certain degree of autonomy that permits to answer more specifically to local needs but that limits the possibilities to the local budget, potentially creating more space for further regional divergence.

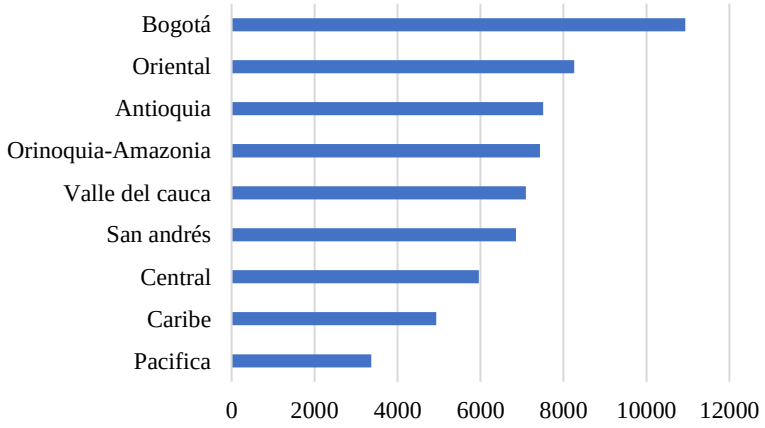

Figure 3: GDP per capita per region, Colombia (2018) (in US$).
Source: Departamento Administrativo Nacional de Estadìstica (DANE). Note: the division in regions of the graph does not refer to an actual administrative division of the country, but uses the same classification used in the ENCV 2018.

2.3.1 Educational system

The educational system in Colombia consists of four levels in which the education is provided by both public and private institutions regulated by the Ministry of Education (*Ministerio de Educación*): preschool (*educación preescolar*) of which only last year is mandatory, basic education (*educación básica*) for children between 6 and 14 years old and divided into primary and secondary education, middle school (*educación media*) for 15 and 16 years old, and finally tertiary education (*educación superior*) which length depends on the type of studies (university, technical or technic) from one to five years plus eventual postgraduate education such as master or Ph.D. In brief, from primary to undergraduate, education in Colombia consists of eighteen years for individuals from 6 to 21 years old (unless of repeated academic years); the education is compulsory for individuals between 5 and 15 years old, thus individuals are supposed to be enrolled from the last year of preschool until the first year of middle school.

The four school cycles are provided by public institutions subsided by public expenditure plus individual fees for the secondary basic education and the tertiary education (with reductions depending on the family income) (Ministerio de la Educación de Colombia, 2020). Nevertheless, a significant number of families with enough economical resources decides to enrol children in private schools and universities because of the gap in quality and environment between public and private institutions; indeed, 19% of the students are enrolled in private schools (the percentage is significantly higher than OECD average of 12%) and the proportion grows to 47% if only tertiary education is considered. This difference in preferences and opportunities between wealthy and poor families is one of the determinants of the high level of school segregation that characterises the Colombian educational system (OECD, 2016). Although the main rules and goals are decided centrally by the Ministry of Education, since the General Law of Education (*Ley General de Educación*) in 1994, education in Colombia is based on the idea of decentralization and large autonomy is given to the single Certified Local Authorities (*Entidades Territoriales Certificadas*) and to each scholar institutions. The decentralization has created more opportunities for innovation and it has permitted to focus more on local needs but it is an aspect that aggravates the geographical differences since the poorest departments and municipalities do not have enough economic resources, tools, and knowledge to guarantee an effective and high-quality education (OECD, 2016).

The last decades represented a period of important commitment by the Colombian government that put universal education among its primary goals. The Constitution of 1991 and the General Law of Education of 1994 state that education is a fundamental right for every Colombian citizen and a priority for the government, in order to achieve the development of both the individual and the society (OECD, 2016). Table 1 shows the level of enrolment for each different school cycle. Those numbers represent an important improvement with respect to the past, in particular for secondary and middle education; the enhancements were possible thanks to different tools the country used such as infrastructures construction, cash transfers, school credits, the increment of teachers' quality, and policies for disadvantaged groups (i.e. indigenous, rural inhabitants, children from poor families) (Government of Colombia, 2019); the improvements across all the

period considered in the analysis can be appreciated looking at the trend of the average number of years of education by year of birth in the sample (figure 4). Nevertheless, despite the improvements of the last years and the declarations stated in the Constitution, education in Colombia is still far from being "universal" (i.e. more than 90% of the individuals enrolled) and the Government still put it as one of the main objectives for the future of the country (Government of Colombia, 2019).

Table 1: Net enrolment rate for educational level in Colombia, 2017 and 2000

Level[1]	Enrolment rate (2017)	Enrolment rate (2000)
Pre-school[2]	55%	36%
Primary	83%	87%[3]
Secondary	72%	59%
Middle	43%	30%
Total	85%	87%[3]

Source: Plan Nacional de Desarrollo 2018-2022, Government of Colombia (2019). Notes. 1: Enrolment rate for tertiary education is excluded from the table because only gross rates are available; the percentage for 2017 and 2000 are respectively 56.43% and 24.49%. 2: The percentages for the preschool refer uniquely to the mandatory year. 3: Data available only for 2003.

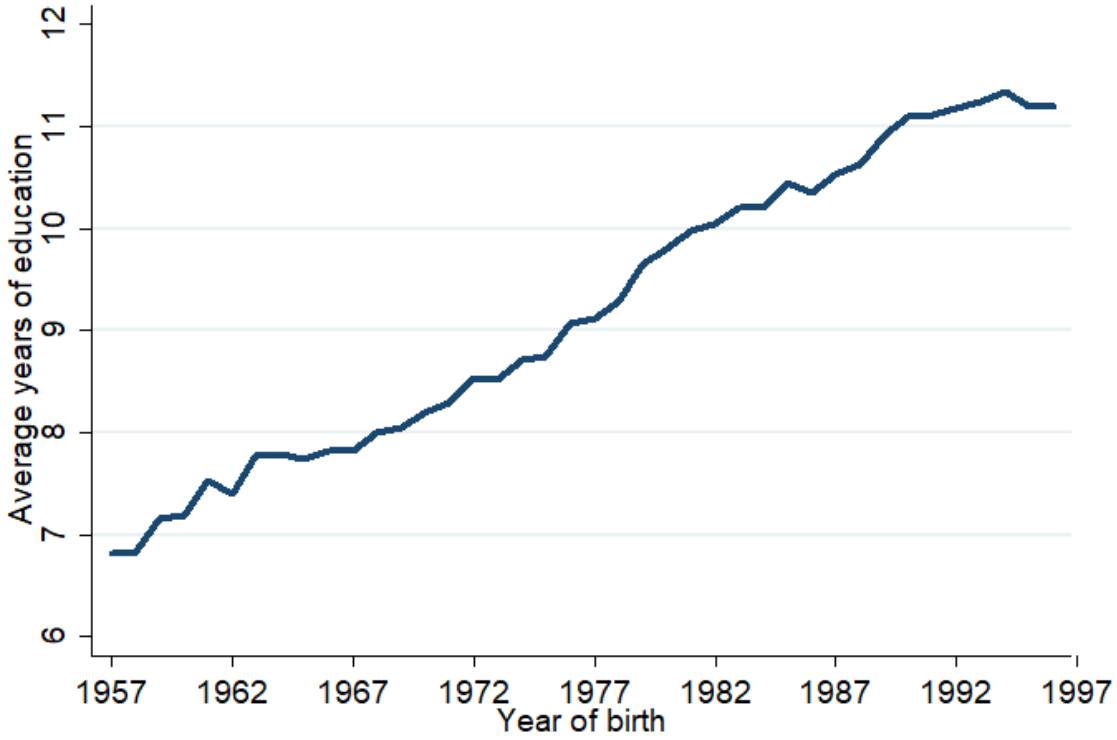

Figure 4: Average years of education per year of birth
Source: ENCV 2018 (DANE)

Similarly, the last decades have seen an improvement in terms of quality of education: in 2015 Colombia was one of the four countries with greatest improvements in the results of the PISA standardized tests; nevertheless, its level is still lower than the OECD average (Government of Colombia, 2019).

Besides the aforementioned problems in terms of access and teaching quality, the country needs to face other issues, which are strictly related to inequality of opportunity. Indeed, one of the main characteristics of the educational system is the high inequality and heterogeneity and the schooling segregation that arises from it. In spite of the public efforts of the recent years and of the several public policies addressed to the disadvantaged individuals, educational achievements still highly depend on personal characteristics that are not in control of the individual (OECD, 2016; Government of Colombia, 2019). For instance, rural regions have an enrolment gap of 9% relative to the cities and the gap is even wider for individuals coming from ethnic minorities (Government of Colombia, 2019). Moreover, the socioeconomic status of the parents is an important determinant of enrolment: children from non-educated parents have fewer incentives to get educated and are less aware of the possibilities and - even tough since 2012 public school is free - they incur high indirect and opportunity costs (OECD, 2016). The unequal geographical distribution of poverty in addition to the high level of private educational expenditure generates a pronounced difference between schools attended by wealthy or poor individuals. The consequences of the segregation are lower performance, higher social conflicts and a reinforcement of the inequalities that already characterize the country: poorest individuals from rural environment receive an education of lower quality, grow up in more violent and less stimulating environments and have fewer choices in terms of educational options (in particular for higher grades) (OECD, 2016).

In its National Plan for Development (2019), the Colombian Government recognizes the aforesaid improvements but also the issues that the educational system still faces, putting education as a fundamental pillar for the future of the country. A better educational system is needed in order to achieve economic growth and a more equal society, where each individual could take advantage of the same opportunities, no matter his socioeconomic status, ethnicity, gender, or origin. In order to achieve such goal, the Government intends to increase educational public expenditure and to realize policies aimed at increasing quality and access (with particular focus on middle school), involving children from rural and indigenous environments, and removing the economic barriers for individuals from low-socioeconomic families (e.g. with credit, cash transfer, lower fees) (Government of Colombia, 2019).

3.Data

3.1 Source material

The data used for the analysis are retrieved from the 2018 *Encuesta Nacional de Calidad de Vida* (i.e. National Survey on life quality, henceforth ENCV) collected in Colombia by the National Administrative Department of Statistics (i.e. *Departamento Adiminastrativo Nacional de Estadistica*, *DANE*) The ENCV originated from a project started in 1991 as an answer to the necessity to recollect and analyse data about the welfare of the population with indicators that would go beyond the mere income or consumption, comprehending different dimensions of welfare (e.g. possession of goods, health, education). In 1991, the survey started in the Capital District of Bogotá and through the years it was expanded adding more information and reaching national coverage. Since 2010 the ENCV has been repeated each year containing eleven permanent questionnaires: data of the dwelling, household's services, characteristics and composition of the household, health, comprehensive care of children under the age of five, education, workforce, information and communication technology, child labor, possession and finance of the dwelling, living conditions of the household and possession of goods.

The survey of 2018 collected information about more than 80.000 households and 280.000 individuals, and it is representative of the entire Colombian population, divided into the 33 departments. The departments are further grouped into nine regions, which do not correspond at any actual administrative division of the country but put together political and geographical divisions in order to create homogeneous territories: Antioquia, Bogotá, San Andrés, Valle del Cauca, Caribe, Oriental, Central, Pacifica, and Orinoquía – Amazonía (Departamento Administrativo Nacional de Estadistica, 2019).

Variables about education are available in the ENCV for all the individuals above the age of 5, nevertheless, in this study I limit the sample to individuals between 22 and 61 years old at the time of the interview, meaning to people born between 1959 and 1996. The lower bound restricts the sample to individuals that are supposed to have finished their academic path until the undergraduate level (although the sample includes individuals that are still enrolled), while the upper-bound avoids possible selection bias given by the fact that education is correlated with health and longevity, with less-educated individuals more likely to be unrepresented in the sample for the older cohorts. Afterward, the sample is further divided into four different cohorts, each of which contains individuals born in ten different years (i.e. 1957-66, 1967-76, 1977-86, and 1987-96).

The inclusion of individuals still enrolled and the possibility that people enrol at the graduate or post-graduate level after being 22 years old could generate a downward bias in the estimations (Hertz, et al., 2007). Nevertheless, I expect the bias to not significantly affects the results since less of 5% of the individuals in the sample are still enrolled and 30% of them are over 30, which means they are likely to already be in the job market.

Furthermore, it has to be noticed that the arbitrariness in the division of the different cohorts could bring to different results depending on how the dummies are defined; nevertheless, in chapter 6 I check that the division into cohorts does not affect the main results since the long-term trend in educational mobility is still captured regardless the division.

For each individual, many variables available in the ENCV give information on the level of education and literacy, on grants or credits, and on eventual reasons behind school drop-out; in particular, the educational level is available both as the number of years of education or the highest level obtained (i.e. pre-school, primary, secondary, middle, or tertiary education). The main dependent variable that will be used in the analysis – individual education (*educ*) – represents academic achievement of each individual in terms of years of completed education (starting from the mandatory year of preschool); for individuals still enrolled the variable takes the value of the year previous the one in which they were enrolled at the moment of the survey.

The dataset provides information about the educational level of both parents, which can be used to assess the intergenerational transmission of education. However, differently from individual education, these variables indicate only the highest grade attended, reducing the accuracy of the analysis. Since the division in school cycles do not coincide in different countries, in order to increase the comparability with other studies I construct a variable indicating the number of years of each parents' education. While the construction of the variable does not create problems when the individual dropped out at the end of a cycle, when the educational level obtained is indicated as "*part of*" a cycle, I assume the number of years to be equal to the middle year of such cycle[11].

The previous literature is inconsistent about the exact variable to use to measure the parental educational background, varying between mother's or father's educations, or a combination of the two such as the highest education between the two or an average of both. I here use a further method developed by Lutbosky and Wittenberg (2006) (henceforth LW) in order to obtain the most efficient estimator when the variable of interest (here parental education) is a combination of multiple proxies (here father's and mother's educations) reducing the attenuation bias. All the different variables cited above can be seen as the weighted average of both proxies; nevertheless, while common measures arbitrarily set the weights, the LW procedure constructs such weights depending on the correlation of each proxy with the variable of interest. The exact computation of the variable is further explained in the method's chapter.

[11] The educational level (stated by the original variable) and the corresponding number of years of education that I assumed are the following: "no education" 0, "some primary education" 3, "primary education" 5, "some secondary education" 7.5, "secondary education" 9, "middle school" 11, "some tertiary education" 14, "tertiary education" 19. In order to check the reliability of my estimations, I used the same method to construct a further variable indicating the individual years of education; the new variable has a correlation of 0.94 with the original variable that is available in the dataset.

Besides the main variable of interest that is parental education, the specifications include other control variables in order to avoid possible omitted variable biases given by the correlation between parental education and other variables that affects individual education in the error term. Therefore, among the controls, there should be all the variables that are correlated with inequality of opportunity and educational attainment of both parents and individual, such as ethnic group, family income, quality and availability of school, household size. Nevertheless, many of those variables are absent in the dataset or they only concern the moment in which the survey was collected and not when human capital decisions were made during childhood. Thus, in order to decrease a possible omitted variable bias, I include regional fixed effects, that proxy for regional characteristics that influence the decisions in human capital, such as school quality and availability, average poverty and inequality levels, access to urban labor markets, etc. The regional characteristics are expected to cover an important part of the unobservable variables, since the location is one of the main determinants of inequality of opportunity in Latin America and regional differences are among the main issues of the country (Paes de Barros, et al., 2008; Ferreira & Gignoux, 2011; OECD, 2016; Government of Colombia, 2014). Moreover, I include gender and ethnic origins since they are important individual determinants of school enrolment (Paes de Barros, et al., 2008); in particular, ethnic differences in enrolment are still an issue in Colombia with people coming from minorities such as indigenous groups less likely to be enrolled in the higher grades in spite of the recent public policies addressing disadvantaged groups (OECD, 2016).

3.2 Descriptive statistics

Table 2 displays the average values of the variables used in the analysis for the different cohorts considered. Clearly, the number of observations is decreasing when considering older cohorts, as a consequence of population growth (around 2% on average during the period) and of higher mortality rate across older cohorts (World Bank Database). Moreover, it has to be noticed that 13,325 observations (9.31% of the total sample) had to be deleted from the sample because no information was available about nor mother's or father's education. This is likely to create a selection bias since missing information about the parents is more likely in disadvantaged families; a robustness check in chapter 6 proves that such bias does not affect the final results. From 80 to 84% of the sample is white, while the remaining part is mostly from indigenous groups (8.6%) or African descendants (9%).

As expected, educational attainment has increased by almost four years across cohorts from 7.68 to 11.15 years of education obtained, which correspond respectively with half of the secondary school and half of the middle school. Moreover, each cohort has average educational attainment that is around the double than the one of their parents: parents' education goes from 3 years to 6 across cohorts, meaning that parents' have on average only primary education. Moreover, it has to be noticed that the standard deviations of individuals' education and parents' educations follow the opposite trend, since the former is decreasing while the latter is increasing across the cohort. The different trends' signs

can be explained using the educational Kuznets curve theory: similarly to the classical Kuznets curve, the relationship between educational attainment and educational dispersion is an inverse U-shape. When the educational attainment is minimum, individuals have a similar number of years of education close to the minimum; with schooling expansion, the average number of years increases as well as the dispersion because educational opportunities are initially available only for part of the population. After a certain number of years – that it has been computed close to 6 from international empirical studies but that depends on each society – there is a turning point after which school expansion goes along with a decrease in educational inequality until reaching its minimum, when most of the population is highly educated (Londoño, 1996; Shukla & Mishra, 2019; Ram, 1990).

Table 2: Summary statistics of key variables of the sample per cohort.

	Cohort 1957-66	Cohort 1967-76	Cohort 1977-86	Cohort 1987-96
Number of individuals	25,579	30,098	34,970	39,134
Years of education	7.68 (5.24)	8.60 (5.03)	10.08 (4.74)	11.15 (3.94)
Father's education	3.14 (3.51)	3.55 (3.71)	4.56 (4.36)	5.86 (4.83)
Mother's education	3.06 (3.14)	3.53 (3.47)	4.71 (4.26)	6.25 (4.87)
Male	0.48 (0.50)	0.47 (0.50)	0.47 (0.50)	0.47 (0.50)
White	0.84 (0.37)	0.83 (0.38)	0.82 (0.39)	0.80 (0.40)

Source: ENCV 2018, Colombia (DANE)
Note: Values show average and standard deviation (in parentheses).

Table 3 and its graphical representation in figure 5 show more in detail the different trends in education across cohorts, indicating for each one the total number of individuals and the percentage of the population that obtained a certain level of education. The most evident change across cohorts is the movement of the modal value from the primary education to middle school: among individuals between 52 and 61 years old more than half obtained only the primary education or less, while in the youngest cohort half of the individuals obtained at least the middle school. Nevertheless, even for the youngest cohort with the mandatory school until the age of fifteen, more than 30% of the individuals dropped out before finishing the middle school; indeed, the same Colombian Government today sees the low enrolment in middle school as one of the main hurdles to overcome in order to increase educational attainment in the country (Government of Colombia, 2019; OECD, 2016). Even if to a lesser degree, the achievement of the tertiary education

(comprehensive of technological, technical and university institutes) has increased too; moreover, it has to be noticed that the values for the tertiary education in the most recent cohort are downward biased since some individuals are still enrolled.

Table 3: Individuals by the highest completed level of education.

	Cohort 1957-66	Cohort 1967-76	Cohort 1977-86	Cohort 1987-96
None	2,755 (10.77%)	2,089 (6.94%)	1,383 (3.95%)	892 (2.28%)
Preschool	47 (0.18%)	56 (0.19%)	65 (0.19%)	75 (0.19%)
Primary	11,392 (44.54%)	11,932 (39.66%)	9,522 (27.23%)	5,667 (14.48%)
Secondary	3,280 (12.82%)	3,980 (13.23%)	4,850 (13.87%)	6,273 (16.03%)
Middle	4,333 (16.94%)	6,922 (23.00%)	11.197 (32.02%)	17.947 (45.86%)
Undergraduate	3,025 (11.82%)	4,380 (14.56%)	7,129 (20.43%)	7.936 (20,27%)
Postgraduate	747 (2.92%)	730 (2.43%)	814 (2.33%)	344 (0.88%)
Number of individuals	25,579 (100%)	30,098 (100%)	34,970 (100%)	39,134 (100%)

Source: ENCV 2018, Colombia (DANE)

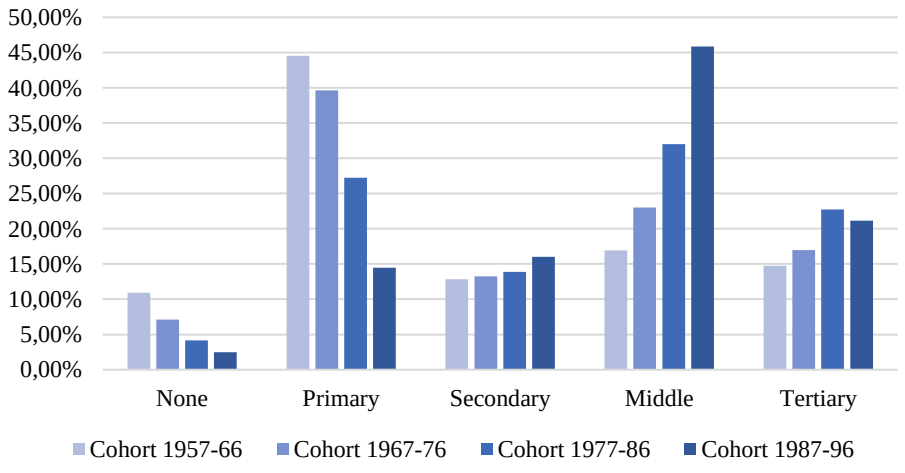

Figure 5: Individuals by highest completed level of education (%), Colombia
Source: ENCV 2018, Colombia (DANE)

4.Methods

The chapter starts with the methodology used to compute average parental education, retrieved from Lutbosky and Wittenberg (2006). Then, the indices used for the analysis are presented, with their respective formula as well as meaning in terms of intergenerational mobility. Finally, the chapter concludes with possible limitations of the analysis that could arise from data quality or the author's choices in terms of variables employed.

4.1 Parental education

As stated in the previous chapter, the independent variable of interest of my analysis – parental education – is computed following the method developed by Lutbosky and Wittenberg (2006) as the weighted average of mother's and father's education (in years of education approved):

$$p\_educ_i = w_i^f * f\_educ_i + w_i^m * m\_educ_i \quad (1)$$

where *p_educ* is parental education, *f_educ* father's education, *m_educ* mother's education (both in years of education). w^f and w^m are the respectively the weights for father's and mother's educations, which are computed as follow:

$$w^j = \frac{\phi^j}{\phi^g + \phi^m \rho^m} \text{ with j = f, m} \quad (2)$$

where Φ are the estimated coefficients of the linear regression of children's education on the two proxies of parental education $educ_i = \phi^f f\_educ_i + \phi^m meduc_i + v_i$ and ρ^m is defined as the ratio of covariances $\rho^m = \frac{cov(educ_i, m\_educ_i)}{cov(educ_i, p\_educ_i)}$. Although the LW variable is usually highly correlated with the other possible indices of parental education, this method allows obtaining the most efficient estimation of the effects of parental education on individual education (Lutbotsky & Wittenberg, 2006; Neidhofer, et al., 2018). Since the LW method is not widely used in the previous literature, I repeat the analysis with the other common proxies for parental education as a robustness check.

4.2 Trends in intergenerational mobility

The empirical analysis aims to explore how the magnitude of the effects of parental education on individual educational achievement has changed across generations in Colombia. Intergenerational mobility can be measured as absolute mobility (structural mobility), which compares the level of education of children and parents and which is related to economic growth and school expansion. Otherwise, other indices of mobility give information on relative mobility (exchange mobility or social fluidity), that focuses on changes within distribution and compares the level of education of disadvantaged individuals to the level of more advantaged ones (Neidhofer, et al., 2018; Biagi & Stuhler,

2018). The choice between a type of index or another depends on the research question because they give information about different phenomena. For instance, given the improvements in term of average school attainments in the country, it is likely that many individuals from low-educated parents reached a higher level of education than their parents (thus improving absolute mobility) but still remaining in the low part of the education distribution (thus without changes in terms of relative mobility). Given the focus of this paper on the concept of inequality of opportunity, the notion of relative mobility is the one that better answers to the research question; nevertheless, indices of both types are going to be computed, in order to give the most exhaustive representation of educational mobility trends in Colombia.

The most common way to measure intergenerational mobility in the literature is computing the slope coefficient on a regression between parents' and children's education (Neidhofer, et al., 2018; Binder & Woodruff, 2002).

$$educ_i = \alpha + \beta_1 p\_educ_i + \beta_{2j} cohort_{ji} + \beta_{3j} cohort_{ji} * p\_educ_i + \beta_4 X + \lambda_i + \varepsilon_i \quad (3)$$

Where *educ* is the level of education of the individual in numbers of years completed, *p_educ* the parents' level of education as weighted averages of both parents' education (following the LW procedure), $cohort_j$ is a vector of dummies indicating to which cohort *j* the individual belongs to, *X* a vector of controls that influence the level of education and of opportunities (i.e. sex and ethnic group), λ is a regional fixed effect, and ε is the robust standard error. Each term is specified for each individual *i*. In addition to a basic linear regression, specification (3) adds interactions between parental education and each age group dummy, which coefficient β_{3j} indicates how the magnitude of the intergenerational transmission of education changed over time. Given the increase of public expenditure and policies in the last decades, I expect β_{3j} to be lower for younger generations, meaning that parental education had decreasing importance on human capital choices. Moreover, model (3) includes regional fixed effects, which allow accounting for unobservable characteristics common to each region, such as public expenditure, school quality, and availability, poverty level, access to urban labor markets (and thus the level of returns to education); in this way, the regional fixed effects avoid having spurious effects originated by unobservable common determinants of parental and children's education.

A second option to measure intergenerational mobility is to take into account the differences in the variance of the level of education between parents and children; in order to do so, the slope coefficient is multiplied by the ratio between the standard deviations σ of parents (*p*) and children (*c*) obtaining the intergenerational correlation coefficient[12]:

$$r_j = (\beta_1 + \beta_{3j}) \frac{\sigma_j^P}{\sigma_j^c} \quad (4)$$

Both indices capture the two facets of educational mobility, increasing for both absolute and relative mobility, but *r* improves β using changes in the distribution of the output of

[12] The index *r* is equal to the Pearson's correlation coefficient and to the square root of R^2 if no control variables are included in the regression

interest; therefore, the focus on one or the other index depends on how important is the dimension of intergenerational transmission of inequality, since the computation of (4) factors out this part of the information. There is no agreement in the previous literature on which one is the better index, and the difference between β and r becomes particularly significant when the country experiences large improvements in educational attainments, thus to report both is the most practical solution (Neidhofer, et al., 2018; Black & Devereux, 2010).

While the previous indices capture both absolute and exchange mobility, the Spearman's rank correlation coefficient (5) captures uniquely the relative mobility, since it is computed using information about the position of each individual and his parents in the distribution of educational achievements of the cohort in which they belong.

$$\rho_j = \frac{cov(rank_{p_{educ_i}}, rank_{educ_i})}{\sigma_{p_{educ_i}}, \sigma_{educ_i}} \quad (5)$$

Where *rank* is the position of each individual in his own distribution of educational achievements, *cov* is the covariance between the two positions, and σ is the standard deviation.

The previous measures do not give any information about the distribution of mobility, assuming that the degree of persistence does not change depending on the position of families in the education distribution. Nevertheless, persistence is likely to be stronger at the tails, generating different conclusions in terms of both individual and society's welfare. In order to capture the differences in mobility for individuals coming from different family backgrounds, it is useful to compute the probability to reach a certain level of education given the circumstances that the individual faces (i.e. different parental education). A transition matrix gives such information for each possible combination between children's and parents' education, indicating the probability that an individual reaches the educational level *e* given the parents' education *p* (Richey & Rosburg, 2017):

$$p_{ep} = \frac{Pr(educ{=}e|p\_educ{=}p)}{Pr(p\_educ{=}p)} \quad (6)$$

where $\sum_{p=1}^{m} p_{ep} = 1$ with m equal the maximum level of education. Although the transition matrix is the most comprehensive tool giving information for each possible combination between parents' and individuals' education, it does not permit overall comparisons across countries or time. With this purpose, starting from the transition matrix, it is possible to compute the Immobility Ratio (IR), which gives information about the clustering around the diagonal of the matrix, i.e. the percentage of the cases in which the level of education does not change from parents' to children; when individuals are divided into five groups (usually quintiles), the IR is equal to 100 with complete immobility and equal to 20 with complete mobility (meaning that each individual has equal probability to be in any position of the distribution, regardless his parents' position). Otherwise, it is possible to focus uniquely on a specific part of the distribution, computing

the bottom upward mobility (*BUM*) or the upper-class persistence (*UCP*), that give information respectively about the lower and the higher tails of the distribution (Neidhofer, et al., 2018):

$$BUM_j = Pr\left(educ_{ij} \geq s \mid p\_educ_{ij} < s\right) \quad (7)$$

$$UCP_j = Pr\left(educ_{ij} \geq s \mid p\_educ_{ij} \geq s\right) \quad (8)$$

where *s* is a certain educational level of interest. The upward mobility indicates the probability to be highly educated for an individual whose parents were poorly educated, while upper-class persistence indicates the probability to be highly educated for an individual whose parents were highly educated as well. I define *s* equal to middle education because it is considered by the OECD as the minimum requirement in order to participate in the socio-economic life of the country, with consequences in the long-term on variables such as income, health, level of violence and employment (OECD, 2016). Thus, "*highly educated*" is defined for a person with middle or tertiary education, whereas "*poorly educated*" is defined for a person with secondary school or lower as the highest completed level of education.

Finally, two more indices give information about the absolute magnitude of intergenerational mobility.

$$M1_j = \frac{1}{N_j}\sum_{i=1}^{N_j}\left|educ_{ij} - p\_educ_{ij}\right| \quad (9)$$

$$M2_j = \frac{1}{N_j}\sum_{i=1}^{N_j}\left(educ_{ij} - p\_educ_{ij}\right) \quad (10)$$

M1 (absolute mobility) and M2 (directional mobility) measure the average difference between children's and parental education, respectively without and with taking into account the direction of such change. The two indices are particularly useful together: M1 summarises both upward and downward movements treating them in the same way, while M2 gives information about educational expansion across generations, thus their difference is a computation of the level of downward mobility (Neidhofer, et al., 2018).

4.3 Limitations

Every empirical analysis that aims to explore the topics of inequality of opportunity and intergenerational mobility needs to face severe limitations in terms of data quality and availability. Indeed, the type of analysis requires information about different generations or about variables that are difficulty observable (Roemer & Trannoy, 2015; Ferreira, et al., 2011). In the computation of the regression coefficient, the omission of important factors that affect individuals' education could generate omitted variable bias giving a biased estimation of the coefficient. In this analysis important missing factors are mostly related to the childhood, when the decisions about investments in human capital were

taken and when the level of parental education was mostly affecting the individual; some examples are the place of residence, family income, household size, availability of school infrastructure in the region. Other limitations related to the data availability are the indication of parents' education in the level of education instead of years of education and the presence of a high percentage of missing observations that could create a selection bias; nevertheless, further robustness checks in this thesis discard the possibility for those biases to affect the results obtained.

Regarding the main variables of the analysis (i.e. parents' and children's education), the most important issue relates to the choice on how to measure them in order to better capture the benefits that education has on the individual. Besides the years of education used here, part of the literature focuses on the highest degree obtained by the individual; moreover, more recently some authors center their analysis on the quality of education rather than on the quantity, assuming that a mere increase does not always translate in actual improvements of children knowledge and abilities (Biagi & Stuhler, 2018). Additionally to not take into account the quality of education, to use the number of years as variable presumes a linear relation between parents' and children's education, assuming that each further year of education has the same effects. Therefore, the usage of this type of proxy for education could bring to overlook effects related to different levels of education, reducing the results to an average value (Hertz, et al., 2007; Biagi & Stuhler, 2018).

5.Empirical analysis

Table 4. Regression of parents' education on individual education in Colombia (1957-1996)

Independent variables	(1)	(2)	(3)	(4)
Parents' education	0.66***	0.62***	0.45***	0.44***
	(0.00)	(0.00)	(0.00)	(0.00)
Cohort 1977-86		-0.20***	-1.25***	-1.26***
		(0.03)	(0.05)	(0.05)
Cohort 1967-76		-1.02***	-2.59***	-2.63***
		(0.03)	(0.05)	(0.05)
Cohort 1957-66		-1.67***	-3.54***	-3.59***
		(0.03)	(0.05)	(0.05)
Parents' education * Cohort 1977-86			0.18***	0.17***
			(0.01)	(0.01)
Parents' education * Cohort 1967-76			0.33***	0.32***
			(0.01)	(0.01)
Parents' education * Cohort 1957-66			0.45***	0.44***
			(0.01)	(0.01)
Male				-0.65***
				(0.02)
White				0.67***
				(0.03)
Constant	6.63***	7.45***	8.44***	8.27***
	(0.02)	(0.03)	(0.03)	(0.05)
Observations	129,772	129,772	129,772	129,772
Adjusted R-squared	0.29	0.31	0.32	0.34

Source: ENCV 2018, Colombia (DANE).

Note: robust standard errors in parentheses. All the coefficients are significant at 1% confidence level.

The results of specification (3) are shown in table 4 that reports the coefficient between parents' education for the entire period considered, for different cohorts and finally for different cohorts adding controls and regional fixed effects. In all four specifications, the coefficient keeps its strong statistical and economical significance. From the first basic specification in column (1), the coefficient indicates that individual and parents' education are strongly correlated since one year more in parental education is associate to an average increase of 0.66 in individual education and parents' education alone explains 29% of the variation in individual education. Nevertheless, looking at columns (3) and (4), it emerges that the magnitude of such relation has changed across time.

doubling between the oldest and the youngest cohorts. Column (4) shows that for individuals between 52 and 61 years old at the time of the interview (i.e. cohort 1957-66)

the correlation with parents' education was almost perfect since one year more is associated with 0.88 years more, while for individuals between 22 and 31 (i.e. cohort 1987-96) the value drops to 0.44.

From table 4 it is also possible to appreciate the school expansion that the country experienced in the last decades: the constant for the youngest cohort is equal to 8.27 years of schooling equal, while the value is equal to 4.68 for the oldest cohort (when including all controls into the model – column 4): the values respectively correspond to part of the secondary and to part of the primary educations. Finally, the controls appear to be statistically significant but with low effects on the final amount of the education: male individuals have on average 0.65 years of education less than females, while Caucasian individuals experience an increase of 0.67 years respect to indigenous, black or mixed-race individuals.

Table 5 reports the coefficients between parents' and individuals' education retrieved from column (4) of table 4, plus the correlation coefficient *r*, and the Spearman's rank correlation coefficient. Given the opposite tendency of the standard deviations for parents' and individuals' education (reported in table 11 in appendix A.1), the trends obtained with the correlation coefficient and the regression coefficient are highly different: while β steadily decreases, *r* remains almost steady across cohorts fluctuating between 0.50 and 0.51. The results of the regression coefficient indicate that each additional year of parents' education corresponds to a smaller average difference in individual education for the youngest cohorts respect to the oldest ones; nevertheless, when taking into account the differences in inequality (i.e. the standard deviations), the steady correlation coefficient *r* shows that one standard deviation difference in parents' education corresponds to a steady difference of 0.50-0.51 standard deviations in individual schooling.

Table 5: Regression coefficient, correlation coefficient r, and Spearman's rank correlation

	Cohort 1957-66	Cohort 1967-76	Cohort 1977-86	Cohort 1987-96
Regression coefficient	0.88	0.76	0.61	0.44
Correlation coefficient *r*	0.51	0.50	0.51	0.50
Spearman's rank correlation	0.50	0.49	0.53	0.54

Source: ENCV 2018, Colombia (DANE).

Similarly, the Spearman's rank correlation coefficient doesn't show any downward trend, but it rather slightly increases between the oldest and the youngest cohorts. Indeed, the index goes from 0.50 to 0.54, meaning that the position in the educational distribution of each individual becomes more correlated to the parents' position for the youngest cohort.

The results are consistent with the previous estimation of the three indices for Latin America and confirm that the mobility experienced by the country has been mainly structural (Neidhofer, et al., 2018).

Table 6 reports the transition matrices for the four cohorts analyzed: each cell shows the probability for an individual with given parents' background to have a certain level of education. The division in groups for both parents and individuals is made following the school cycles: zero education (that comprehends zero or one year of school attainment in the preschool), primary school, secondary school, middle school, and tertiary education (that comprehends both undergraduate and postgraduate levels). Although the matrices are not as directly intuitive as an index, it is possible to notice higher values at the top-left and low-right extremes, that represent immobility at the tails: individuals with parents with tertiary education have elevated probability to be highly educated themselves, while individuals with non-educated parents difficulty will go beyond the primary education. Although such immobility at the tails still persists, some level of improvement across cohorts can be appreciated by the matrices, particularly for mobility at the bottom part of the distribution. For instance, the probability to have no education or primary education for an individual whose parents were not educated goes from 78.58% to 45.06%.

Table 6: Transition matrices by the level of education per cohort.

		PARENTS' EDUCATION									
		Cohort 1957-66					Cohort 1967-76				
		None	Primary	Secondary	Middle	Tertiary	None	Primary	Secondary	Middle	Tertiary
INDIVIDUAL EDUCATION	None	25.62	6.15	0.94	1.94	0.54	19.32	4.19	1.15	0.90	0.87
	Primary	52.96	47.16	11.09	6.31	1.80	51.66	42.72	8.93	3.92	1.64
	Secondary	9.78	14.84	12.23	3.40	2.16	11.13	14.93	11.45	4.22	1.53
	Middle	8.03	17.70	25.35	22.82	10.99	13.14	22.98	31.42	19.88	12.45
	Tertiary	3.61	14.14	50.40	65.53	84.50	4.75	15.17	47.05	71.08	83.52
		Cohort 1977-86					Cohort 1987-96				
		None	Primary	Secondary	Middle	Tertiary	None	Primary	Secondary	Middle	Tertiary
INDIVIDUAL EDUCATION	None	14.10	2.85	0.96	0.56	0.90	10.97	2.03	0.91	1.06	0.48
	Primary	44.66	31.84	5.62	2.51	1.05	34.09	19.40	3.30	1.47	0.40
	Secondary	15.26	16.29	9.43	4.46	1.24	21.81	20.86	11.15	4.40	1.56
	Middle	19.61	31.32	36.51	25.38	9.86	25.98	39.96	43.03	32.27	15.25
	Tertiary	6.38	17.70	47.48	67.09	86.95	7.15	17.75	41.62	60.80	82.31

Source: ENCV 2018, Colombia (DANE).

The matrices can be graphically represented as in figure 6. The immobility at the tails cited above becomes clearer in the graphic representation: more than 80% of the children of graduated parents enrolled in the tertiary education no matter the cohort they belong to. For the lower tail, the strength of the correlation decreases over time but remaining

significantly high: almost 80% of the children of non-educated parents have maximum primary education in the oldest cohort, while the same percentage decreases at 45% for the youngest one. The most remarkable trend across cohorts is the increase of people with middle school degree among the individuals which parents have reached secondary education or less, whereas lower levels of education decrease for the same group of individuals.

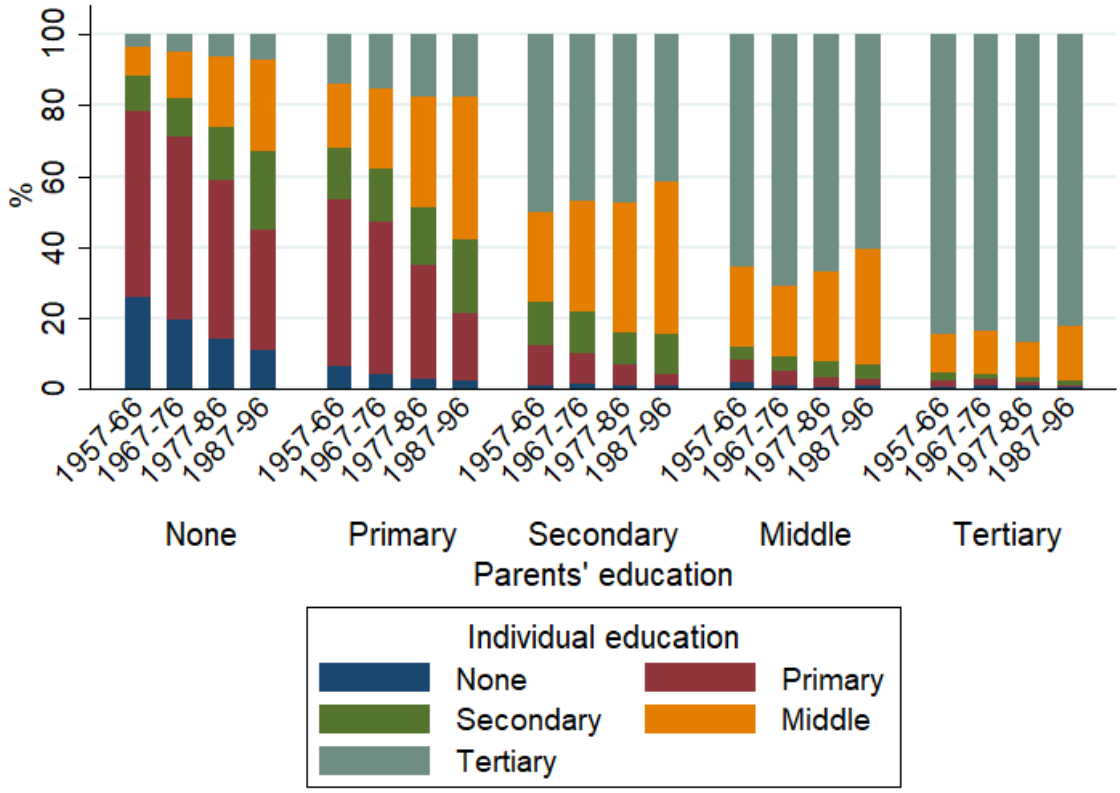

Figure 6: Graphic representation of the transition matrices by the level of education
Notes. Each group of bars represents a different level of parents' education, the different colours represent the probability to have a certain level of education, given the parental background. Source: ENCV 2018, Colombia (DANE). Own estimations.

The previous matrices and their graphic representation do not differentiate between absolute and relative mobility. Indeed, the average schooling achievements have increased across cohorts, no matters the family background, and this could explain the particular increase in the probability to reach middle education. In order to extrapolate the average trend and to obtain uniquely the relative mobility, I repeat the exercise dividing both individuals and parents in quintiles[13] depending on their position in the

[13] Since the two variables individual and parents' educations are discrete variables, it is not possible to construct proper quintiles. Figure 6 is created grouping the observations in homogenous groups each of which comprehends 20% of the population of interested; nevertheless, this requires that in some cases individuals with the same level of education are randomly assigned to adjacent quintiles. In the appendix A.2, table 13 showa the same exercise computing five groups depending on the position in the distribution and grouping individuals with the same level of education in the same group; in this way each group contains a different percentage of observations but always around the 20% (the lowest is 13% while the highest is 33%). Although not optimal, the two methods are second-best options in lacking the possibility to compute proper quintiles and both show the same trend, confirmed by the other indices here introduced.

education distribution of their cohort (in terms of year of education); the results are represented in figure 7 and table 12 (in the appendix A.2). Using the quintiles, the absence of trend across cohorts becomes clearer: indeed, the probability to be in a certain quintile given the family background does not significantly change across cohorts. On the other hand, the stronger immobility at the tails is confirmed with this different division of the observations, for both bottom and upper tail: around 40% of the individuals with parents' in the first quintile and 50% of the individuals with parents' from the top quintile do not change their position in the distribution.

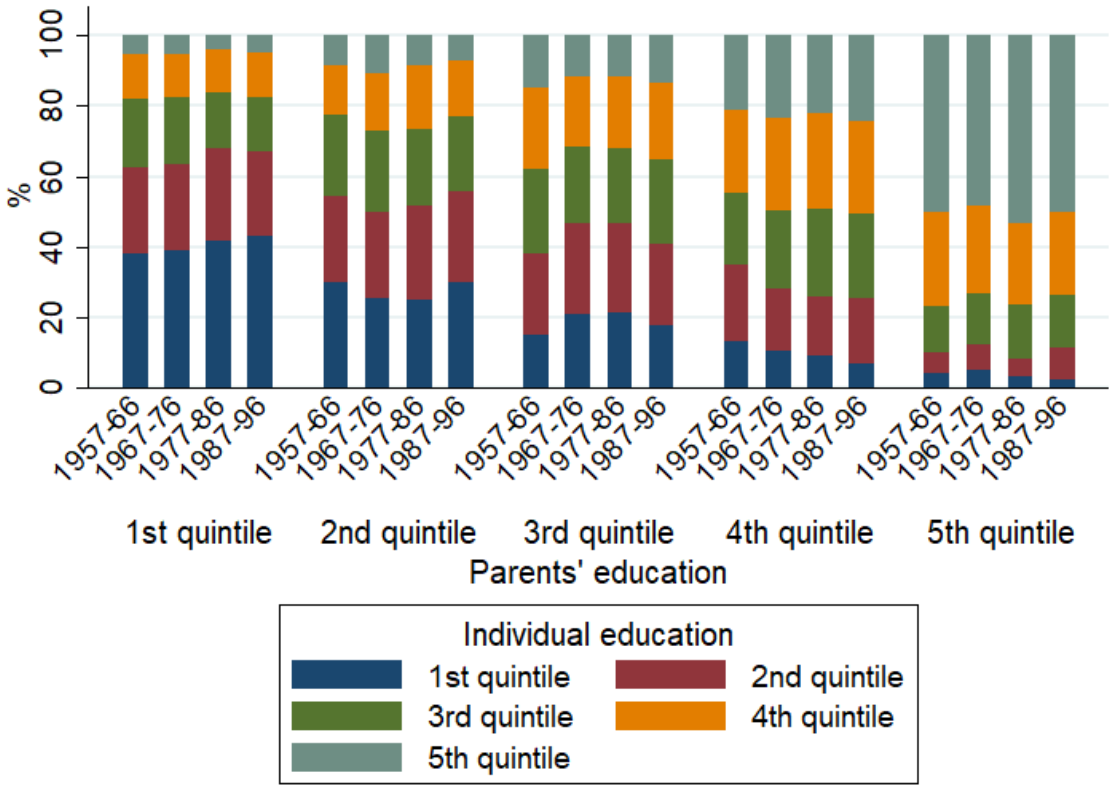

Figure 7: Graphic representation of the transition matrices by quintile

Notes. Each group of bars represents a different quintile of parents' education, the different colours represent the probability to be in a certain quintile (depending on own education), given the parental background. Source: ENCV 2018, Colombia (DANE). Own estimations.

The picture drawn by the previous figures can be synthesized using an index computed from the transition matrices: the immobility ratio (IR) is computed as the percentage of the people that lay in the diagonal of the matrix (IR_1) or in the adjacent cells (IR_2), i.e. the people that have the same or similar amount of education of their parents (or that are in the same or adjacent quintile). The index goes from 20 to 100, representing respectively perfect mobility and perfect immobility. Using the level of education obtained to divide the observations (i.e. as was made in the transition matrices in table 6), the IR shows across cohorts a percentage always higher of individuals obtaining a degree different (higher or lower) than the one obtained by the parents; in particular, IR_2 shows greater improvements going from 0.70% for the oldest cohort to 0.55% for the youngest. As already seen with the index *r* or with the matrices, the situation is different when taking

into account the average improvements that the country experienced in the last decades: when dividing the observations into quintiles, both IR_1 and IR_2 show no improvements across cohorts in terms of intergenerational mobility but rather a slight increase in immobility for the two youngest cohorts. Indeed, for the four cohorts, around 71% of the individuals are in a quintile close to the one of the parents (the same or the adjacent one) and around 33% are in the same exact quintile.

Table 7: Immobility ratios per cohort

	By degree obtained			By quintiles	
	IR_1	IR_2		IR_1	IR_2
Cohort 1957-66	0.39	0.70		0.32	0.70
Cohort 1967-76	0.35	0.64		0.32	0.70
Cohort 1977-86	0.29	0.56		0.34	0.72
Cohort 1987-96	0.24	0.55		0.34	0.72

Source: ENCV 2018, Colombia (DANE).

Note: The first two columns report the computation done dividing the individuals into five groups depending on the maximum level of education obtained; the last two columns divide the population into quintiles.

Table 8 reports the values of the bottom upward mobility and the upper-class persistence for the four generations of interest; the first index shows the probability that an individual with low-educated parents reaches a high level of education, whereas the second index represents the probability that individuals from highly-educated parents are themselves highly educated. For both indices, the threshold between a low and a high level of education is represented by the obtainment of the middle education degree. The bottom upward mobility index shows that, among the people between 52 and 61 years old, only 30% of the individuals with low-educated parents could obtain higher education, whereas the same percentage has more than doubled across generations until 63% for individuals between 22 and 31years old. On the other side, the upper-class persistence is particularly strong across all cohorts: almost the totality of individuals with highly educated parents (around 94 and 96%) was highly educated themselves; the high percentage shows that to have parents' with middle or tertiary education is an almost perfect predictor for high individual education.

The last two columns of table 8 report the same indices but using the two highest quintiles in order to identify "*high education*"; the magnitude of the indices is not comparable with the first two columns because the thresholds are different, but the different trends are insightful. Indeed there are almost no changes across cohorts for both indices, showing that the increase in the bottom upward mobility obtained comparing the levels of education represents a phenomenon of absolute mobility given by a general increase in school achievements in the country rather than higher mobility across the distribution.

Table 8: Bottom upward mobility and upper-class persistence per cohort.

	By degree obtained		By quintiles	
	Bottom upward mobility	Upper-class persistence	Bottom upward mobility	Upper-class persistence
Cohort 1957-66	0.30	0.94	0.26	0.61
Cohort 1967-76	0.38	0.95	0.26	0.62
Cohort 1977-86	0.51	0.96	0.25	0.63
Cohort 1987-96	0.63	0.96	0.25	0.62

Source: ENCV 2018, Colombia (DANE).
Note: The first two columns report the computation done dividing the individuals into five groups depending on the maximum level of education obtained; the last two columns divide the population into quintiles.

Figure 8 shows the trends of absolute (M1) and directional (M2) mobility (the values are reported in table 14 in the appendix A.3). The pattern of both curves is an inverted U-shaped in accordance with previous results and theory: the country experienced remarkable school expansion in the last decades, but rising parental education reduces the place for further increases, creating an inverted U-shaped curve despite the country keeps improving (Neidhofer, et al., 2018). The gap between the two indices does not significantly change across cohorts, meaning that the downward mobility has been stable across time around 0.4-0.36; just for the youngest cohort, the value is slightly higher, around 0.58, showing a slightly higher value of downward mobility.

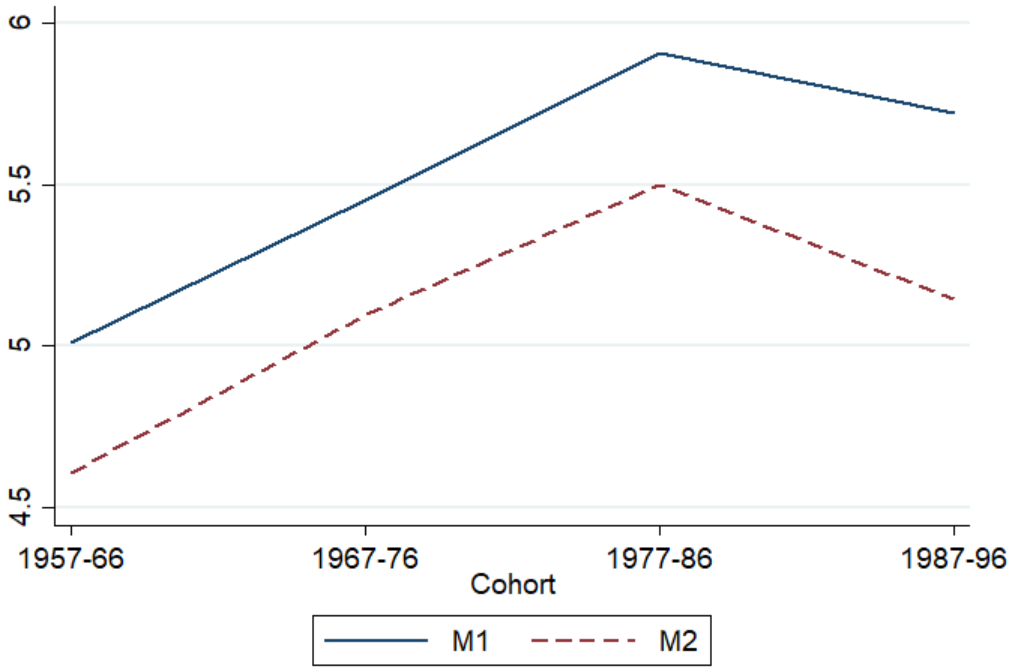

Figure 8: M1 and M2 per cohort.
Source: ENCV 2018, Colombia (DANE). Own estimations.

6.Robustness checks

6.1 Parental education proxies

There is no consensus in the previous literature about which variable better proxies the parental educational background with the choice being usually arbitrary or dictated by data limitations. As explained previously in the method chapter, the variable used in this analysis is computed as the weighted average of both parental education assigning the weights depending on the importance that each proxy has of individual education; the method minimizes the possible bias generating the most efficient proxy for the parental educational level (Lutbotsky & Wittenberg, 2006; Neidhofer, et al., 2018). This chapter shows the main results of the analysis obtained using the other four proxies common in the literature. The exercise allows to test whether the results could be biased by the proxy chosen and it permits to compare the results with the previous literature, since the LW index used here is not commonly adopted by other authors.

One of the most common variables used by the previous literature is the average between the parents' educations, which is close to the LW variable used here but assigning equal weights to the two variables; that is the case for instance of Hertz et al. (2007) whose results have been cited in the literature review. The second possible proxy relies on the assumption of dominance principle, following which it is not the total average level of education that affects the educational choices about the children but different educational decisions depend on the highest level of education in the family; following this idea, the proxy used for the parental educational background is the education of the parent that reached the highest degree (Neidhofer, et al., 2018; Heineck & Riphahn, 2007). Eventually, uniquely father's or mother's educations could be used because of data limitations or because one of the two is considered to affect more children's education in the setting analysed (it is usually the case that fathers' education is more important for sons educational choices while mothers' education has more influence on daughters' education) (OECD, 2018).

Table 9 reports the mean values and the standard deviations (into parenthesis) of the five proxies for parental education described above for each cohort used in the analysis. Both values do not change significantly across proxies besides for the variable that exploits the maximum level of education between the parents that – by definition – has a higher mean for all cohorts considered. Since father's and mother's educations are positively correlated all the five proxies – which are basically weighted averages of the two that exploit different weights – are strongly correlated among them (the minimum correlation coefficient is around 0.70 for mother's and father's educations while the maximum is 0.99 for the LW variable and the average between parental educations).

Table 9: Summary statistics of the proxies for parental educational background per cohort

	Cohort 1957-66	Cohort 1967-76	Cohort 1977-86	Cohort 1987-96
LW	3.05 (3.05)	3.50 (3.27)	4.59 (3.94)	6.03 (4.43)
Max	3.60 (3.63)	4.16 (3.92)	5.41 (4.64)	7.09 (5.16)
Average	3.06 (3.07)	3.50 (3.28)	4.59 (3.95)	6.02 (4.43)
Father	3.14 (3.51)	3.55 (3.72)	4.56 (4.36)	5.86 (4.83)
Mother	3.06 (3.14)	3.53 (3.47)	4.71 (4.26)	6.25 (4.87)

Source: ENCV 2018, Colombia (DANE).

Notes. The values show the averages and standard deviations (into parentheses). The variables presented are the following: LW variable, maximum parental level of education, average parental education, father's education, and mother's education.

Figure 8 shows the trends of the regression coefficient across cohorts for the five different proxies for parental education described above. From the figure, it becomes clear that the choice of the variable does not affect the tendency of the relationship between individual and parental educations. On the other hand, the magnitude of the coefficient slightly changes for some of the proxies. While "*average parental education*" produces an identical result (indeed, the weights for mother's and father's education are close to 0.5 for the LW proxy), the coefficients obtained with the other proxies have lower magnitude for all cohorts. Indeed, the use of "*maximum parental education*" and "*father's education*" reduces the coefficient by 0.12 on average, while for "*mother's education*" the average reduction is equal to 0.08; the differences are slightly higher for the older cohorts. The fact that the coefficients are higher when both parents' levels of education are taken into account could suggest that the general level of education in the family is important in order to generate or not a beneficial environment for the children rather than the value of education of just one of the two parents. Finally, all estimated coefficients keep high significance for all cohorts.

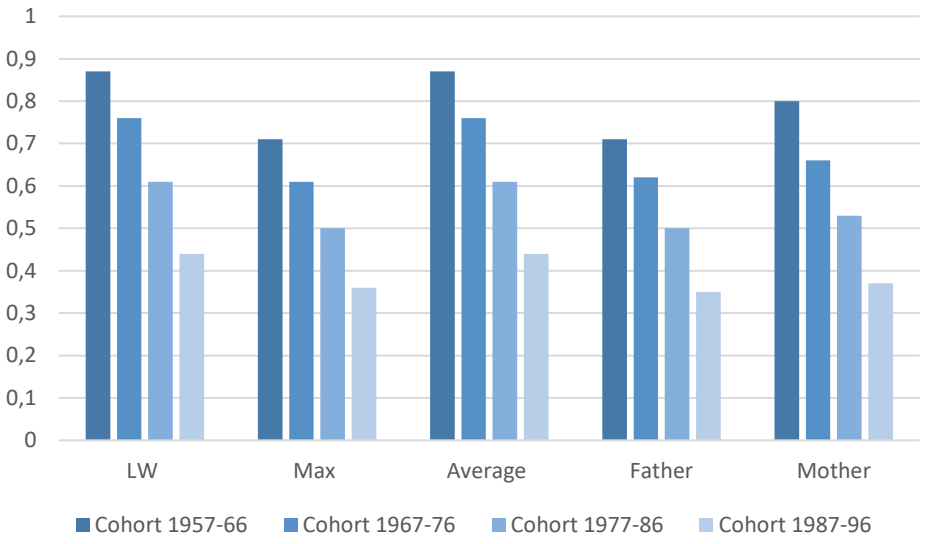

Figure 9: Regression coefficient of the five proxies for parents' education per cohort
Source: ENCV 2018, Colombia (DANE). Own estimations.

Table 10 presents a similar analysis for the Spearman's rank correlation, showing again that the results obtained do not depend on the choice of the proxies used for parental education. Indeed, as with the LW proxy, the Spearman's rank coefficient does not significantly vary across cohorts.

Table 10: Spearman's rank correlation between the five proxies for parents' education

	Cohort 1957-66	Cohort 1967-76	Cohort 1977-86	Cohort 1987-96
LW	0.50	0.49	0.53	0.54
Max	0.50	0.48	0.52	0.53
Average	0.51	0.49	0.53	0.54
Father	0.49	0.46	0.49	0.49
Mother	0.49	0.47	0.51	0.51

Source: ENCV 2018, Colombia (DANE).
Note: the values show the averages and standard deviations (into parentheses). The variables presented are the following: LW variable, maximum parental level of education, average parental education, father's education, and mother's education.

Tables reporting all the indices computed in the previous chapter for each one of the proxies are showed in the appendix B.1. Similarly to above, the trend for each one of the indices does not vary depending on the proxy used, and the magnitudes only slightly change of few hundredths. Thus, the conclusions obtained with the analysis in the previous chapter have not been affected by the choice of the variable used to represents the educational family background.

6.2 Cohorts division

Since the decision on how to divide the population of interest into different cohorts is mostly just arbitrary, I test here that such choice does not affect the final results. Figures 10 and 11 show the regression coefficient and the Spearman's rank correlation when different cohorts' subdivisions are used. The graphs above show the original results with cohorts that comprehend individuals that were born in a time frame of 10 years, while the interval is lowered to 5 years in the graphs below (which is the same choice made by Hertz et al. (2007) among other authors). Both figures confirm that the results are not affected by different cohorts' division. Indeed, the regression coefficient keeps a downward trend across cohorts from 0.90 to 0.39 (it was from 0.87 to 0.44 in the original analysis) while the Spearman's rank coefficient is almost steady going from 0.50 to 0.54. The same conclusion is obtained looking at the other indices computed (in table 22 in appendix B.2).

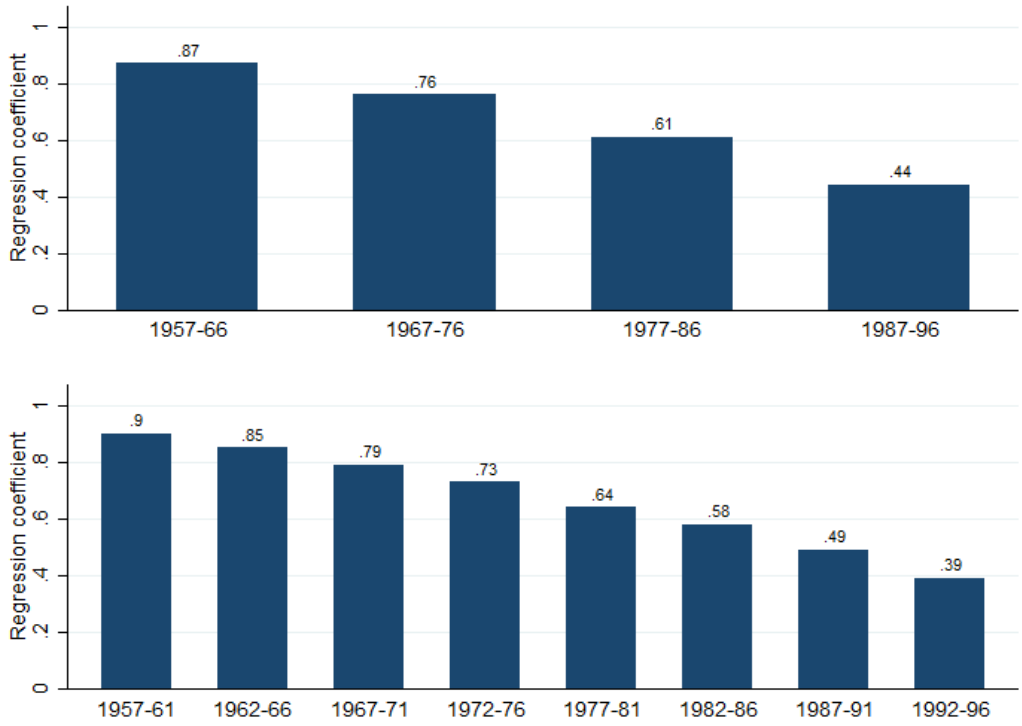

Figure 10: Regression coefficient per cohort.

Source: ENCV 2018, Colombia (DANE). Own estimations. Note: Each cohort groups individuals born every 10 (above) or 5 (below) years.

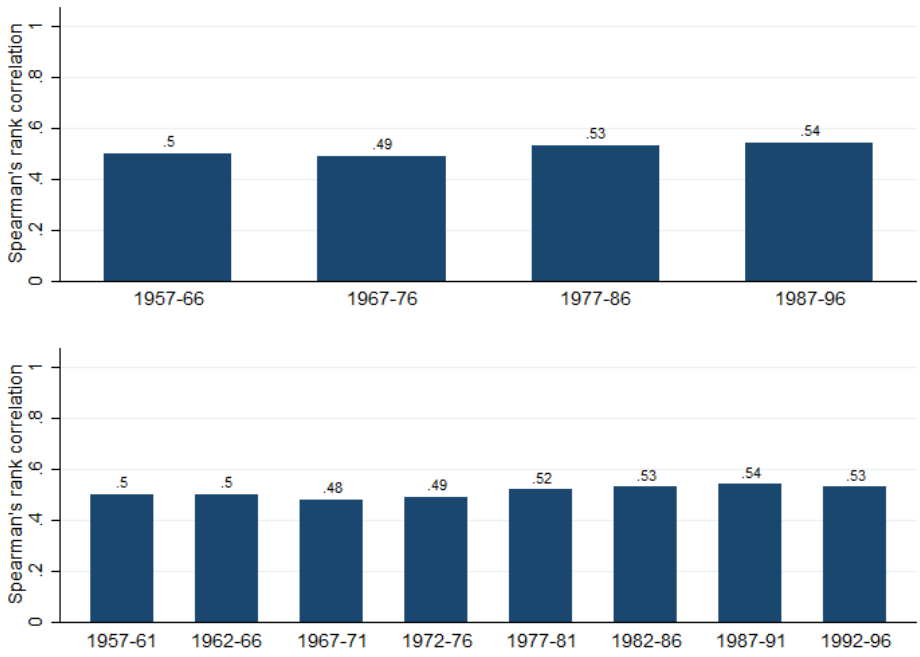

Figure 11: Spearman's rank correlation per cohort.
Source: ENCV 2018, Colombia (DANE). Each cohort groups individuals born every 10 (above) or 5 (below) years.

6.3 Missing values

The sample used in the analysis comprehends 129,772 individuals although the ENCV 2018 is composed of 143,097 individuals between 22 and 61 years old. The 13,325 missing observations are due to the fact that information about the education of both parents is missing. The missing values correspond to 9.31% of the original sample and are likely to be non-random: indeed, missing information in the survey is more likely to appears in families with a low socioeconomic status. For instance, individuals with no information about parental education are on average less educated, and they are more likely to be from an ethnic minority. Therefore, the exclusion of those individuals from the sample used for the analysis could create a selection bias, underrepresenting individuals with a lower level of education. For this reason, I here show the results obtained repeating the analysis assuming that all the individuals excluded from the sample were indeed from a family with low socioeconomic status (i.e. with the lowest possible level of parental education, which is zero).

Figures 12 and 13 compare the regression coefficient and the Spearman's rank correlation excluding (i.e. original analysis) and including the individuals whose information about parental education is missing. Both indices slightly decrease for each cohort when the missing observations are included, but the overall trends do not change; similar results are obtained with the other indices (table 23 in appendix B.3). Although the lack of information in a survey is more likely for people from low socioeconomic status, replacing all the missing values in parental education with the minimum level possible (i.e. zero years of education) relies on a strong assumption that is unlikely to be true. Nevertheless, the robustness check is informative because it shows that even in the most

extreme scenario the bias that is generated by excluding the observations with missing values does not invalidate the results obtained.

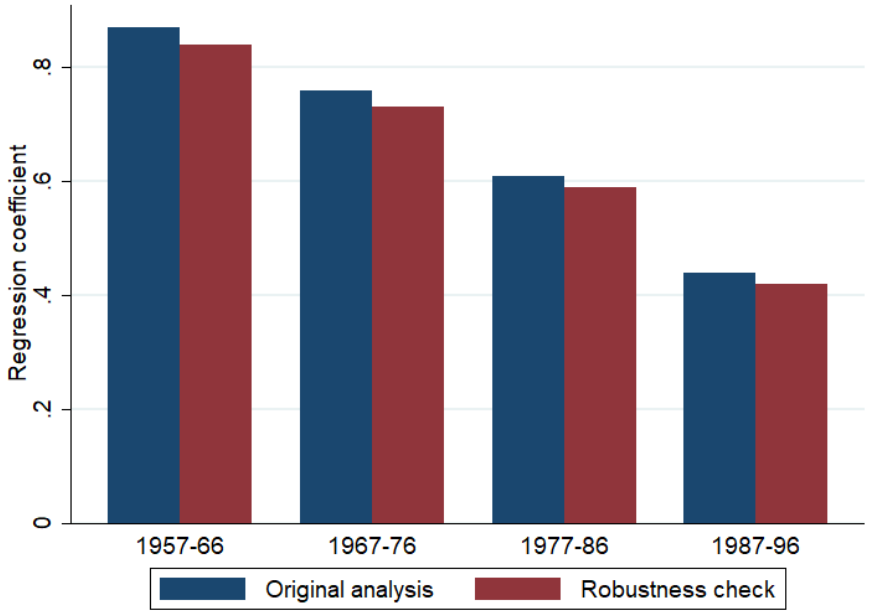

Figure 12: Regression coefficient per cohort excluding and including the observations with missing values.
Source: ENCV 2018, Colombia (DANE).

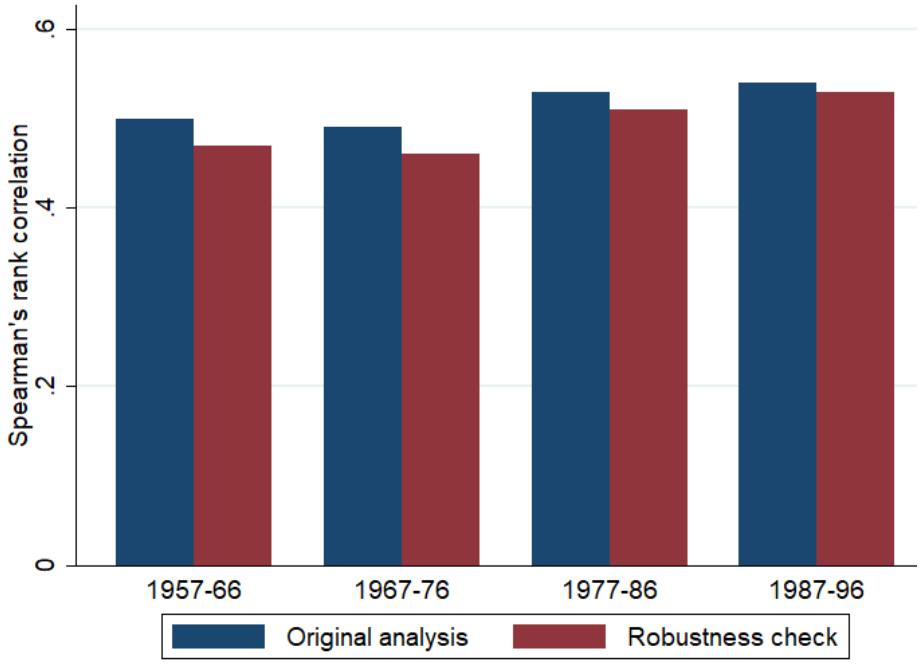

Figure 13: Spearman's rank correlation per cohort excluding and including the observations with missing values.
Source: ENCV 2018, Colombia (DANE)

7.Conclusions

This thesis presents a comprehensive picture of the trends in intergenerational mobility of education in Colombia during the last decades, with a focus on the concept of inequality of opportunity. With this purpose in mind, the thesis contributes to the previous literature computing several indices in order to go beyond the mere regression coefficient commonly used by the other authors. In this way, it was possible to discern among the different patterns of absolute (structural) and exchange (positional) mobility, distinguishing between the improvements generated by a general increase in schooling achievements and the ones that have their roots in an expansion of the opportunities for the most disadvantaged groups. Therefore, the work permits to expand the previous scarce literature on the topics of inequality of opportunity and intergenerational transmission that usually focused on developed countries or in cross-countries analysis (Hertz, et al., 2007; Neidhofer, et al., 2018). The focus on Colombia over a span of 40 years and the computation of several different indices permit to have a broad picture of the trend of intergenerational mobility in the country, issue not extensively studied by the previous literature.

The regression coefficient between parents' and children's educations is found to be decreasing across cohorts, but still keeping its high statistical and economical significance even for the youngest individuals (born between 1987 and 1996). Indeed, one additional year in average parental education is still correlated with 0.44 years more in children's education (whereas the value was equal to 0.88 for individuals born between 1957 and 1966). The result is in line with previous literature about Latin America (Neidhofer, et al., 2018) and it is expected because of the improvements Colombia went through in terms of public expenditure in education, school enrolment rates, and public policies towards more disadvantaged groups (i.e. indigenous, families with low socioeconomic status, individuals living in rural areas) (Government of Colombia, 2014).

Nevertheless, the situation appears more complex when other indices are considered, such as the index r (i.e. the Pearson's correlation coefficient when no controls are included) and the Spearman's rank correlation. These indices, more than the regression coefficient, permit to highlight the concept of exchange mobility rather than the one of absolute mobility. Indeed, the index r controls for changes in inequality between the two generations, whereas the Spearman's rank correlation uses information about the ranks in their respective distributions, capturing uniquely changes in exchange probability. Differently to the regression coefficient, the index r does not change across cohorts, while the Spearman's rank correlation slightly increases by 0.04 points; both indices lay between 0.49 and 0.54 in all cohorts considered. The different trend of the latter two indices compared to the regression coefficient suggests that the mobility experienced in Colombia in the last years has been mostly structural, while no important improvements were made in terms of inequality of opportunity. Indeed, Colombia has experienced a general increase in schooling achievements in the last decades, meaning that children are usually more educated than their parents, and this has particularly benefited individuals from disadvantaged families since the regression coefficient decreases over time (Hertz,

et al., 2007). Nevertheless, the improvements did not decrease the relationship between parents' and children education, with the family background still being a strong predictor for individual schooling achievements: one standard deviation in parents' education corresponds to a change of 0.50 standard deviations in individual education, and the correlation has not changed for individual born in a time frame of forty years.

Eventually, the transition matrices between parents' and children's educations and the indices derived from them gave further information about the distribution of intergenerational mobility. Indeed, average values as the previous indices could overlook important phenomena that happen in different parts of the distribution. This is the case of the so-called "*sticky floor*" and "*sticky ceiling*" that represent higher immobility respectively at the bottom and the top of the distribution (OECD, 2018). For instance, the upper-class persistence index showed that in Colombia there is a high persistence in the top of the distribution that has not decreased over time: individuals with highly educated parents (i.e. with middle education or more) have an extremely high probability (over 90%) to be highly educated themselves. On the other hand, although still significant, the immobility for individuals from non-educated parents has decreased across time, with a probability to be highly educated going from 30% to 63% for individuals from non-educated families. Nevertheless, when the analysis of these indices is repeated dividing the population into quintiles instead of levels of education, the lack of changes in the level of inequality of opportunity becomes clear again: while the first analysis showed an improvement for the individuals from non-educated parents, the indices computed through the quintiles display almost no changes across cohorts. Once again, this discrepancy validates the two different trends in absolute and exchange mobility found with the first indices.

To sum up, the results obtained with the different indices present a picture that confirms the improvements that Colombia experienced in terms of average schooling achievements in the last decades. Moreover, both the decrease of the regression coefficient and the increase of the bottom-upward mobility are consistent with the substantial improvements in the first cycles of the educational system, meaning that the public policies towards education were particularly beneficial for the most disadvantaged groups. Nevertheless, the indices show no improvement in terms of exchange mobility. The final result is that individuals from low-educated parents are today more likely to be more educated than their parents, but the probability to be in the same position in the distribution than their parents has not changed across generations. Thus, the lack of opportunities for individuals from disadvantaged families creates a poverty trap that has not been modified by the recent public policies in the country, with almost no changes in terms of inequality of opportunity.

Future research on the topic is needed in order to understand the causes behind the phenomena described above. Indeed, the previous empirical literature on other countries found several variables to significantly interact on the relationship between parents' and children education but the importance of each factor varies depending on the society of study (Behrman, et al., 2000); therefore, it becomes important to understand which

elements could decrease the level of transmission of education across generations in Colombia. The issue would be particularly important in order to design future public policies in line with the objectives of the Colombian Government of more equality and educational opportunities for everyone in the society regardless of their origins (Government of Colombia, 2019).

References

Acemoglu, D., Johnson, S. & Robinson, J. A., 2002. Reversal of fortune: geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics,* 117(4), pp. 1231-1294.

Banejeree, A. V. & Newman, A. F., 1993. Occupational choice and the process of development. *Journal of Political Economy,* 101(2), pp. 274-298.

Behrman, J., 1996. The impact of health and nutrition on education. *World Bank Research Observer,* 11(1), pp. 23-37.

Behrman, J. R., Birdsall, N. & Skékely, M., 2000. *Intergenerational mobility in South America.* Washington DC, Stiglitz Summer Research Workshop on poverty and development .

Biagi, F. & Stuhler, J., 2018. *A review of intergenerational mobility and its drivers*, Luxembourg: Publications Office of the European Union.

Binder, M. & Woodruff, C., 2002. Inequality and intergenerational mobility in schooling: the case of Mexico. *Economic development and cultural change,* 50(2), pp. 249-267.

Black, S. E. & Devereux, P. J., 2010. Recent developments in intergenerational mobility. *Institute for the Study of Labor (IZA), IZA discussion papers.*

Blanden, J., 2013. Cross-country rankings in intergenerational mobility: a comparison of approaches from economics and sociology. *Journal of Economic Surveys,* 27(1), pp. 38-73.

Bouillon, C. P. & Tejerina, L., 2007. *Do we know what works? A systematic review of impact evaluations of social programs in Latin America and the Caribbean,* Washington D.C.: Inter-American Development Bank.

Brunori, P., Ferreira, F. H. G. & Peragine, V., 2013. Inequality of opportunity, income inequality and economic mobility: some international comparisons. *IZA Discussion paper.*

Carneiro, P. & Heckman, J. J., 2003. *Human capital policy,* Bonn: Insitute for the Study of Labor (IZA).

Corpoeducación, 2006. *Informe de progreso educativo de Colombia,* s.l.: s.n.

Currie, J., 2009. Healthy, wealthy and wise: socioeconomic status, poor health in childhood and human capital development. *Journal of Economic Literature,* 47(1), pp. 87-122.

Currie, J. & Vogl, T., 2013. Early-Life health and adult circumstance in developing countries. *Annual review of economics,* 5(1), pp. 1-36.

De Gregorio, J., 1996. Borrowing constraints, human capital accumulation, and growth. *Journal of Monetary Economics,* 37, pp. 49-71.

Deininger, K. & Squire, L., 1998. New way of looking at old issues: inequality and growth. *Journal of Development Economics,* 57, pp. 259-287.

Demirguc-Kunt, A. & Levine, R., 2008. Finance, financial sector policies, and long-run growth. *Commission on growth and development working paper,* 11.

Departamento Administrativo Nacional de Estadistica, 2019. *Metodologia general encuesta nacional de calidad de vida - ECV,* Bogotá D.C.

Ferreira, F. H. G., Gignoux, J. & Aran, M., 2011. Measuring inequality of opportunity with imperfect data. The case of Turkey. *The Journal of Economic Inequality,* 9, pp. 651-680.

Ferreira, F. H. & Gignoux, J., 2011. The measurement of inequality of opportunity: theory and an application to Latin America. *Review of Income and Wealth,* 57(4), pp. 622-658.

Fogel, R. W., 2004. *The escape from hunger and premature death, 1700-2100. Europe, America and the Third World.* Cambridge: Cambridge University Press.

Galor, O. & Zeira, J., 1993. Income distribution and macroeconomics. *The Review of Economic Studies,* 60(1), pp. 35-52.

Government of Colombia, 2014. *Bases del Plan Nacional de Desarrollo 2014-2018. Todos por un nevo país: paz, equidad, educación,* Bogotá, D.C., Colombia: Departamento Nacional de Planeación.

Government of Colombia, 2019. *Bases del Plan Nacional de Desarrollo 2018-2022. Pacto por Colombia, pacto por la equidad,* Bogotà, D.C., Colombia: Departamento Nacional de Planeaciòn.

Heckman, J. J., 2012. Promoting social mobility. *Boston Review,* 1 September.

Heineck, G. & Riphahn, R. T., 2007. Intergenerational transmission of educational attainment in Germany: the last five decades. *DIW Discussion Papers,* Volume 738.

Hertz, T. et al., 2007. The inheritance of educational inequality: international comparisons and fifty-year trends. *Advances in Economic Analysis and Policy,* 7(2).

Legovini, A. & Regalia, F., 2001. *Targeted Human Development Programs: investing in the next generation,* Washington, D.C.: Inter-American Development Bank.

Londoño, J. L., 1996. *Poverty, inequality, and human capital development in Latin America, 1950-2025,* Washington, D.C. : The World Bank.

Lutbotsky, D. & Wittenberg, M., 2006. Interpretation of regressions with multiple proxies. *The Review of Economics and Statistics,* 88(3), pp. 549-562.

Malthus, T. R., 1803. *An essay on the principle of population.* 2nd ed. Cambridge: Winch, D.

Marrero, G. A. & Rodríguez, J. G., 2013. Inequality of opportunity and growth. *Journal of Development Economics,* 104, pp. 107-122.

Ministerio de la Educación de Colombia, 2020. *Ministerio de la Educación.* [Online] Available at: https://www.mineducacion.gov.co/portal/

Neidhofer, G., Serrano, J. & Gasparini, L., 2018. Educational inequality and intergenerational mobility in Latin America: a new database. *Journal of Development Economics* , Volume 134, pp. 329-349.

OECD, 2016. *Educaciòn en Colombia,* Paris: Organisation for Economic Cooperation and Development (OECD).

OECD, 2018. *A broken social elevator? How to promote social mobility,* Paris: OECD Publishing.

Paes de Barros, R., Ferreira, F. H., Molinas Vega, J. R. & Saavedra Chanduvi, J., 2008. *Midiendo la desigualdad de oportunidades en América Latina y el Caribe,* Washington DC: World Bank.

Prados de la Escosura, L., 2007. Inequality and poverty in Latin America: a long-run exploration. In: T. J. Hatton, K. H. O'Rourje & A. M. Taylor, eds. *The new comparative economic history.* Cambridge, Mass.: MIT Press, pp. 291-315.

Ram, R., 1990. Educational expansion and schooling inequality: international evidence and some implications. *Review of Economics and Statistics,* 72(2), pp. 266-274.

Richey, J. & Rosburg, A., 2017. Decomposing economic mobility transition matrices. *Journal of Applied Econometrics,* 33(1), pp. 91-108.

Robeyns, I., 2005. The Capability Approach: a theoretical survey. *Journal of Human Development,* 6(1), pp. 93-115.

Roemer, J., 1998. *Equality of opportunity.* Cambridge, M.A.: Harvard University Press.

Roemer, J., 2004. Equal opportunity and intergenerational mobility: going beyon intergenerational income transition matrices. In: *Generational Income Mobility in North America and Europe.* New York: Corak, M., pp. 217-300.

Roemer, J. E. & Trannoy, A., 2015. Equality of Opportunity. In: *Handbook of Income Distribution.* s.l., pp. 218-300.

Royuela, V. & Garcìa, G. A., 2015. Economic and social convergence in Colombia. *Regional Studies,* 49(2), pp. 219-239.

Sen, A., 1980. Equality of what?. In: *The Tannel Lectures on Human Values.* Salt Lake City, Utah: S. McMurrin.

Shukla, V. & Mishra, U. S., 2019. Educational expansion and schooling inequality: testing educational Kuznets curve for India. *Social Indicators Research,* 141, pp. 1265-1283.

The World Bank, 2006. *World development report. Equity and development,* Washington DC: Oxford University Press.

Appendix

Appendix A.1

Table 11: Regression coefficient, standard deviations of parents' and individual's education and correlation coefficient r per cohort.

	Cohort 1957-66	Cohort 1967-76	Cohort 1977-86	Cohort 1987-96
Regression coefficient	0.87	0.76	0.61	0.44
sd(parents' education)	3.05	3.27	3.94	4.43
sd(Individuals' education)	5.23	5.03	4.73	3.92
Correlation coefficient *r*	0.51	0.50	0.51	0.50

Source: ENCV 2018, Colombia (DANE).
Note: The variables presented are the following: LW variable, maximum parental level of education, average parental education, father's education, and mother's education.

Appendix A.2

Table 12: Transition matrices by quintile per cohort.

		PARENTS' EDUCATION									
		Cohort 1957-66					Cohort 1967-76				
		None	Primary	Secondary	Middle	Tertiary	None	Primary	Secondary	Middle	Tertiary
INDIVIDUAL EDUCATION	None	38.19	29.93	14.91	13.14	3.83	38.85	25.24	20.62	10.37	4.92
	Primary	24.43	24.46	22.99	21.86	6.26	24.59	24.39	25.92	17.80	7.29
	Secondary	19.43	22.95	24.32	20.20	13.12	18.81	23.10	21.70	22.00	14.39
	Middle	12.41	14.37	22.75	23.77	26.69	12.30	16.40	20.07	26.17	25.06
	Tertiary	5.53	8.29	15.03	21.04	50.11	5.45	10.87	11.68	23.66	48.34
		Cohort 1977-86					Cohort 1987-96				
		None	Primary	Secondary	Middle	Tertiary	None	Primary	Secondary	Middle	Tertiary
INDIVIDUAL EDUCATION	None	41.85	24.74	21.40	9.05	2.96	42.92	30.08	17.74	6.81	2.44
	Primary	25.99	26.92	25.29	16.67	5.12	23.98	25.50	23.11	18.37	9.04
	Secondary	15.99	21.93	21.19	25.22	15.67	15.66	21.51	23.94	24.26	14.63
	Middle	11.98	17.92	20.33	26.79	22.98	12.66	15.59	21.78	26.23	23.75
	Tertiary	4.19	8.49	11.78	22.26	53.27	4.78	7.32	13.43	24.34	50.14

Table 13: Transition matrices by quintile (second computation) per cohort.

		PARENTS' EDUCATION									
		Cohort 1957-66					Cohort 1967-76				
		None	Primary	Secondary	Middle	Tertiary	None	Primary	Secondary	Middle	Tertiary
INDIVIDUAL EDUCATION	None	42.57	24.98	17.89	10.78	4.09	43.32	28.85	22.97	13.09	5.58
	Primary	16.32	17.47	17.52	10.92	3.36	27.65	30.99	29.40	23.49	8.92
	Secondary	25.29	32.53	30.88	26.41	13.73	11.77	14.65	16.28	16.35	12.16
	Middle	12.22	18.02	23.29	29.97	29.44	12.51	16.62	19.85	25.51	25.54
	Tertiary	3.61	6.99	10.43	21.92	49.37	4.75	8.89	11.50	21.56	47.81
		Cohort 1977-86					Cohort 1987-96				
		None	Primary	Secondary	Middle	Tertiary	None	Primary	Secondary	Middle	Tertiary
INDIVIDUAL EDUCATION	None	37.33	20.08	17.51	7.14	2.52	44.40	29.92	17.12	6.22	2.47
	Primary	23.66	24.22	21.03	12.50	3.05	19.67	21.05	17.77	11.92	5.07
	Secondary	13.36	15.44	14.82	13.43	5.87	26.96	34.53	40.45	40.06	23.03
	Middle	19.93	29.08	31.76	39.77	29.27	6.06	9.06	14.53	21.12	22.72
	Tertiary	5.73	11.19	14.89	27.17	59.28	2.91	5.44	10.12	20.68	46.71

Appendix A.3

Table 14: Absolute (M1) and directional mobility (M2) per cohort

	Cohort 1957-66	Cohort 1967-76	Cohort 1977-86	Cohort 1987-96
M1	5.01	5.45	5.91	5.72
M2	4.61	5.10	5.50	5.14

Source: ENCV 2018, Colombia (DANE).

Appendix B.1

Table 15: Regression coefficient of the five proxies for parental educational background per cohort

	Cohort 1957-66	Cohort 1967-76	Cohort 1977-86	Cohort 1987-96
LW	0.87	0.76	0.61	0.44
Max	0.71	0.61	0.50	0.36
Average	0.87	0.76	0.61	0.44
Father	0.71	0.62	0.50	0.35
Mother	0.80	0.66	0.53	0.37

Source: ENCV 2018, Colombia (DANE).

Table 16: Correlation coefficient r of the five proxies for parents' education per cohort

	Cohort 1957-66	Cohort 1967-76	Cohort 1977-86	Cohort 1987-96
LW	0.51	0.49	0.51	0.50
Max	0.49	0.48	0.49	0.47
Average	0.51	0.50	0.51	0.50
Father	0.48	0.46	0.46	0.43
Mother	0.48	0.46	0.48	0.46

Table 17: Immobility ratios of the five proxies for parents' education per cohort (level of education).

	Cohort 1957-66		Cohort 1967-76		Cohort 1977-86		Cohort 1987-96	
	IR_1	IR_2	IR_1	IR_2	IR_1	IR_2	IR_1	IR_2
LW	0.39	0.70	0.35	0.64	0.29	0.56	0.24	0.55
Max	0.39	0.71	0.36	0.65	0.30	0.58	0.26	0.58
Average	0.39	0.70	0.35	0.64	0.28	0.55	0.23	0.53
Father	0.36	0.68	0.32	0.62	0.25	0.51	0.20	0.47
Mother	0.35	0.68	0.31	0.61	0.26	0.53	0.22	0.52

Source: ENCV 2018, Colombia (DANE).

Note: The computation is done dividing the individuals in five groups depending on the maximum level of education obtained.

Table 18: Immobility ratios of the five proxies for parents' education per cohort (quintiles)

	Cohort 1957-66		Cohort 1967-76		Cohort 1977-86		Cohort 1987-96	
	IR_1	IR_2	IR_1	IR_2	IR_1	IR_2	IR_1	IR_2
LW	0.32	0.70	0.32	0.70	0.34	0.72	0.34	0.72
Max	0.32	0.70	0.32	0.70	0.33	0.71	0.32	0.71
Average	0.32	0.70	0.32	0.70	0.33	0.71	0.33	0.72
Father	0.31	0.69	0.31	0.68	0.32	0.69	0.32	0.69
Mother	0.31	0.69	0.31	0.69	0.34	0.70	0.32	0.70

Source: ENCV 2018, Colombia (DANE).
Note: The computation is done dividing the individuals into quintiles depending on the number of years of education.

Table 19 BUM and UCP of the five proxies for parents' education per cohort (level of education)

	Cohort 1957-66		Cohort 1967-76		Cohort 1977-86		Cohort 1987-96	
	BUM	UCP	BUM	UCP	BUM	UCP	BUM	UCP
LW	0.30	0.94	0.38	0.95	0.51	0.96	0.63	0.96
Max	0.29	0.91	0.37	0.92	0.51	0.94	0.62	0.94
Average	0.30	0.94	0.38	0.95	0.52	0.96	0.63	0.96
Father	0.31	0.94	0.39	0.94	0.53	0.96	0.65	0.96
Mother	0.31	0.91	0.39	0.94	0.53	0.94	0.63	0.95

Source: ENCV 2018, Colombia (DANE).
Note: The computation is done dividing the individuals in five groups depending on the maximum level of education obtained.

Table 20: BUM and UCP of the five proxies for parents' education per cohort (quintiles)

	Cohort 1957-66		Cohort 1967-76		Cohort 1977-86		Cohort 1987-96	
	BUM	UCP	BUM	UCP	BUM	UCP	BUM	UCP
LW	0.26	0.61	0.26	0.62	0.25	0.63	0.25	0.62
Max	0.25	0.62	0.25	0.60	0.27	0.60	0.26	0.61
Average	0.25	0.62	0.25	0.62	0.25	0.62	0.25	0.62
Father	0.27	0.61	0.28	0.60	0.27	0.61	0.28	0.60
Mother	0.29	0.61	0.27	0.61	0.27	0.61	0.27	0.60

Source: ENCV 2018, Colombia (DANE).
Note: The computation is done dividing the individuals in five groups depending on the maximum level of education obtained.

Table 21: M1 and M2 of the five proxies for parents' education per cohort.

	Cohort 1957-66		Cohort 1967-76		Cohort 1977-86		Cohort 1987-96	
	M1	M2	M1	M2	M1	M2	M1	M2
LW	5.01	4.61	5.45	5.10	5.91	5.50	5.72	5.14
Max	4.73	4.06	5.10	4.43	5.49	4.68	5.31	4.09
Average	5.00	4.60	5.45	5.09	5.91	5.50	5.73	5.16
Father	5.12	4.63	5.61	5.16	6.20	5.66	6.20	5.41
Mother	5.13	4.67	5.58	5.13	5.99	5.44	5.83	4.97

Source: ENCV 2018, Colombia (DANE).

Appendix B.2

Table 22: Mobility indices dividing the sample into ten cohorts

	Cohort 1957-61	Cohort 1962-66	Cohort 1967-71	Cohort 1972-76	Cohort 1977-81	Cohort 1982-86	Cohort 1987-91	Cohort 1992-96
Regression coefficient	0.90	0.85	0.79	0.73	0.64	0.58	0.49	0.39
r	0.51	0.51	0.49	0.50	0.50	0.51	0.51	0.48
Spearman's rank correlation	0.50	0.50	0.48	0.49	0.52	0.53	0.54	0.53
IR_1	0.40	0.38	0.36	0.34	0.30	0.27	0.25	0.23
IR_2	0.72	0.69	0.66	0.62	0.57	0.54	0.54	0.55
BUM	0.27	0.31	0.35	0.40	0.48	0.55	0.60	0.65
UCP	0.93	0.94	0.92	0.96	0.95	0.96	0.97	0.96
IR_1 (quintiles)	0.31	0.32	0.31	0.32	0.33	0.33	0.34	0.32
IR_2 (quintiles)	0.71	0.70	0.79	0.70	0.70	0.71	0.72	0.70
BUM (quintiles)	0.26	0.26	0.26	0.25	0.26	0.26	0.26	0.26
UCP (quintiles)	0.60	0.61	0.60	0.62	0.61	0.61	0.61	0.60
M1	4.86	5.13	5.33	5.57	5.84	5.97	5.88	5.58
M2	4.44	4.74	4.98	5.22	5.44	5.56	5.38	4.93

Source: ENCV 2018, Colombia (DANE).

Note: The indices in the table are regression coefficient, correlation coefficient r, Spearman's rank correlation, immobility ratios, bottom-up mobility, upper-class persistence, absolute mobility, and directional mobility. The IR, BUM, and UPC indices are computed first dividing the sample in classes depending on the highest level of education obtained and then dividing it in quintiles.

Appendix B.3

Table 23: Mobility indices including observations with missing values.

	Cohort 1957-66	Cohort 1967-76	Cohort 1977-86	Cohort 1987-96
Regression coefficient	0.84	0.73	0.59	0.42
r	0.43	0.48	0.50	0.48
Spearman's rank correlation	0.47	0.46	0.51	0.53
IR_1	0.36	0.33	0.27	0.22
IR_2	0.70	0.64	0.55	0.53
BUM	0.28	0.36	0.50	0.62
UCP	0.94	0.94	0.96	0.96
IR_1 (quintiles)	0.31	0.30	0.32	0.33
IR_2 (quintiles)	0.68	0.69	0.70	0.71
BUM (quintiles)	0.27	0.26	0.26	0.27
UCP (quintiles)	0.59	0.60	0.61	0.60
M1	5.13	5.59	6.10	5.94
M2	4.78	5.27	5.72	5.40

Source: ENCV 2018, Colombia (DANE).
Note: The indices in the table are regression coefficient, correlation coefficient r, Spearman's rank correlation, immobility ratios, bottom-up mobility, upper-class persistence, absolute mobility, and directional mobility. The IR, BUM, and UPC indices are computed first dividing the sample in classes depending on the highest level of education obtained and then dividing it in quintiles.